Guest Editorial: Networking Challenges in Cloud Computing Systems and Applications

David S. L. Wei, Sarit Mukherjee, Kshirasagar Naik, Amiya Nayak, Yu-Chee Tseng, and Li-Chun Wang

CLOUD computing is an innovative Internet-based computing paradigm that offers users a scalable, elastic, and cost-effective computing environment. On the other hand, this promising computing paradigm also poses various challenges to cloud service providers, and the associated research issues have attracted numerous researchers. This research work can be classified into research on computational aspects of cloud computing and research on networking aspects of cloud computing. While research on computational aspects of cloud computing is actively being conducted, studies on networking challenges in cloud computing have been lagging behind. In the following we will explain the urgency for addressing the research on networking aspects. The goal of this special issue is to feature recent research in this area.

Network performance is the key to cloud computing performance as cloud networks are the infrastructure for cloud services. A cloud network connects servers and storage in multiple locations to create a pool of resources, and contains a set of routers and switches that transport traffic between the servers and to the outside world. Due to the complexity of interconnection networks, large number of users, user's mobility, and a large variety of application services, cloud networks pose several challenges. The key challenges include the data casting and routing problem, cost-effective data center scalability, unpredictable traffic patterns and variable demand, dynamic network resource allocation, workload and IP mobility, to name a few. These challenges can be addressed by, for example, traffic control, novel interconnection networks with scale-out property, protocols for supporting live migration of VMs, as well as using emerging technologies such as network virtualization and novel network addressing schemes to tackle service and VM migration. The papers that appear in this special issue address these important issues in cloud computing systems and applications.

Though research on cloud networking is still in its nascent stage, this special issue still attracted 35 submissions of high-quality works from around the world. Due to limited slots, only the following 11 papers were selected for publication after a rigorous review process. These papers present results of analysis, experimentation, simulation, and system implementation. They cover the topics of Network Virtualization and VM Placement, Data Casting and Routing, Traffic Management, Resource Management, Distributed Storage System, Big Data, and Cloud Applications. We summarize these papers in the order of the listed topics in what follows.

D. S. L. Wei is with the Dept. of Computer and Information Sciences, Fordham University, Bronx, NY 10458, USA (e-mail: dsl.wei01@gmail.com).

S. Mukherjee is with the Center for Networking Research, Alcatel-Lucent Bell Labs, Holmdel, NJ 07733, USA.

K. Naik is with the Dept. of Electrical and Computer Engineering, University of Waterloo, Ontario N2L 3G1, Canada.

A. Nayak, SITE, University of Ottawa, Ontario K1N 6N5, Canada.

Y.-C. Tseng is with the Dept. of Computer Science, National Chiao-Tung University, Hsinchu 300, Taiwan, R.O.C.

L .-C. Wang is with the Dept. of Electrical Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan, R.O.C.

A. Leon-Garcia is the J-SAC board representative for this issue of IEEE Journal on Selected Areas in Communications.

Digital Object Identifier 10.1109/JSAC.2013.131201

I. Network Virtualization and VM Placement

The first three papers are about network virtualization or VM placement. In the paper "SR-IOV Based Network Interrupt-Free Virtualization with Event Based Polling," Guan, Dong, Tian, and Li propose a network interrupt-free virtualization mechanism to improve the network throughput of a datacenter. Their approach eliminates interrupts in the critical I/O handling path, and instead replaces them by a smart event-based polling model. Their proposed mechanism can be implemented either at the guest OS kernel level or at the Virtual Machine Monitor level. In the paper "Enhancing Survivability in Virtualized Data Centers: A Service-aware Approach," Xu, Tang, Kwiat, Zhang, and Xue propose to use a service-aware approach to enhance survivability in virtualized data centers. They consider the optimization problem of Survivable Virtual Infrastructure Mapping and present a general optimization framework. They then propose an efficient algorithm for VM Placement (VMP), a polynomial-time optimal algorithm for Virtual Link Mapping (VLM), and an effective heuristic algorithm that jointly solve VMP and VLM problem. In the paper "Automating Cloud Network Optimization and Evolution," Wu, Zhang, Singh, Jiang, and Wang present a cloud network performance optimization framework, a topology independent resource allocation and optimization approach, to achieve continuous and cost effective data center maintenance. Based on a swarm intelligence optimization model, their approach improves the scalability of the cloud network by relocating VMs and matching resource demand and availability.

