# A Distributed Job Scheduler Virtual Machine-based Data Centres

**K. Suseela[1] Mr. T. Vinoth Kumar[2]**
[1]M. E. [2]M. B. A. (Assistant Professor)
[1]Computer Science Engineering
[1, 2]RatnaVel Subramaniam College of Engineering & Technology

*Abstract*---The existing techniques by turning on or off servers with the help of virtual machine (shortly called as VM) migration are not enough. Instead, finding an optimized dynamic resource allocation method to solve the problem of on-demand resource provision for VMs is the key to improve the efficiency of data centers. However, the existing dynamic resource allocation methods only focus on either the local optimization within a server or central global optimization, limiting the efficiency of data centers. A two-tiered on-demand resource allocation mechanism consisting of the local and global resource allocation with feedback to provide on-demand capacities to the concurrent applications. On-demand resource allocation using optimization theory based on the proposed dynamic resource allocation mechanism and model, It propose a set of on-demand resource allocation algorithms. This algorithm preferentially ensures performance of critical applications named by the data center manager when resource competition arises according to the time-varying capacity demands and the quality of applications. Using Rainbow, a Xen-based prototype implemented, it evaluates the VM-based shared platform as well as the two-tiered on-demand resource allocation mechanism and algorithms. The experimental results show that Rainbow without dynamic resource allocation (Rainbow-NDA) provides twenty six to three twenty four percent improvements in the application performance, as well as twenty six percent higher average CPU utilization than traditional service computing framework, in which applications use exclusive servers. The two-tiered on-demand resource allocation further improves performance by nine to sixteen percent for those critical applications, seventy five percent of the maximum performance improvement, introducing up to five percent performance degradations to others, with one to five percent improvements in the resource utilization in comparison with Rainbow-NDA.

**Keywords:** Data centers, virtual machines, on-demand resource allocation, optimization, algorithm, model.

## I. INTRODUCTION

The rapid growth of data centers in the past few years, and expect the number of datacenters will triple by 2020. Fighting for green data centers is the biggest challenge faced by computer scientists and practitioners. Google argues that more requests served by the same platform is another path to green data centers. Thus, one of the efficient solutions for the green data centers is improving the throughput as well the resource utilization by server consolidation based on virtualization technology, which is verified by several previous efforts and our previous work. With effective isolation and agile resource management provided by virtualization technology, virtualized data center is also the infrastructure of most cloud platforms. In such a shared virtual computing environment, dynamic load changes as well as different quality requirements of applications in their lifetime give rise to dynamic and various capacity demands which result in lower resource utilization and lower application quality using the existing static resource allocation.

Furthermore, the total required capacities of all the hosted applications in current enterprise data centers may surpass the capacities of the platform. Thus, the existing techniques by turning on or off servers with the help of virtual machine (VM) migration is not enough Instead, finding an optimized dynamic resource allocation method in server consolidation scenarios to solve the problem of on-demand resource allocation is the key to improve the resource utilization and throughput, so as to optimize power efficiency of data centers.

## II. PROBLEM STATEMENT

Nowadays, real-time data packet sources are in care of providing security services to their real-time applications, making them robust against different security threats especially in local-area network (LAN), where most of the hacking processes occurred at the network's edge. In order to provide such security services on the network's data streams, different security protocols were implemented such as the secure sockets layer protocol (SSL), the transport layer security protocol (TLS), and the internet protocol security (IP sec). With the current security protocols, any dynamic change in the network can not affect the pre-negotiated security level. Therefore, network performance issues are not taken into account and the QoS may not be guaranteed for different classes of real-time data streams. Such results may lead into a catastrophe especially for those hard real-time network applications.

While providing such security services to its real-time network applications, service provider should keep a balance between guaranteeing such security services and preserving the overall performance of the network.

The overall performance of a real-time network could be measured by different network performance metrics such as miss rate, total average packets delay, functionality, jitter, and throughput. A key factor that affects such NPMs, and thus controls the overall performance of the network is the best utilization of the network's queuing system, which regulates the total amount of traffic load in the network and thus, limits the maximum throughput in the network. Accordingly, different network-based algorithms were implemented based on network buffer estimation models such as routing, scheduling, maintenance, load balancing and security.

## III. EXISTING METHOD

The lower resource utilization and application quality using the existing static resource allocation. The existing techniques by turning on or off servers with the help of virtual machine (VM) migration are not enough. The existing dynamic resource allocation methods only focus on either the local optimization within a server (or) central global optimization. Dynamic load changes as well as different quality requirements of applications in their lifetime give rise to dynamic and various capacity demands. The total required capacities of all the hosted applications in current enterprise data centers. To optimize the usage of resources and improve the quality of the hosted applications. These algorithms dynamically allocate resources to VMs according to the time-varying capacity demands and the quality requirements of applications.

### *A. Drawbacks:*

The lower resource utilization and application quality using the existing static resource allocation. The existing virtualization technology in such shared environment. The existing dynamic resource allocation methods only focus on either the local optimization. It is low complexity and slow response of resource allocation. Any scheduler's failure could not lead to change the failure of the resource allocation in the system.

