

# Identifying, Categorizing and Mitigating Performance Interferences in Virtualized Environments – A Survey

A. P. Nirmala  
Sr. Asst. Prof.,

New Horizon College of Engg., Bangalore, India  
Research Scholar,  
Karpagam Academy of Higher Education,  
Coimbatore, India

Dr. S. Veni  
HOD,

Department of Computer Science,  
Karpagam Academy of Higher Education,  
Coimbatore, India

**Abstract**—Virtualization technology consolidates multiple systems on a shared infrastructure by running co-located virtual machines (VMs) on a single physical machine. Though virtualization enables high utilization of hardware resources and adopted in many enterprises, there are still lots of technical challenges that need to be addressed. Specifically, the performance of applications running in one VM affects the performance of application running on its neighbor VMs in the same physical host. Hence, it is a need of the hour to analyze the performance interference in order to mitigate it. This motivates to present a survey on prediction models, types of performance interferences and approaches to mitigate them in virtualized environment.

**Keywords**— *Virtualization, Prediction model, Performance influencing factors, Scheduling algorithm, Performance measures, Performance Interference.*

## I. INTRODUCTION

In the modern era, virtualization is an essential technology adopted in datacenters as well as in novel computing paradigms like Cloud computing [1]. It offers server consolidation through tools such as VMware, XEN, KVM which are widely used in data centers. In virtualization technology, applications with different characteristics and requirements can run inside Virtual Machines (VMs) on the same physical multi-core host, with increased levels of security, fault tolerance, isolation, dynamicity in resource allocation and ease of management.

Virtualization techniques offer security isolation, fault isolation and environment isolation in an effective manner but they do not provide effective performance isolation between virtual machines. Due to this, a number of issues arise such as the degradation of the performance of the applications [3].

Resource sharing promises a more efficient usage of the available hardware. However, as all VMs share the same physical resources, they also mutually influence each other's performance. Specifically, a user running the same virtual machine on the same hardware at different times will see wide disparity in performance based on the work performed by other VMs on that physical host. This phenomenon is referred as performance interference [3]. This influences the analysis of bringing out the reason for performance degradation and also methods for predicting the performance of applications running in virtualized environment.

The proposed survey presented in this paper provides an analysis of performance interferences with the present state of the art of research from academic and industrial perspective. This paper is organized as follows: Section II provides the related work, Section III discusses the proposed survey and Section IV highlights some of the future directions.

## II. RELATED WORK

Y. Koh et al. [3] analyzed performance interference effects caused by system-level characteristics of different workloads. The authors in their work have discussed prediction mechanisms for different combinations of workloads. They highlight the workloads that create intense performance interference Q. Huang et al. in [25] addressed the cascade attacks to find the performance interferences due to various types of hardware resources such as CPU, memory and I/O. The authors in their work have suggested few approaches to defend the attacks. However the research efforts in [3] and [25] do not address the techniques to mitigate interference effects.

From the above literature, it is evident that a survey analyzing different aspects of performance Interference is presently unavailable. This fact becomes the motivation for the proposed survey to discuss the issues which are presented in the next section. It is believed that it will benefit the user community and researchers.

