

# Guest Editors' Introduction: Special Issue on Cloud Computing

Vojislav B. Misić, *Senior Member, IEEE*, Rajkumar Buyya, Dejan Milojicic, *Fellow, IEEE*, and Yong Cui

CLOUD is a novel computing paradigm that builds on the foundations of distributed computing, grid computing, networking, virtualization, service orientation, and market-oriented computing. Cloud provides flexible resource allocation on demand with the promise of realizing elastic, Internet-accessible computing on a pay-as-you-go basis. Cloud services include infrastructure such as computing and storage servers, platforms such as operating systems, and application software; these are often referred to as infrastructure-as-a-service or IaaS, platform-as-a-service or PaaS, and software-as-a-service or SaaS, respectively. Cloud offers or, rather, promises numerous benefits from both the technology perspective such as increased availability, flexibility, and functionality, and the business perspective such as reduced capital and operational expenditure and shorter turnaround times for new services and applications.

But, before this paradigm can be widely accepted, many issues have yet to be resolved, including flexible architectural solutions, efficient resource virtualization techniques (CPU, storage, and link virtualization), performance modeling and optimization, modeling and development techniques for Cloud-based systems and applications, reliability modeling, techniques and policies for ensuring security and privacy, and many others. Additional areas that need attention due to recent developments in the area of smart devices (phones, tablets, and the like) include mobile and roaming access, mobile and dynamic applications, incorporation of modern wireless and cellular technologies into the Cloud paradigm, and the development of new virtualization, scheduling, and transport schemes to achieve energy saving and enable green computing to become an intrinsic part of the Cloud.

To help gain additional and much needed insights into those issues, the *IEEE Transactions on Parallel and Distributed Systems (TPDS)* has created a special issues devoted to

recent advances in Cloud computing from the parallel and distributed systems perspective. The Call for Papers was issued in late 2011, with the submission deadline set to 1 March 2012. The response was overwhelming, which caused some delays in the decision making process, but ultimately we have been able to accept the 19 best papers out of more than 100 high quality submissions.

The papers can be broadly classified according to their primary focus. Papers in the first group deal with a broad spectrum of cloud implementation issues, mostly those related to resource management and performance of cloud-based systems. One subgroup of papers deals with general resource management issues.

The paper “*Anchor: A Versatile and Efficient Framework for Resource Management in the Cloud*” by Hong Xu and Baochun Li presents a resource management architecture that allows clients and operators to express and enforce a variety of resource management policies which are then used in the process that matches virtual machines (VMs) with specific resource requirements to servers with appropriate resources in an efficient and cost-effective manner.

Another resource management framework which relies on request partitioning based on Iterated Local Search in the networked environment is presented in the paper “*Efficient Resource Mapping Framework over Networked Clouds via Iterated Local Search-Based Request Partitioning*” by Aris Leivadreas, Chrysa Papagianni, and Symeon Papavassiliou. A distributed intracloud resource mapping is then used to allocate virtual to physical resources in an efficient and balanced manner. The performance of the proposed approach is found to compare favorably against an exact request partitioning solution as well as another common intradomain virtual resource allocation approach.

The paper “*Optimal Multiserver Configuration for Profit Maximization in Cloud Computing*” by Junwei Cao, Kai Hwang, Keqin Li, and Albert Y. Zomaya examines the problem of optimal multiserver configuration in a cloud computing environment with the goal of maximizing the profit of cloud providers. The multiserver system is treated with a queuing model upon which the optimization problem is based and subsequently solved. The impact of various system parameters such as service quality, service level agreement, application workload, and costs of renting, energy, and other resources, upon the optimal solution is analyzed and discussed.

Resource allocation is also discussed in the paper “*Error-Tolerant Resource Allocation and Payment Minimization for Cloud System*” by Sheng Di and Cho-Li Wang, using a deadline-driven problem formulation which is solved in

- V.B. Misić is with the Department of Computer Science, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada. E-mail: vmisic@ryerson.ca.
- R. Buyya is with the Cloud Computing and Distributed Systems (CLOUDS) Laboratory, Department of Computing and Information Systems, The University of Melbourne, Doug McDonnell Building, Parkville Campus, Melbourne, VIC 3010, Australia. E-mail: rbuyya@unimelb.edu.au.
- D. Milojicic is with HP Labs, 1501 Page Mill Road, MS 1183, Palo Alto, CA 94304. E-mail: dejan.milojicic@hp.com.
- Y. Cui is with the Computer Science Department, Tsinghua University, Room 4-104, FIT Building, Beijing, 100084, China. E-mail: cuiyong@tsinghua.edu.cn.