II. Data Casting and Routing

The next two papers are about data casting and routing. In the paper "Datacast: A Scalable and Efficient Reliable Group Data Delivery Service for Data Centers," Cao, Guo, Lu, Xiong, Zheng, Zhang, Zhu, Chen, and Tian propose a data casting protocol, named Datacast, for Reliable Group Data Delivery (RGDD), a pervasive traffic pattern in data centers. Datacast explores two design spaces: 1) Datacast uses

multiple edge-disjoint Steiner trees for data delivery acceleration, and 2) Datacast leverages in-network packet caching and introduces a simple soft-state based congestion control algorithm to address the scalability and efficiency issues of RGDD. In the paper "A Unified Unicast and Multicast Routing and Forwarding Algorithm for Software-Defined Datacenter Networks," Jia and Wang propose a multiple membership query algorithm based on the prime theory, such as Chinese Remainder Theorem, for the scalability problem associated with software-defined datacenter. Their approach outperforms the Bloom filter approach in terms of memory consumption, hardware cost and delivery accuracy in a large-scale datacenter networks.

III. Traffic Management

The next two papers are about traffic management. In the paper "Data Centers as Software Defined Networks: Traffic Redundancy Elimination with Wireless Cards at Routers," Cui, Xiao, Liao, Stojmenovic, and Li present a mechanism for traffic redundancy elimination in data center networks. Their proposed mechanism adds wireless network cards to both servers and routers to achieve the 'logically centralized' control over the physically distributed states in emerging software defined networks (SDN) paradigm. In the paper "Scalable Multi-Class Traffic Management in Data Center Backbone Networks," Ghosh, Ha, Crabbe, and Rexford investigate two alternative traffic management designs for large online service providers. Their work achieves scalability by distributing computation across multiple tiers of optimization machinery. Using optimization, they show that their designs provably maximize the aggregate utility over all traffic classes.

IV. Resource Management

The next paper addresses the issue of resource management. In the paper "A Framework for Cooperative Resource Management in Mobile Cloud Computing," Kaewpuang, Niyato, Wang, and Hossain propose a decision making framework for mobile cloud service providers in a mobile cloud computing environment. The proposed framework is composed of methods for resource allocation to mobile applications, revenue management, and cooperation formation among the providers. The objective of the framework is to obtain the best decisions for the mobile cloud service providers given that they are rational and interested in maximizing their own benefits.

V. Distributed Storage System

The next paper is about Distributed Storage Systems. In the paper "Capacity and Security of Heterogeneous Distributed Storage Systems," Ernvall, Rouayheb, Hollanti, and Poor study distributed heterogeneous storage systems, whose nodes in these systems can have different storage capacities and different repair bandwidths. They prove an upper bound on the capacity that depends on the average resources available per node, and develop an expression for the system capacity when all the nodes' parameters are known. They also investigate the case in which the system is compromised by an active or passive adversary, and provide bounds on the system secure capacity.

VI. Big Data

The next paper "Moving Big Data to The Cloud: An Online Cost-Minimizing Approach," Zhang, Wu, Li, Guo, Chen, and Lau study timely, cost-minimizing upload of massive, dynamically-generated, geo-dispersed data into the cloud, for processing using a MapReduce-like framework. A polynomial-time optimal offline algorithm is proposed based on dynamic programming, and two online algorithms are designed to practically guide data migration in an online fashion.

