



# High-performance computing systems and applications for AI

Gangman Yi<sup>1</sup> · Vincenzo Loia<sup>2</sup>

Published online: 24 June 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

## 1 Introduction

Recently, the integration of High-Performance Computing (HPC) and Artificial Intelligence (AI) has become an exciting innovative paradigm for industry and life. These two emerging technologies are rapidly growing field of research that cover powerful computation required for big data analysis. The combination of AI and HPC has already been used to test various modules of the software without creating the prototype model. HPC-enabled AI could provide optimization of supply chains, complex logistics, manufacturing, simulating and underpin modeling to resolve any problem. For example, the science and academia used HPC-enabled AI to provide data-intensive workloads by data analytic and simulating for a long time. Industries, such as finance, automotive, manufacturing healthcare, require powerful computing and incentive data analyses; therefore, they are turning to HPC and AI application. Contributions to theoretic research offering advanced concepts and technologies, reporting of experiences and tutorials on emerging trends, analyzing experiences related to the implementation as well as application of theories are needed for these fields of research. As such, the goal of this special section is to deliver an overview of the existing technologies and proposed solution guidelines for HPC and AI application.

This special section covers pure applications and research within innovative and emerging scopes related to HPC system and AI application. In addition, it deals with new paradigms, theories, specific mathematical models, hardware/software technologies, efficient data processing.

---

✉ Gangman Yi  
gangman.yi@mme.dongguk.edu

Vincenzo Loia  
loia@unisa.it

<sup>1</sup> Dongguk University-Seoul, Seoul, Korea

<sup>2</sup> Università degli Studi di Salerno, Fisciano, Italy

## 2 High-performance computing systems and applications for AI

In this special section, ‘High-Performance Computing Systems and Applications for AI’ is presented in the twelve accepted articles [1–14]. Each accepted article provides the novel contribution in the field of high-performance computing systems and applications for AI.

Daeyong Jung et al. [1] proposed a manager methodology for synchronizing data and heterogeneous cluster computing that consist of various computing nodes. The memology employed various Windows Subsystem for Linux (WSL) package. The cluster scheduler engine was used for Son of Grid Engine (SGE). Under the certain condition, the manager synchronizes with Linux daemon. The jobs for synchronizing the data executed in SGE are performed by manager. The authors demonstrated the effectiveness of the proposed methodology using the experimental evaluation. The upgradation of the in-engineering SW-based job submission can provide job monitoring environment in the SGE.

Sangdo Lee et al. [2] studied the past cyber-attack incidents and hacking attempts that involved the attacks against nuclear power stations. The authors described two examples of attacks: One is cyber threat against Korea Hydro and Nuclear Power (NHNP) in December 2015, and another is the Stuxnet attack against an Iranian nuclear facility. Based on the study, they proposed a security measure against past attack incidents and mechanism of big data exploded in the attacks. The evaluation of effective measure demonstrates that the security should be strengthened. The security measure can be used to provide security against control systems.

Yin et al. [3] addressed the issue in the finite-state automata-based deep packet inspection systems that need higher operating speed and regular engines for reducing the memory consumption. The authors analyzed the shortcomings of employing finite automata and its significant feature. Based on the analysis, they improved the nondeterministic finite automata (NFA) for reducing the memory utilization by decreasing conversion edge. The experimental results demonstrated that the utilization of memory was reduced more than half over the real-world dataset.

Hwang et al. [4] realized the necessity of artificial intelligence for Internet of things (IoT) in maintaining the big data in various fields of smart transportation, healthcare, metropolitan-scale deployments, infrastructure management, smart manufacturing, home automation and so on. Based on the realization, they proposed a network clock model for time awareness in the integration of IoT and AI. The proposed system was implemented on an actual test platform of MICAz-compatible sensor nodes, and the experimental evaluation results showed that the IoT devices can serve and maintain a standard timestamp for any AI application.

Lee et al. [5] introduced a hybrid mechanism of implementation in parallel on the simultaneous integration of GPU and CPU to support high-speed ECDSA signature verification. The authors exploited various modern computing environments to implement the proposed mechanism for edge computing device and RSU. The experimental evaluation concluded that GPU can provide the significant support to increase the performance of the ECDSA signature verification in RSU platform.

Park et al. [6] studied the effect of the automated enforcement systems (AES) on urban traffic collisions. Based on the study, they categorized the road sections into various clusters by employing spatial collision occurrence patterns and road environment characteristics, respectively. The authors showed that the number of collisions tends to reduce at HCCLs past enforcement sites. The AES installation supports the alleviation in the most cluster of severity of collisions.

Addressing the evolving heterogeneity and distributive features of an IoT environment, Ryu et al.'s research [7] proposed a new digital forensics framework. Based on blockchain technology, the new structure strengthened IoT communications data security by storing them as block transactions. The distributed public ledger of the blockchain makes it easier for forensic investigators to conduct investigation transparently.