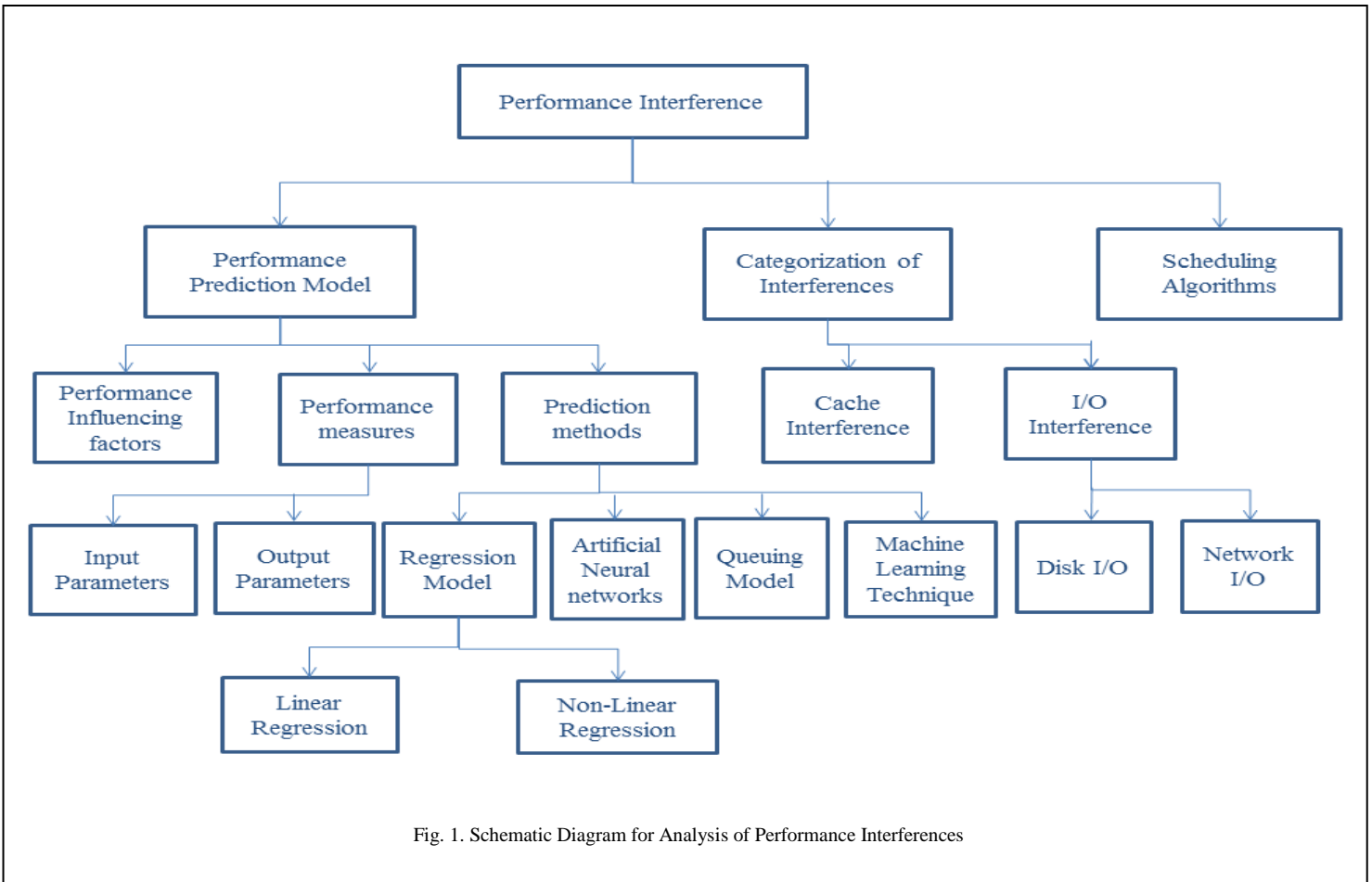


Fig. 1. Schematic Diagram for Analysis of Performance Interferences

### III. IDENTIFYING, CATEGORIZING AND MITIGATING PERFORMANCE INTERFERENCES

The proposed survey discusses performance influencing factors, methods and measures used in performance prediction models, levels of performance interference and variety of Scheduling algorithms to mitigate them are shown in schematic diagram Fig.1.

#### A. Performance Prediction model

Though a lot of research work has been carried out like [17] on performance prediction model in non-virtualized environment, however it does not address the issues arising from virtualized environments. The effects of virtualization on response time, throughput and resource utilization as well as the massive number of resources available, requires an analysis on prediction models.

1) *Performance influencing factors* : N. Huber et al. in [19] have explored that factors like scheduling parameter and workload type which influence performance. The authors in their work propose a generic model which aims at predicting the performance overheads by evaluating these factors on different virtualization platforms.

In this regard, Hauck M. et al. in [18] have proposed an approach to elaborate performance relevant properties for virtualized environments. The authors in their work highlight research challenges that have to be addressed in terms of

performance properties, performance analysis and detecting measurements in virtualized environments. Benevenuto et al. proposes a model in [5], emphasizes on how to build performance models and also have provided metrics to predict the performance of applications running on Xen environment.