For information on obtaining reprints of this article, please send e-mail to: [tpds@computer.org](mailto:tpds@computer.org).

polynomial time using a novel solution approach. The solution is then augmented with an error-tolerant approach that ensures tasks will be completed by the specified deadline even under inaccurate workload prediction, which is validated through extensive experiments.

Virtualization is a technique commonly used in cloud data centers to leverage the power of modern server architectures; however, it does pose many challenges of its own. The papers in the next subgroup highlight some of these challenges and offer efficient solutions to them.

Dynamic resource allocation based on virtualization technology is analyzed in the paper "Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment" by Zhen Xiao, Weijia Song, and Qi Chen, who measure the unevenness of resource utilization through a proxy metric of "skewness." By minimizing skewness, different workloads can be accommodated with ease, thus leading to improvements in server resource utilization. A set of heuristics is also developed to prevent system overload while improving energy efficiency.

A different look at virtualization is presented in the paper "Performance Enhancement for Network I/O Virtualization with Efficient Interrupt Coalescing and Virtual Receive-Side Scaling" by HaiBing Guan, YaoZu Dong, RuHui Ma, Dongxiao Xu, Yang Zhang, and Jian Li, where a comprehensive optimization-based approach is described to address both the challenges of network I/O virtualization and efficient use of multicore processors. Experiments on a Xen virtualization platform confirm that performance challenges can be successfully solved for substantial performance improvements.

The paper "A New Disk I/O Model of Virtualized Cloud Environment" by Dingding Li, Xiaofei Liao, Hai Jin, Bingbing Zhou, and Qi Zhang focuses on problems posed by virtualization from the standpoint of disk I/O performance and proposes a novel I/O model in which the guest file system uses synchronous I/O operations while the host file system uses asynchronous ones. Experimental results with the prototype system in Xen hypervisor environment demonstrate the advantages of this approach over the conventional solutions.

The paper "Improving Data Center Network Utilization Using Near-Optimal Traffic Engineering" by Fung Po Tso and Dimitrios P. Pazaros focuses on multipath routing in data center networks, with the goal of constructing a routing algorithm that simultaneously provides simple path discovery, path diversity even in the presence of nonuniform cost links, and minimal link utilization. One such routing algorithm is described in the paper and it is shown that the use of this algorithm, together with a simple modification of the current canonical tree data center architecture, can significantly reduce maximum link utilization and increase the efficiency of the network.

Energy efficiency is also an important issue in cloud data centers, as exemplified by the techniques proposed in the following two papers.

The paper "Electricity Cost Saving Strategy in Data Centers by Using Energy Storage" by Yuanxiong Guo and Yuguang Fang addresses the possibility of using energy storage in data centers with the goal of reducing the cost of

electricity usage under a wide range of fluctuations of electricity price and data center workloads. The online algorithm presented in the paper achieves an explicit tradeoff between energy storage capacity and cost savings, and allows for effective energy management and, consequently, reductions in operating cost of cloud data centers.

The energy efficiency of cloud data centers can be improved by powering down unnecessary servers during time periods with reduced workload; however, the savings depend on the predictability of the future workload. The paper "Simple and Effective Dynamic Provisioning for Power-Proportional Data Centers" by Tan Lu, Minghua Chen, and Lachlan L.H. Andrew discusses novel decentralized dynamic provisioning algorithms that allow cloud operators to reduce power consumption in this way, and shows that effective solutions can be obtained without excessive knowledge of future workload requirements.

The second group of papers deals with various issues related to cloud applications, first and foremost security and privacy.

As cloud resources may be available to a large number of users, protecting the confidentiality of customer data is among the top priority requirements for cloud systems. The paper "Harnessing the Cloud for Securely Outsourcing Large-Scale Systems of Linear Equations" by Cong Wang, Kui Ren, Jia Wang, and Qian Wang investigates a secure outsourcing mechanism that allows customers to securely harness the cloud while keeping the sensitive input and output data private. Furthermore, an efficient result verification mechanism is proposed to protect data from tampering. While the mechanism is demonstrated in the context of solving large systems of linear equations, it can be easily modified to other similar applications that use the cloud.

