

## High performance networked computing in media, services and information management

Gang Kou

Published online: 24 April 2013  
© Springer Science+Business Media New York 2013

Nowadays, as the web applications are becoming extremely resource-consuming, the researches in the field of High Performance Networked Computing have become active.

This Special Issue is focusing on Technologies, Architecture, Systems, Applications, Investigation/Development of relevant tools, algorithms, and paradigms and Service issues on Media, Services and Information Management on the high performance Next-generation networked computing environment. The issue will provide information for interested researchers to exchange new ideas, and reporting chances on progress of current developments and applications of effective techniques.

This special issue seeks articles from researchers and practitioners from all related disciplines, and it will focus on theoretical and practical issues about the high performance computing environment. Topics of this SI include (but are not limited to) the following:

- Interactive Networked Media on Distributed Environments
- Media Content Management on Grid/Distributed Computing Environments
- Networked Service Issues on U-biz./M-biz./E-biz.
- Service System on Grid/Distributed Computing Environments
- Ubiquitous Media Streaming with Mobility, QoS on High performance Network
- Intelligent Data Processing/Management/service Technology on u-City and BI
- Persistent and Autonomous Media Sharing
- Flexible and Transparent Protocols & Middleware for Networked Media
- Intelligent Information Processing/Management/service Technology

---

G. Kou (✉)  
School of Management and Economics, University of Electronic Science and Technology of China,  
Chengdu 610054, China  
e-mail: [kougang@yahoo.com](mailto:kougang@yahoo.com)

This special issue seeks both high-quality academic (theoretical or empirical) and practical papers presenting the current state of the art from researchers and practitioners from all related disciplines. All papers will be required to have some numerical or experimental illustrations. In particular, the numerical experiments must have a scientific value of its own. Furthermore, experimental comparisons with other approaches are strongly encouraged.

After several rounds of peer-review, nine papers out of 33 submissions were accepted. These papers were contributed from 35 authors from six countries or regions: Canada, China, Hong Kong, Iran, South Korea, and USA. In the following, the nine accepted papers are introduced.

Resource allocation is a complicated task in cloud computing environment because there are many alternative computers with varying capacities. The goal of the first paper is to propose a model for task-oriented resource allocation in cloud computing environment. Resource allocation task is ranked by the pairwise comparison matrix technique and the Analytic Hierarchy Process giving the available resources and user preferences. The computing resources can be allocated according to the rank of tasks. Furthermore, an induced bias matrix is further used to identify the inconsistent elements and improve the consistency ratio when conflicting weights in various tasks are assigned. Two illustrative examples are introduced to validate the proposed method.

The appearance of media applications with high bandwidth and quality of service requirements has made a significant impact in telecommunications technology. In this direction, the IEEE802.16 has defined wireless access systems called WiMAX. These systems provide high speed communications over a long distance. For this purpose some service classes with QoS requirements are defined; but QoS scheduler is not standardized in IEEE802.16. Scheduling mechanism has a significant effect on the performance of WiMAX systems for use of bandwidth and radio resources. Some scheduling algorithms are introduced by researchers; but they only provide some limited aspects of QoS. An intelligent decision support system is therefore necessary for scheduling. In the second paper, a fuzzy based scheduling system is proposed for compounds of real-time and non-real-time polling services which provide QoS requirements and fairness in dynamic conditions. A series of simulation experiments have been carried out to evaluate the performance of the proposed scheduling algorithm in terms of latency and throughput QoS parameters. The results show that the proposed method performs effectively regarding both of these criteria and achieves proportional system performance and fairness among different types of traffic.

The rapid development of the Internet and the Internet of Things accelerates the emergence of the hyper world. It has become a pressing research issue to realize the organic amalgamation and harmonious symbiosis among humans, computers, and things in the hyper world, which consists of the social world, the physical world and the information world (cyber world). In the third paper, the notion of Wisdom Web of Things (W2T) is proposed in order to address this issue. As inspired by the material cycle in the physical world, the W2T focuses on the data cycle, namely “from things to data, information, knowledge, wisdom, services, humans, and then back to things.” A W2T data cycle system is designed to implement such a cycle, which is, technologically speaking, a practical way to realize the harmonious symbiosis of humans, computers, and things in the emerging hyper world.

Efficient multimedia and specifically image authentication is critical and in demand to protect data vulnerability in Wireless Multimedia Sensor Networks (WMSN). This is to prevent malicious intruders from modifying and forging image contents over a network. Watermarking technique has been widely used to assert an image data authentication over wired networks; however, resource constraints (e.g. processing power, communication energy) in small sensors and the state of error-prone wireless channels result in fundamental challenges for developing efficient watermarking schemes in WMSN. These challenges include how to embed/protect/extract watermark efficiently and robustly in low-cost sensors and how to transmit authenticated image and multimedia with high energy efficiency. In the fourth paper, the authors propose a communication-resource-aware and adaptive watermarking scheme for multimedia authentication in WMSN. Their contribution is two fold. First, the transmission quality for the watermark as well as watermarked multimedia authenticity by embedding watermark with adaptive coding redundancies, and by unequally allocating network resources to protect the image and multimedia packets with the watermark information. Second, communication energy efficiency and real-time performance are achieved with the watermark being adaptive to the network condition and the processing delay reduced due to the exploited inter-frame correlation. The simulation and experimental results demonstrate that the proposed adaptive watermarking system can achieve considerable gains in terms of energy saving, image transmission quality and multimedia authentication performance

Since software development has become an essential investment for many organizations recently, both software industry and academic communities are more and more concerned about reliable and accurate estimation of software development effort. The fifth paper puts forward six widely used case-based reasoning (CBR) methods with optimized weights derived from the particle swarm optimization (PSO) method to estimate software effort. Meanwhile, four combination methods are adopted to assemble the results of independent CBR methods. The experiments are carried out using two datasets of software projects from Desharnais dataset and Miyazaki dataset. Experimental results show that different CBR methods can get the best results in different parameters settings, and there is not a best method for software effort estimation among the six different CBR methods. Currently, combination methods proposed in this study outperform independent methods, and the weighted mean combination (WMC) method can get the better result.