A few parameters presented by Kousiouris et al. in [8] influence the performance of virtual machines (VMs). These parameters include allocation percentages, real-time scheduling decisions and co-placement of VMs. The authors in their work experiment with different combinations of VM workload types in order to find the optimal allocation strategies.

2) *Performance Measures*: Performance prediction is the process of estimating performance measures of a computer system for a given set of parameters. Typical performance measures include input and output parameters. The output parameters include response time, throughput, resource utilization and resource queue length. The input parameters to such a model fall into one of the three categories: workload, basic software and hardware parameters [5]. Y.Koh et al. [3] have studied the effects of performance interference between two virtual machines hosted on the same hardware platform by considering system-level workload characteristics. It is observed from their work that performance of the applications could be predicted and thus can further identify the existence of performance interference based on different kinds of workload.

3) *Prediction Methods*: In a virtualized environment, identifying methods that adequately model the behavior of virtualized applications is a challenge. Some literatures have addressed the Regression, Artificial Neural Networks, Machine Learning Techniques and Queuing models to predict the performance of applications running in VMs.

a) *Regression Model* : Regression analysis is widely used statistical tool to model relationship between variables. It estimates the mathematical relationship between a dependent variable and an independent variable. A simple form of regression analysis is linear regression, in which dependent variable is assumed as a linear function of explanatory variables.

*Linear Regression Model*: Performance modeling and evaluation techniques are essential to avoid performance related issues and guarantees quality service. G. Casale et al. in [16] proposed simple linear prediction model to predict throughput, response time and mix of read/write requests due to the impact of consolidated storage workloads in a virtualized environment. Y.Koh et al. presented a technique in [3], for predicting performance degradation of co-located applications based on their resource usage statistics, similar to program similarity analysis by Hoste et.al.in [11]. It is observed by A. Tikotekar et al. in [2] that prediction error for most of the applications are significantly high. Q. Noorshams et al. in [15] proposed a general heuristic search algorithm and created optimized performance models using different regression methods. The accuracy of regression models are measured with optimized regression and standard parameters. The authors in their work conclude that there is a considerable improvement in model accuracy of regression optimization algorithm as compared to regression techniques.

*Non-Linear Regression Model*: S. Kundu et al. addressed in [4] that conventional mathematical model which can predict a virtualized application's performance with acceptable accuracy is extremely difficult. So there is a need for a technique which is adaptive and efficient in modeling complex non-linear relationships between inputs and outputs. Evolutionary approaches such as fuzzy logic, artificial neural networks and genetic algorithms provide such ability. According to Y. Koh et.al in [3], high prediction error in linear regression analysis leads to the adoption of non-linear regression analysis.

Xiaoli Sun et al in [14] applied multivariable exponent interference (MCEI) model to identify interfering factors and explored the interference by understanding the relationships between them. The authors in their work have used least squares method in MCEI to predict the performance and have proved that prediction accuracy of MCEI model is more precise than linear and quadratic model.

b) *Artificial Neural Network*: Artificial Neural Networks (ANNs) are preferred to predict in order to provide the resources proactively to the application on demand. It also used in choosing the suitable configuration with least overhead.

S. Kundu et al. in [4] have identified the set of parameters that can directly impact application performance including CPU, memory, disk and network I/O bandwidths. The authors in their work propose an iterative model training technique based on ANN for dynamic resource allocation in consolidated virtualized environments. This model is intended to reflect the influencing factors that can be used to predict the performance overhead for services deployed in, e.g., Cloud Computing or virtualized environments in general.

The authors in [8] consider a wide range of parameters that affect the application performance when running on consolidated virtualized infrastructures and quantify the overhead. Therefore, Infrastructure as a Service (IaaS) providers may perform intelligent decisions regarding the placement of VMs on physical nodes. This will mitigate the aforementioned problem and will aid the IaaS layer to minimize the need for over-provisioning.

c) *Queuing Model* : Time domain queuing model is presented by A.Chandra et al. in [10], for shared server environment. This model allocates the resources for applications based on workload characteristics. It offers dynamic resource allocation through online monitoring and prediction techniques which estimates the workload parameter of applications.

d) *Machine Learning Technique* : Y. Amannejad et al. in [13] have used machine learning technique to detect cloud platform induced performance interference. The authors in their work discuss the performance interference issues faced by the cloud users. In particular, for cloud users handling web services, there is long response time due to performance interference. Hence the authors have applied collaborative filtering, a machine learning technique on transaction response time data collected by a web service.