On the other hand, there may be situations in which multiple users need to share data through the cloud—but in a controlled manner. The problems arising from such a setup are analyzed in the paper "Mona: Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud" by Xuefeng Liu, Yuqing Zhang, Boyang Wang, and Jingbo Yan, where group signatures and dynamic broadcast encryption techniques are leveraged to create a secure multi-owner data sharing scheme dubbed "Mona." Mona allows anonymous data sharing with very little overhead, in terms of both storage and computation, even with frequent changes of membership in the owner set.

The paper "A Privacy Leakage Upper Bound Constraint-Based Approach for Cost-Effective Privacy Preserving of Intermediate Data Sets in Cloud" by Xuyun Zhang, Chang Liu, Surya Nepal, Suraj Pandey, and Jinjun Chen deals with the problem of preserving the privacy of intermediate datasets—data produced during computation and saved in order to avoid recomputation and improve efficiency. The approach proposed in the paper avoids encryption of all such datasets by exploiting privacy leakage constraints to determine which data sets should be encrypted and which can be left in the clear. The proposed approach is shown to reduce computational cost while still satisfying the privacy requirements of data holders.

The paper "A Truthful Dynamic Workflow Scheduling Mechanism for Commercial Multicloud Environments" by Hamid Mohammadi Fard, Radu Prodan, and Thomas Fahringer presents a pricing model and a scheduling mechanism that compensate for selfish behavior of individual users and thus allow for optimal behavior with respect to the efficiency of the entire system to be achieved.

Issues related to costing are the topic of the following papers.

The paper "QoS Ranking Prediction for Cloud Services" by Zibin Zheng, Xinmiao Wu, Yilei Zhang, Michael R. Lyu, and Jianmin Wang discusses QoS ranking of cloud services and proposes a prediction framework that takes advantage of past service usage experiences. On the basis of this framework, personalized QoS ranking prediction approaches are proposed and analyzed; experiments using real-world data show that the performance of the framework exceeds that of other competing approaches.

The paper "Cloudy with a Chance of Cost Savings" by Byung Chul Tak, Bhuvan Uргаonkar, and Anand Sivasubramaniam discusses a range of hosting options for cloud-based systems and their impact on economic indicators, most notably the cost of deployment. The analysis considers a range of important application characteristics, as well as workload variance and cloud elasticity, using familiar applications from the TPC (Transaction Processing Council) benchmark suite.

The paper "A Highly Practical Approach toward Achieving Minimum Data Sets Storage Cost in the Cloud" by Dong Yuan, Yun Yang, Xiao Liu, Wenhao Li, Lizhen Cui, Meng Xu, and Jinjun Chen focuses on storage strategies for large application datasets to be stored in the cloud and their associated cost. Various alternatives are analyzed and the tradeoffs thereby incurred are discussed, and it is shown that the cost effectiveness can be optimized in a wide range of parameter values.

Finally, the two papers described below deal with diagnostics and benchmarking.

Diagnostics of cloud systems requires efficient diagnostics tools capable of unsupervised operation at a desired level of granularity. The paper "Toward Fine-Grained, Unsupervised, Scalable Performance Diagnosis for Production Cloud Computing Systems" by Haibo Mi, Huaimin Wang, Yangfan Zhou, Michael Rung-Tsong Lyu, and Hua Cai discusses these requirements and proposes the architecture and implementation details for the tool dubbed "CloudDiag" that relies on advanced statistical techniques and a fast matrix recovery algorithm to effectively pinpoint the causes of performance problems without requiring extensive domain-specific knowledge of the host system.

The paper "C-MART: Benchmarking the Cloud" by Andrew Turner, Andrew Fox, John Payne, and Hyong S. Kim discusses the problems arising from the need to benchmark cloud systems in order to ensure maximum resource utilization and thus minimize the cost for the cloud providers and operators. They present "C-MART," a benchmark designed to emulate the characteristics of modern web applications executing in a cloud environment. Through flexible design and comprehensive architecture, C-MART can vastly increase the accuracy of

detecting various resource usage problems, and thus improve the performance of cloud systems using it.