VII. Cloud Applications

The next paper is about cloud computing applications. In the paper "Dynamic Request Splitting for Interactive Cloud Applications," Hajjat, Shankaranarayanan, Maltz, Rao, and Sripanidkulchai present a system, named Dealer, that helps geo-distributed, interactive and multi-tier applications meet their stringent requirements on response time despite the high variability in performance of cloud services. Dealer can enable applications to meet their SLA requirements by dynamically splitting requests for each component among its replicas in different data-centers.

In conclusion, we sincerely hope that this special issue provides an up-to-date and valuable research information for the researchers currently conducting research in cloud networking. We would also like to take this opportunity to thank the authors who submitted their quality work to this special issue, and thank all reviewers for their efforts and valuable reviews. We are grateful to Dr. Martha Steenstrup for her advice and encouragement in the initial stage of our proposal. Our special thanks go to Dr. Alberto Leon-Garcia who rendered prompt advice and assistance in the preparation of this special issue. We also thank Laurel Greenidge and Sue Lange for their quick responses to our various queries and the timely processing of the final manuscripts.

David S.L. Wei received his Ph.D. degree in Computer and Information Science from the University of Pennsylvania in 1991. He is currently a Professor of Computer and Information Science Department at Fordham University. From May 1993 to August 1997 he was on the Faculty of Computer Science and Engineering at the University of Aizu, Japan (as an Associate Professor and then a Professor). Dr. Wei has authored and co-authored more than 90 technical papers in the areas of distributed and parallel processing, wireless networks and mobile computing, optical networks, peer-to-peer communications, and cognitive radio networks in various archival journals and conference proceedings. He served on the program committee and was a session chair for several reputed international conferences. He was a lead guest editor of IEEE Journal on Selected Areas in Communications for the special issue on Mobile Computing and Networking, and was a guest editor of IEEE Journal on Selected Areas in Communications for the special issue on Peer-to-Peer Communications and Applications. He was the chair of Intelligent Transportation Forum of Globecom 2010, the general chair of Intelligent Transportation Workshop of ICC 2011, and the chair of Cloud Security Forum and Intelligent Transportation Forum of Globecom 2011. He is currently a lead guest editor of IEEE Journal on Selected Areas in Communications for the special issue on Networking Challenges in Cloud Computing Systems and Applications, a lead guest editor of IEEE Transactions on Cloud Computing for the special issue on Cloud Security, and an Associate Editor of Journal of Circuits, Systems and Computers. Currently, Dr. Wei focuses his research efforts on cloud computing, wireless sensor networks, and cognitive radio networks.

Sarit Mukherjee is a Senior Manager in the Bell Laboratories' Center for Networking Research where he has been leading the research and development of the networked cloud technologies. Before this he held a Technical Manager position with Lucent Technologies managing the research and development of Internet content distribution appliances. Prior to this he managed the design and development of streaming appliances in a start-up company, led the video-networking group in Panasonic Information and Networking Technology Lab at Princeton. He received his M.S. and Ph.D. in Computer Science from the University of Maryland, College Park. He has published more than 50 research papers in several renowned technical journals and conferences, served in the technical committees of number of international conferences, and holds dozens of US patents. His research interests include high-speed networking, cloud computing and multimedia applications.

Kshirasagar (Sagar) Naik received the B. Sc. Engineering degree from Sambalpur University, India and M. Tech. degree from the Indian Institute of Technology, Kharagpur, respectively. He worked as a software developer in Wipro Technologies, Bangalore for three years. Next, he received the M. Math degree in computer science from the University of Waterloo and Ph.D. degree in electrical and computer engineering from Concordia University, Montreal, respectively. He worked as a faculty member at the University of Aizu in Japan and Carleton University in Ottawa. Currently, he is a full professor in the Department of Electrical and Computer Engineering at the University of Waterloo. He was a co-guest editor of three special issues of IEEE Journal on Selected Areas in Communications. He is an associate editor of Journal of Peer-to-Peer Networking and Applications, International Journal of Parallel, Emergent and Distributed Systems, and IEEE Transactions on Parallel and Distributed Systems. In addition, he is serving as the Editor (America) of Journal of Circuits, Systems, and Computers. His research interests include dependable wireless communication, resource allocation in wireless networks, mobile computing, peer-to-peer communication, vehicular networks, energy efficiency of smartphones and tablet computers, energy performance testing of mobile apps, and communication protocols for smart power grids. He has published more than 140 research articles in international journals and conferences. His book entitled Software Testing and Quality Assurance: Theory and Practice (Wiley, 2008) has been adopted as a text in many universities around the world.