## B. Categorization of Performance Interference :

In virtualized centers, Virtual Machine Monitor (VMM) or hypervisor manages multiple VMs sharing the storage device. While the hypervisor slices resources and allocate shares to different VMs, the behavior of one VM can still affect the performance of another adversely due to the shared use of resources in the system and thus leads to significant performance degradation.

The performance interferences are categorized as cache, I/O interference. I/O interference is further divided into Disk and Network I/O interferences. According to Q. Zhu et al. in [20], a lot of research efforts focus on the impact of performance interference among consolidated VMs with respect to CPU, Cache and I/O.

1) *Cache Interference*: Nathuji et al. presented an approach in [6], that uses an online MIMO (multi-input multi-output) model to capture the performance interference effects in terms of resource allocations. The method adjusts the processor allocation for each application based on the required performance level, in effect compensating for performance degradation by allocating additional resources.

2) *I/O Interference* : Read and write requests may affect system performance in different ways. Modeling the performance disk requests is very challenging due to the joint interaction of I/O flows issued by several VMs. This complexity has been tackled by introducing simple model for I/O performance prediction in consolidated environments where multiple VMs can share access to a remote storage server.

a) *Disk I/O Interference* : The performance of I/O-bound applications is controlled based on a few factors such as time taken by the operating system to schedule read, write operations and response times of storage devices in completing such requests. Latency of disk I/O requests is affected by the changes in workload as well as in the software and hardware environments. It is often useful to define performance models to anticipate the effects of a change. This is especially important in Cloud environments that employ virtualization technologies.

A large amount of I/O research literature proposes scheduling algorithms for disk I/O in virtualized environments. The main challenges in scheduling are to provide fair access to the shared storage resource for all VM applications, while maintaining performance isolation, i.e. disk accesses by one application should not affect the I/O performance of another [7]. More recently, A. Gulati et al. in [9] characterizes disk workloads and performance metrics in a consolidated virtualized environment.

b) *Network I/O Interference*: According to Xing et al. in [12], their work gives a dedicated performance interference study on network I/O workloads running in separate VMs under shared hardware platform. Experimentation on the performance interference among VMs is implemented in Xen VMs. The authors have analyzed interference factors and presented their findings from the experiments based on network I/O workload characteristics.

### C. Scheduling Algorithms

Saurabh Bilgaiyan et al. in [26] have highlighted the need of proper scheduling in cloud computing environment, as the process of scheduling is significant to manage resources, to reduce the idle time and to improve the performance of systems. In their work, the authors have analyzed the different evolutionary and swarm based algorithms for task scheduling by discussing their advantages and the mechanism used in those algorithms.

Shengjun Xue et. al [27] proposed improved ant colony optimization scheduling algorithm to allocate the resources for tasks and improve the utilization rate. This algorithm balances the allocation of tasks to virtual machines based on its capability, in order to reduce the execution time and avoid the resource wastage.

CloudScope [21], a lightweight prediction model applies discrete-time Markov chain to predict performance degradation based on system behavior and different workloads. The performance degradation is measured using virtualization slowdown (V-slowdown) factor. This model uses interference aware VM scheduler which is compared with default scheduler to analyze the performance of CPU, disk and network intensive applications.

Q-Clouds proposed by R.Nathuji et al. in [6], a model that allocates the resources to identify the interference between CPU applications. It uses multi-input multi-output (MIMO) model and applies feedback control methods to manage the resources for mitigating the interference effects. Similar to Q-Clouds, Q. Zhu et al. in [20] proposed consolidation algorithm which offers application QoS as guaranteed. This algorithm optimizes the number of successfully executed application as provider's metric. It aims at predicting the performance degradation based on time variance in application's resource usage.