In conclusion, we may safely say that the papers presented in this special issue demonstrate the breadth and diversity of research in the field of cloud computing. We wish to thank both the authors and the reviewers for their hard work and the effort they have invested in helping us assemble this special issue. We would also like to express our sincere gratitude to the Editor-in-Chief, Professor Ivan Stojmenovic, for extending this opportunity and for his tireless guidance throughout the process, and the editorial staff of *TPDS* for their continuous support and professionalism.

Vojislav B. Misić  
Rajkumar Buyya  
Dejan Milošević  
Yong Cui  
*Guest Editors*



**Vojislav B. Misic** received the PhD degree in computer science from the University of Belgrade, Serbia, in 1993. He is a professor of computer science at Ryerson University in Toronto, Ontario, Canada. His research interests include software engineering, cloud computing, and performance evaluation of wireless networks and systems. He has authored or coauthored six books, 18 book chapters, and more than 200 papers in archival journals and at

prestigious international conferences. He serves on the editorial boards of the *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Cloud Computing*, *Ad Hoc Networks*, *Peer-to-Peer Networks and Applications*, and the *International Journal of Parallel, Emergent, and Distributed Systems*. He is a senior member of the IEEE and a member of the ACM and AIS.



**Rajkumar Buyya** is a professor of computer science and software engineering, future fellow of the Australian Research Council, and director of the Cloud Computing and Distributed Systems (CLOUDS) Laboratory at the University of Melbourne, Australia. He is also serving as the founding CEO of Manjrasoft, a spin-off company of the university, commercializing its innovations in Cloud Computing. He has authored more than 425 publications and four text books including

*Mastering Cloud Computing* (McGraw Hill and Elsevier/Morgan Kaufmann, 2013) for the Indian and international markets, respectively. He also edited several books, including *Cloud Computing: Principles and Paradigms* (Wiley Press, February 2011). He is one of the highly cited authors in computer science and software engineering worldwide. Microsoft Academic Search Index ranked Dr. Buyya as the world's top author in distributed and parallel computing between 2007 and 2012. Recently, ISI has identified him as a "Highly Cited Researcher" based on citations to his journal papers. Software technologies for Grid and Cloud computing developed under Dr. Buyya's leadership have gained rapid acceptance and are in use at several academic institutions and commercial enterprises in 40 countries around the world. Dr. Buyya has led the establishment and development of key community activities, including serving as founding chair of the IEEE Technical Committee on Scalable Computing and five IEEE/ACM conferences. These contributions and international research leadership of Dr. Buyya are recognized through the award of the 2009 IEEE Medal for Excellence in Scalable Computing from the IEEE Computer Society. Manjrasoft's Aneka Cloud technology developed under his leadership has received the 2010 Asia Pacific Frost & Sullivan New Product Innovation Award and the 2011 Telstra Innovation Challenge, People's Choice Award. He is currently serving as the founding Editor-in-Chief of *IEEE Transactions on Cloud Computing*. For further information on Dr. Buyya, please visit his cyberhome: [www.buyya.com](http://www.buyya.com).



**Dejan Milojevic** received the PhD degree from the University of Kaiserslautern, Germany, in 1993; and the MSc/BSc degrees from Belgrade University, Serbia, in 1983 and 1986, respectively. He has been a senior researcher and research manager at HP Labs, Palo Alto, California since 1998. He also worked at the OSF Research Institute, Cambridge, Massachusetts (1994-1998), and the Institute "Mihajlo Pupin," Belgrade, Serbia (1983-1991). He is

the IEEE Computer Society 2014 President. He was a founding Editor-in-Chief of *IEEE ComputingNow* (2008-2012). He has been on many conference program committees and journal editorial boards. He is an IEEE fellow, an ACM Distinguished Engineer, and a USENIX member. He has published more than 130 papers and two books. He has 11 patents and 25 patent applications.



**Yong Cui** received the BS and PhD degrees, both in computer science, from Tsinghua University in 1999 and 2004, respectively. He is a professor in the Computer Science Department of Tsinghua University, cochair of IETF IPv6 Transition WG Software, and a council member of the China Communication Standards Association. Having published more than 100 papers in refereed journals and conferences, he received Best Paper Awards from ACM ICUIMC 2011 and WASA 2010.

Holding more than 40 patents, he won the National Science and Technology Progress Award of the China and the Influential Invention Award of China Information Industry. He is one of the authors of IETF RFC 5747 and RFC 5565 for his proposal on IPv6 transition technologies. His major research interests include mobile wireless Internet and computer network architecture.