Many emerging online stream processing services require the consideration of quality of service (QoS), which is highly dependent on the placement of services at various hosts. The sixth paper investigates the QoS-aware placement problems of stream processing services under different contexts. On condition that the client demands are stable, the QoS-aware placement problem aiming to minimize the cost when servers are CPU uncapacitated, is equivalent to the set cover problem, and can be solved by a greedy algorithm with approximation factor  $O(\log n)$ , where  $n$  is the number of clients. However, when CPU capacity constraints on servers are taken into account, the QoS-aware placement problem cannot be approximated unless  $P = NP$ . Therefore, the authors propose two heuristic algorithms: (1) ISCA (Iterated Set Cover-based Algorithm) and (2) KBA (Knapsack-Based Algorithm). They also consider the placement problem of client demands increasing over time.

Two objectives, called extension factor and system lifetime, are proposed for demand increment-blind and increment-aware models, respectively. Both of them can be solved by extending ISCA and KBA. The experimental results show that ISCA and KBA have distinct effects on different demand sizes. ISCA is more efficient when client demands are relatively smaller, while KBA performs better for larger demands.

Recent development in Graphics Processing Units (GPUs) has enabled inexpensive high performance computing for general-purpose applications. Compute Unified Device Architecture (CUDA) programming model provides the programmers adequate C language like APIs to better exploit the parallel power of the GPU. Data mining is widely used and has significant applications in various domains. However, current data mining toolkits cannot meet the requirement of applications with large-scale databases in terms of speed. In the seventh paper, the authors propose three techniques to speedup fundamental problems in data mining algorithms on CUDA platform, scalable thread scheduling scheme for irregular pattern, parallel distributed top- $k$  scheme, and parallel high dimension reduction scheme. They play a key role in our CUDA-based implementation of three representative data mining algorithms, CU-Apriori, CU-KNN and CU-Kmeans. These parallel implementations outperform the other state-of-the-art implementations significantly on HP xw8600 workstation with a Tesla C1060 GPU and a Core-quad Intel Xeon CPU. Our results have shown that GPU + CUDA parallel architecture is feasible and promising for data mining applications.

Traditional DHT-based P2P networks cannot provide fuzzy search for resources. On the contrary, inverted index based DHT network can provide semantic-based fuzzy search for resources similar to search engines over the Internet. However, inverted indexing in P2P networks may bring about extra costs for the maintenance of the index network, due to the dynamic nature of the P2P network. In the eighth paper, the authors present ROIN, an efficient index network framework for the P2P network. In the ROIN framework, an inverted index network for the fully distributed P2P network is established by assigning nodes with high reputations as index nodes. A group of experiments concerning the search coverage rate, the search latency, the load of index nodes and the network disturbance of the ROIN framework are carried out. A series of experimental evaluations and comparisons are made between ROIN, PCIR, Pastry, and GlusterFS. Upon the experimental results and evaluations, the authors conclude that applying invert index and selecting a suitable number of nodes with high reputations as index nodes will bring about a robust and efficient distributed network that supports fuzzy search for resources.

A two-dimensional (2D) Petersen-torus network is a mesh-class fixed-degree network designed using a Petersen graph, which has a maximum of 10 nodes when the degree is 3 and the diameter is 2 in a  $(d, k)$ -graph problem. In the last paper, the authors propose a new three-dimensional (3D) Petersen-torus network that extends the 2D Petersen-torus network without increasing the degree. The 3D Petersen-torus has the same number of nodes ( $N$ ). The 3D Petersen-torus is better than the well-known 3D torus and 3D honeycomb mesh in terms of diameter and network cost. The 3D Petersen-torus network is better than the hypercube-like and star graph-like networks in terms of extendibility. Hence, the proposed network may serve as the foundation for realizing a high-performance multicomputer. In the paper, the optimal routing algorithm, Hamilton cycle, and several basic attributes are discussed. Furthermore, a

comparison with a mesh-class fixed-degree 3D network is made for degree, diameter, and network cost.

The guest editors hope that the papers published in this special issue would be of interest to academic and industrial communities by promoting high quality, novel and daring research findings, and would provide solutions and tools to challenging problems facing by fellow researchers, practitioners, and students, as well as research challenges and initiatives for further research.

**Acknowledgements** We would like to thank Professor Hamid R. Arabnia, the Editor-in-Chief of The Journal of Supercomputing, for giving us the chance to edit this special issue. We would like to thank all reviewers for their time, energy, and efforts to review the submissions. Last but not least, the guest editor would like to thank the authors of all the submissions to this special issue for their contribution. We do appreciate having had the opportunity to supervise the reviewing process of the submissions. Without the support of all of them, it would have been impossible to compile this special issue for our readers. The guest editor has been supported by grants from the National Natural Science Foundation of China (#70901015 and #71222108).