Cucinotta, T. et al. in [22] have showed that it is possible to achieve isolation between the co-located VMs. The authors in their work provide a soft real-time scheduling algorithm to obtain good throughput of the VMs by minimizing interference effects. Altino Sampaio et al. in [23] have proposed new algorithms to minimize the performance interference from hardware resources like last-level cache (LLC) sharing by scheduling VMs dynamically. These algorithms aim at increasing the rate of completed tasks. C. Delimitrou et al. in [24] presents Paragon, a system that uses previously scheduled applications to allocate new application based on profiling. It uses a benchmark to detect interferences and impact of interferences on co-located applications.

TRACON, is a framework is proposed by Chiang et al.in [29] with the aim of mitigating the interference effects from co-located data-intensive applications, and can thus improve the overall system performance. Similar to this approach, fast genetic k-means++ and PSO based k-means++ algorithms proposed in [28], schedule and allocate the tasks to various VMs with minimized interference occurring from co-located applications.

## IV. CONCLUSION AND FUTURE WORK

Research on virtualization is highly complex by nature. Further, it becomes more challenging when it involves effective management of cloud environment for service providers as well as customers. Keeping this view in mind, the survey presented in this paper has attempted to consolidate the different performance interference and prediction methods. It is believed that this survey would definitely benefit developers to have complete knowledge about the impact of performance interferences on applications. The future work will be on different scheduling algorithms to mitigate interference and improve system performance.

## REFERENCES

- [1] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. 2010. A view of cloud computing. *Commun. ACM* 53, 4 (April 2010), 50-58. DOI=10.1145/1721654.1721672 <http://doi.acm.org/10.1145/1721654.1721672>
- [2] A. Tikotekar, G. Vall'ee, T. Naughton, H. Ong, C. Engelmann, S. L. Scott, and A. M. Filippi, "Effects of virtualization on a scientific application running a hyperspectral radiative transfer code on virtual machines," in *Proc. of HPCVirt'08. USA: ACM, 2008*, pp. 16-23.
- [3] Y. Koh, R. Knauerhase, P. Brett, M. Bowman, Z. Wen, and C. Pu. An analysis of performance interference effects in virtual environments. In *Proceedings of the IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2007*