Han et al. [8] proposed an efficient job management of computing service (EJM-CS) which assesses idle virtual machine (VM) resources in OpensStack and assigns users computing services. Based on the type of computing work, latency and workload, work priority scheduling allows EJM-CS to perform multiple work processes at the same time.

Yang et al. [9] proposed two methods: dynamic resource allocation method (DRA) and energy saving (ES) for efficient utilization of computing and storage resources in terms of energy consumption in cloud computing. Implementing the DRA and ES algorithms, the power distribution unit monitors and records power consumption. The results show 39.89% of power saving and a release of 20 Gb of memory.

Tian et al.'s research [10] proposed a 3D recognition system to enable swift and accurate recognition of objects in autonomous driving. Multiple valid object features are selected for training and testing the classifier. Various down-sampling ways and 3D convolutional filters are used for weighting and pooling object feature model which serve as input data for the neural network. Parallel computing is deployed to increase time efficiency in real-time object recognition.

Growing crowdsourcing services such as crowd photography require an effective pricing strategy. Hao et al.'s research [11] analyzes the correlation between tasks, their pricing and location of members based on which an improved pricing model is proposed. The new pricing model examines the set of multiple tasks which can be combined in terms of their location pattern.

Kim et al.'s research [12] proposed a mobile cloud infrastructure simulator (MCIS) for offloading high computational tasks, resource management, mobile deployment and information from the mobile device to cloud. MCIS enables visualization of the working state to analyze the performance of the user and identify any problems that may occur during its implementation.

Lee et al.'s research [13] introduces fault-tolerant scheduling algorithms considering check points and replication mechanisms to actively cope with faults. Firstly, job execution model is introduced considering registration, job submission, task allocation, task execution, task result return, and job result return, given that mobile cloud based on a mobile agent is similar to the traditional cloud from the viewpoint of job execution. Then, mobile device groups and scheduling groups are described. Fault-tolerant scheduling algorithms are based on the groups.

Lim et al.'s research [14] describes the differential data processing (DDP) method, which is employed to reduce the size of the transmission data and prove the performance of wireless sensor networks. For the description of the method, node ID-based DDP, merge processing in cluster headers, cluster header-based DDP and difference processing in member nodes are introduced.

### 3 Conclusions

Special thanks go to Journal of Supercomputing's Editor-in-Chief, as well as to all editorial teams for their invaluable support throughout the preparation and publication of this special section. In addition, we thank the external reviewers for their invaluable help in reviewing the papers.

### References

1. Jung D, Lee D, Kim M, Kim J (2018) Efficient data synchronization method on integrated computing environment. *J Supercomput*. <https://doi.org/10.1007/s11227-018-2445-z>
2. Lee S, Huh JH (2018) An effective security measures for nuclear power plant using big data analysis approach. *J Supercomput*. <https://doi.org/10.1007/s11227-018-2440-4>
3. Yin C, Wang H, Yin X, Sun R, Wang J (2018) Improved deep packet inspection in data stream detection. *J Supercomput*. <https://doi.org/10.1007/s11227-018-2685-y>
4. Hwang S (2019) A network clock model for time awareness in the Internet of things and artificial intelligence applications. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02774-0>
5. Lee S, Seo H, Kwon H, Yoon H (2019) Hybrid approach of parallel implementation on CPU–GPU for high-speed ECDSA verification. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02744-6>
6. Park SH, Park SH, Kwon OH, Sung Y (2019) Continuous risk profile and clustering-based method for investigating the effect of the automated enforcement system on urban traffic collisions. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02752-6>
7. Ryu JH, Sharma PK, Jo JH, Park JH (2019) A blockchain-based decentralized efficient investigation framework for IoT digital forensics. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02779-9>
8. Han SH, Kim HW, Jeong YS (2019) An efficient job management of computing service using integrated idle VM resources for high-performance computing based on OpenStack. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02769-x>
9. Yang C, Chen S, Liu J, Chan Y, Chen C, Verma VK (2019) An energy-efficient cloud system with novel dynamic resource allocation methods. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02794-w>
10. Tian Y, Song W, Sun S, Fong S (2019) A 3D object recognition method with multiple features extraction from LiDAR point clouds. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02830-9>
11. Hao F, Guo H, Park D, Kang J (2019) An efficient pricing strategy of sensing tasks for crowdphotographing. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02808-7>
12. Kim H, Kang J, Jeong Y (2019) Simulator considering modeling and performance evaluation for high performance computing of collaborative-based mobile cloud infrastructure. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02882-x>
13. Lee JH, Gil JM (2019) Adaptive fault-tolerant scheduling strategies for mobile cloud computing. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02745-5>
14. Lim KK, Park JS, Shon JG (2019) Differential data processing technique to improve the performance of wireless sensor networks. *J Supercomput*. <https://doi.org/10.1007/s11227-019-02932-4>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.