## IV. PROPOSED METHOD

It proposes a two-tiered on-demand resource allocation mechanism consisting of the local and global resource allocation with feedback to provide on-demand capacities to the concurrent applications. A novel two-tiered on-demand resource allocation mechanism with feedback for VM-based data. It designs a set of algorithms to control the dynamic resource allocation among VMs according to the time-varying resource demands and quality goals of the hosted applications. It will propose and evaluate distributed resource allocation algorithms to control on-demand resource allocation among VMs each of which may use resources in different servers.

### *A. Avantage*

It proposes a virtual-appliance-based autonomic resource provisioning framework for large virtualized data centers. Its define a non-linear constrained optimization model for dynamic resource provisioning and present its novel analytic solution. The result in the mismatch between the on-demand resource allocation and the workloads dispatch.

## V. SYSTEM ARCHITECTURE

This paper has the following contributions:

(1) We propose a novel two-tiered on-demand resource allocation mechanism with feedback to optimize the resource allocation for VM-based data centres.
(2) In order to guide the design of the on-demand resource allocation algorithm, we model the resource allocation using optimization theory.
(3) Base on the two-tiered on-demand resource allocation mechanism and model, we propose local and global resource allocation algorithms to optimize the dynamic resource provision for VMs.

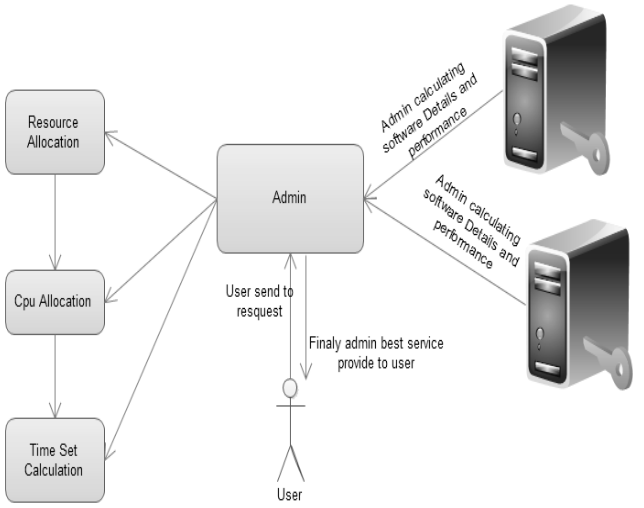

Fig. 1: The system Architecture

## VI. TWO-TIERED ON-DEMAND RESOURCE ALLOCATION MECHANISM AND CONTROL MODEL

We aim to propose on-demand resource allocation strategies. A resource allocation strategy should solve our problems: Which resource will be allocated? When will such resource be allocated? Which VMs will be the source or target of allocating? How many resources will be allocated? To answer these problems, we first design the resource allocation mechanism, then we use control theory to model and design our feedback-driven two-tiered resource allocation.