Amiya Nayak received his B.Math. degree in Computer Science and Combinatorics & Optimization from University of Waterloo in 1981, and Ph.D. in Systems and Computer Engineering from Carleton University in 1991. He has over 17 years of industrial experience in software engineering, avionics and navigation systems, simulation and system level performance analysis. He is in the Editorial Board of several journals, including IEEE Transactions on Parallel & Distributed Systems, International Journal of Parallel, Emergent and Distributed Systems, International Journal of Computers and Applications, and EURASIP Journal of Wireless Communications and Networking. Currently, he is a Full Professor at the School of Electrical Engineering and Computer Science at the University of Ottawa. His research interests are in the area of mobile computing, wireless sensor networks, and fault tolerance.

Yu-Chee Tseng got his Ph.D. in Computer and Information Science from the Ohio State University in January of 1994. He is/was Professor (2000-present), Chairman (2005-2009), and Associate Dean (2007-present), Department of Computer Science, National Chiao-Tung University, Taiwan, and Chair Professor, Chung Yuan Christian University (2006-2010). Dr. Tseng received Outstanding Research Award (National Science Council, 2001, 2003, and 2009), Best Paper Award (Intl Conf. on Parallel Processing, 2003), Elite I. T. Award (2004), and Distinguished Alumnus Award (Ohio State University, 2005), and Y. Z. Hsu Scientific Paper Award (2009). His research interests include mobile computing, wireless communication, and parallel and distributed computing. Dr. Tseng serves/served on the editorial boards of Telecommunication Systems (2005-present), IEEE Trans. on Vehicular Technology (2005-2009), IEEE Trans. on Mobile Computing (2006-present), and IEEE Trans. on Parallel and Distributed Systems (2008-present).

Li-Chun Wang (M96SM06F11) received the B.S. degree from National Chiao Tung University, Taiwan, R. O. C., in 1986, the M.S. degree from National Taiwan University in 1988, and the Ms.Sci. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, in 1995, and 1996, respectively, all in electrical engineering. From 1990 to 1992, he was with the Telecommunications Laboratories of the Ministry of Transportation and Communications in Taiwan (currently the Telecom Labs of Chunghwa Telecom Co.). In 1995, he was affiliated with Bell Northern Research of Northern Telecom, Inc., Richardson, TX. From 1996 to 2000, he was with ATT Laboratories, where he was a Senior Technical Staff Member in the Wireless Communications Research Department. Since August 2000, he has joined the Department of Electrical and Computer Engineering of National Chiao Tung University in Taiwan and is the current Chairman of the same department. His current research interests are in the areas of radio resource management, crosslayer optimized techniques for heterogeneous wireless networks, and cloud computing for mobile applications. Dr.Wang was elected to the IEEE Fellow grade in 2011 for his contributions in cellular architecture and radio resource management in wireless networks. He won the Distinguished Research Award of the National Science Council, Taiwan, in 2012, and was a co-recipient (with Gordon L. Stuber and Chin-Tau Lea) of the 1997 IEEE Jack Neubauer Best Paper Award for his paper "Architecture Design, Frequency Planning, and Performance Analysis for a Microcell/Macrocell Overlaying System," IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, vol. 46, no. 4, pp. 836848, 1997. He has published over 150 journal and international conference papers. He served as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2001 to 2005, the Guest Editor of the Special Issue on "Mobile Computing and Networking" for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 2005 and on "Radio Resource Management and Protocol Engineering in Future IEEE Broadband Networks" for IEEE Wireless Communications Magazine in 2006. He holds 10 US patents.