- [4] S. Kundu, R. Rangaswami, K. Dutta, and M. Zhao, "Application performance modeling in a virtualized environment," in HPCA, 2010, pp. 1–10
- [5] Benevenuto, F., Fernandes, C., Santos, M., Almeida, V., Almeida, J., Janakiraman, G., Santos, J.: Performance models for virtualized applications. In: ISPA, vol. 4331. LNCS, pp. 427–439. Springer, Berlin (2006)
- [6] R. Nathuji, A. Kansal, and A. Ghaffarkhah. Q-clouds: managing performance interference effects for qos-aware clouds. In Proceedings of the 4th ACM European conference on Computer systems (EUROSYS), 2010.
- [7] W. Jin, J. S. Chase, and J. Kaur. "Interposed proportional sharing for a storage service utility," ACM PEVA, vol. 32, no. 1, pp. 7–48, 2004.
- [8] George Kousiouris, Tommaso Cucinotta, Theodora Varvangou, The Effects of Scheduling, Workload Type and Consolidation Scenarios on Virtual Machine Performance and their Prediction through Optimized Artificial Neural Networks, Journal of systems and software, volume 84, issue 8, August 2011, pages 1270-1291
- [9] A. Gulati, C. Kumar, and I. Ahmad, "Storage workload characterization and consolidation in virtualized environments," in VPACT, 2009.
- [10] A. Chandra, W. Gong, and P. Shenoy, "Dynamic resource allocation for shared data centers using online measurements," in IWQoS. Springer, 2003, pp. 381–398
- [11] K. Hoste, A. Phansalkar, L. Eeckhout, A. Georges, L. K. John, and K. D. Bosschere. Performance prediction based on inherent program similarity. In Proceedings of the conference on Parallel architectures and compilation techniques (PACT), 2006.
- [12] Xing Pu, Ling Liu, Yiduo Mei, Sankaran Sivathanu, Younggyun Koh, Calton Pu, Understanding Performance Interference of I/O Workload in Virtualized Cloud Environments., IEEE 3rd International Conference on Cloud Computing . 2010.
- [13] Y .Amannejad, D. Krishnamurthy and B. Far, "Detecting performance interference in cloud-based web services", Proc. IFIP/IEEE Int. Symp. IM , pp.423-431, 2015
- [14] Xiaoli Sun; Qingbo Wu; Yusong Tan; Fuhui Wu, "MVEI: An Interference Prediction Model for CPU-intensive Application in Cloud Environment", Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2014 13th International Symposium on, On page(s) 83 - 87
- [15] Q. Noorshams, D. Bruhn, S. Kounev, and R. Reussner, "Predictive Performance Modeling of Virtualized Storage Systems using Optimized Statistical Regression Techniques," Proc. ACM/SPEC ICPE, 2013
- [16] G. Casale, S. Kraft, and D. Krishnamurthy, "A model of storage i/o performance interference in virtualized systems," in Proceedings of the 31st International Conference on Distributed Computing Systems Workshops(ICDCSW'11), Minneapolis, Minnesota, Jun. 2011, pp. 34–39.
- [17] Koziolok, H. (2010). Performance Evaluation of Component-based Software Systems: A Survey. Performance Evaluation, 67(8):634–658
- [18] Hauck M, Happe J, Reussner R (2011) Towards performance prediction for cloud computing environments based on goal-oriented measurements. In: Proceedings of the 1st international conference on cloud computing and services science, pp 616–622
- [19] Huber N., von Quast, M., Hauck, M., Kounev, S.: Evaluating and modeling virtualization performance overhead for cloud environments. In: CLOSER 2011 (2011)
- [20] Q. Zhu and T. Tung, "A performance interference model for managing consolidated workloads in qos-aware clouds," in Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on. IEEE, 2012, pp. 170–179
- [21] X. Chen, L. Rupprecht, R. Osman, W. Knottenbelt and F. Franciosi (Oct 2015). CloudScope: Diagnosing Performance Interference for Resource Management in Multi-Tenant Clouds. 23rd IEEE MASCOTS. Oct 5-7, 2015, Atlanta, Georgia, USA
- [22] Cucinotta T., Giani, D., Faggioli, D., Checconi, F., 2010b. Providing performance guarantees to virtual machines using real-time scheduling. In: Proceedings of the 5th Workshop on Virtualization and High-Performance Cloud Computing (VHPC 2010), Ischia (Naples), Italy
- [23] Altino Sampaio, Jorge G. Barbosa, Parallel & Cloud computing, PCC Vol. 2 Iss. 4, 2013PP. 116-125 www.vkingpub.com © 2013 American V-King Scientific Publishing
- [24] Christina Delimitrou and Christos Kozyrakis Paragon: QoS-aware scheduling for heterogeneous datacenters. In ASPLOS, 2013.
- [25] Q. Huang, P. P. Lee, "An experimental study of cascading performance interference in a virtualized environment," ACM SIGMETRICS Performance Evaluation Review, vol. 40, no. 4, 2013, pp. 43-52.
- [26] Saurabh Bilgaiyan, Santwana Sagnika, Madhabananda Das, "An Analysis of Task Scheduling in Cloud Computing using Evolutionary and Swarm-based Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 89 – No.2, March 2014.
- [27] Shengjun Xue, Mengying Li, Xiaolong Xu, and Jingyi Chen, "An ACO-LB Algorithm for Task Scheduling in the Cloud Environment", JOURNAL OF SOFTWARE, VOL. 9, NO. 2, FEBRUARY 2014
- [28] Nirmala, A. P., and R. Sridaran. "Towards Performance Improvement of Cloud Applications under Virtualized Environment using PSO based K-means++ Algorithm.", International Journal of Engineering and Technology (IJET) ISSN : 0975-4024 Vol 7 No 5 Oct-Nov 2015, (1557-1563)
- [29] Ron C. Chiang and H. Howie Huang, "TRACON: Interference-Aware Scheduling for Data-Intensive Applications in Virtualized Environments", IEEE Transactions On Parallel And Distributed Systems, IEEE-2013.