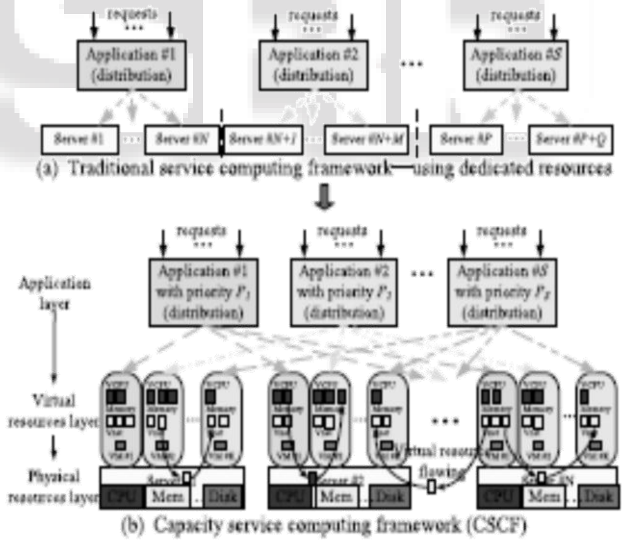

Fig. 2: The evolution of service computing framework

Inspired by the multilevel optimization-based techniques for managing distributed computing systems (e.g., Cluster Reserves, Neptune, MBRP, and Kandasamy et al.'s work), we introduce the multilevel management idea into the on-demand resource allocation in VM-based computing environment. Virtualization Technology can encapsulate an application into an operating system and dynamically allocate resources to the hosted applications agilely with more abundant feedback information, for example, characteristics of application workloads and resource utilizations of VMs, compared with the traditional large-scaled distributed computing environment. Kandasamy et al. propose a structure of three level controllers aimed at operating the cluster in energy efficient fashion while satisfying the QoS goal. In their work, only one level controller controls the resource allocation in each computer, and other two levels are responsible for workload distribution. However, we do not care about the workload distribution, our two-tiered controllers both focus on the resource allocation in a virtualized computing environment. Two-Tiered Resource Allocation Mechanism The goal of on-demand resource allocation is to optimize the dynamic resource provision for VMs. illustrates the two-tiered on-demand resource allocation mechanism as well as the traditional resource management. After comparing, we can easily find that the two-tiered on-demand resource allocation mechanism differs from the traditional resource management in adding a resource management level for VMs. Each application ("application #1" ..."application #S" has multiple instance copies each of which is encapsulated in a VM. The VMs hosting instance copies of the same application constitute an application domain. Each server hosts VMs belonging to multiple application domains. In such a scenario, workloads in VMs residing in the same server are time varying and different from each other, resulting in the requirement of dynamic resource allocation among VMs within a server. Furthermore, workloads of each application running on various servers are also time varying and different from other applications, resulting in the requirement of on-demand resource allocation among applications. Based on the technical support on dynamic resource allocation to VMs within a server provided by the current VMMs, we can independently control resource allocation to VMs in each server. However, only such independent control in each server may readily result in unbalance of resource allocation among applications so as to limit the efficiency of data centers. Currently, there is no technological support on the resource allocation by VMM on a server to a VM residing in another server. Thus, it is necessary to provide a global resource optimization based on the existing virtualization technology in such shared environment. We propose a two tiered on-demand resource allocation mechanism implemented by the local resource scheduler and global resource scheduler. The local resource scheduler controls resource allocation to VMs within a server. It adds a set of on-demand resource allocation algorithms based on the technical support on dynamic resource allocation provided by the existing VMMs. To maintain high resource utilization as well as guarantee the quality of applications, the local resource scheduler automatically optimizes the resource allocation to VMs via adjusting CPU time slots and memory assigned to each VM, according to its resource utilization as well as quality and activity of the application hosting in the VM. (Activity denotes the threshold of resource allocation, e.g., if the CPU activity of an application is 90 percent, some CPU resource will be allocated to it when its CPU utilization reaches 90 percent.) To allocate resources in a timely manner, the local resource scheduler works at short intervals. In the local resource scheduler, the activity of an application is the resource utilization threshold. When the resource utilization of a VM hosting such an application reaches the threshold, some resources should be allocated to the application. Thus, activities are the key parameters effecting the resource allocation. The global resource scheduler indirectly controls resource allocation among applications in the entire system by adjusting activities of the applications in each local scheduler. In our scenario, each application may have several copies running on multiple servers so that each application can indiscriminately use resources of these servers via on-demand resource allocation controlled by the local scheduler. Adjusting activities of applications influences resource allocation among VMs within each server, resulting in dynamic resources allocation among these applications. The global resource scheduler systematically optimizes resource allocation among applications by periodically adjusting the activity of each application. Thus, the activities of applications are the bridge between the local resource scheduler and the global resource scheduler in our mechanism.

## VII. CONCLUSION

It evaluates the VM-based shared platform as well as the on-demand resource allocation algorithms using a Xen-based prototype, Rainbow. These results confirm that the two-tiered on-demand resource allocation gains its goal of improving the resource utilization as well as ensuring quality of the hosted applications. The local and global resource schedulers reallocate resources by evaluating the arriving workloads but unaware of the request dispatch of each application, which may result in the mismatch between the on-demand resource allocation and the workloads dispatch. It will propose and evaluate distributed resource allocation algorithms to control on-demand resource allocation among VMs each of which may use resources in different servers concurrently based on the K-VM-N-PM model and our on-going DVMM project. Analyzing the potential of each tier in the resource allocation mechanism is also our future work.

## REFERENCES

[1] Y. Song, H. Wang, and Y. Li, "Multi-tiered on--demand resource scheduling for VM-based data centres," in Proc. IAPR Conf. Mach. Vis. Appl., Tsukuba Science City, Japan, May 2005, pp. 347–350.
[2] Z. Wang and et al, "Capacity and performance overhead in dynamic resource allocation to virtual containers," in Proc. 3rd Int. Conf. BTAS, Washington, DC, Sep. 2007, pp. 149-158.
[3] Z. Wang, C. Weng, and M. Li , "The Hybrid Scheduling Framework for Virtual Machine System," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 1, pp. 111–120, Jan. 2009.
[4] W. Zhao and Z. Wang, "Dynamic Memory Balancing for Virtual Machines," in Proc. IEEE ICPR, 2009, pp. 21-30.
[5] M. Nelson et al, "Fast Transmission Migration for Virtual Machines," IEEE Trans. Biomed. Circuits Syst., vol. 2, no. 4, pp. 391-394, Dec. 2005.
[6] C. Weng. and Y. Zhou, "Characterizing SMP Virtual Machine Scheduling in Virtualization Environment," in Proc. 3rd Int. Conf. BTAS, Washington, DC, Sep. 2008, pp. 1–6.
[7] X. Lu, S. Chatterjee, , M. Nelson, and M. Goldbaum, "Shares and utilities based power consolidation in

virtualized server environments," IEEE Trans. Med. Image., vol. 8, no. 3, pp. 263–269, Sep. 1989.
[8] Encyclopedia of Biometrics, S. Z. Li, Ed.. New York: Springer-Verlag, 2009.
[9] E. C. Lee and K. R. Park, "Appliance-Based Autonomic Provisioning Framework for Virtualized Outsourcing Data Center," Electron. Lett., vol. 45, no. 21, pp. 1074–1076, Oct. 2009.