SOFTWARE: PRACTICE AND EXPERIENCE
Softw. Pract. Exper. 2017; **47**:345–347
Published online 17 January 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.2475

Editorial

Special issue on Big Data and Cloud of Things (CoT)

Cloud computing and Internet of Things (IoT) are two technologies that are already becoming part of our daily lives and are attracting significant interest from both industry and academia. The *Cloud of Things* (CoT) is a vision inspired from the IoT paradigm where everyday devices, namely, '*smart objects*', are fully connected to the internet and are integrated with the cloud. It is expected the IoT will grow to 35 billion units by 2020, making it one of the main sources of 'Big Data' with characteristics such as volume, heterogeneity, complexity, velocity, and value. In recent years, IoT has given rise to a number of new CoT paradigms (but not limited to) including: Sensing-as-a-Service, Sensing- and Actuation-as-a-Service, Video-Surveillance-as-a-Service, Big Data Analytics-as-a-Service, Data-as-a-Service, Sensor-as-a-Service, and Sensor-Event-as-a-Service. Cloud computing is a more mature technology compared to IoT. It can offer virtually unrestricted capabilities (e.g., storage and computation) to support IoT services and application that can exploit the data produced from IoT devices. The cloud essentially acts as a transparent layer between the IoT and applications providing flexibility, scalability, and hiding the complexities between the two layers (IoT and applications). However, the integration of cloud and IoT into Cloud of Things is not straightforward and imposes several challenges. These challenges include IoT device and service discovery, IoT device integration, big data management and analytics, cloud monitoring and orchestration for distributed IoT applications, mobility issues in cloud access, privacy and security, and SLA management for both cloud and IoT. Specific attention must be paid to address a range of issues from IoT data collection, storage, processing, analytics on demand to automatic provision and management of cloud resources to support the growing population of things.

Hence, this special issue solicits paper related to topics including CoT architectures and models for smart provision of CoT applications, data management challenges facing CoT applications, software and tools to monitor, manage, deploy and deliver CoT applications, quality of service and related SLA management and policies for CoT applications, and security and privacy challenges facing CoT applications. The call for special issues received a number of submissions. After a two-phase peer review process, we have accepted 10 high-quality papers related to the aforementioned areas of interest.

The first paper titled *Using adaptive resource allocation to implement an elastic MapReduce framework* by Jiaqi Zhao, Changlong Xue, Xinlin Tao, Shugong Zhang, and Jie Tao addresses the runtime resource demand challenge faced by application running on MapReduce frameworks. The proposed approach is capable of making the map reduce application, aware of overloading or under-loading situations with the resources allocated. They have extended the existing Hadoop MapReduce resource manager to implement the proposed strategy and validated the concept on an high-performance computing cluster with standard benchmark applications. Experimental results show a significant performance gain, for example, an up to 45% improvement in execution time for running multiple applications.

The second paper titled *A traffic hotline discovery method over cloud of things using big taxi GPS data* by Xiaolong Xu, Wanchun Dou, Xuyun Zhang, Chunhua Hu, and Jinjun Chen addresses the challenge of discovering traffic hotline in CoT environments. Traffic hotlines are identified as the traffic lines with intensive traffic flows among traffic spots. They propose a hotline discovery method over CoT by establishing a hotline discovery principle. They have implemented their approach on SAP HANA cloud and tested it using big taxi global positioning system data under two application scenarios.

The third paper *$PM_{2:5}$ forecasting with hybrid LSE model-based approach* by Yunliang Chen, Fangyuan Li1, Ze Deng, Xiaodao Chen and Jijun He addresses the challenges in forecasting over $PM_{2.5}$ time series. They propose a hybrid approach using local mean decomposition and support vector regression–Elman. They use this approach to analyze 5 days ahead $PM_{2:5}$ concentrations for forecasting in Wuhan, China. Experimental results show that, compared with multiple linear regression, autoregressive integrated moving average, backpropagation neural network, and support vector regression models, the proposed hybrid Elman model-based approach exhibits the best performance in terms of R2, mean absolute error, mean absolute percentage error, root mean squared error, while it is applied for forecasting in real datasets.

The fourth paper *Is the data on your wearable device secure? An Android Wear smartwatch case study* by Quang Do, Ben Martini, and Kim-Kwang Raymond Choo addresses the security challenge in cloud of things. In particular, they focus on studying the vulnerability of sensitive data exposure produced by wearable devices. The study first analyzes the types of sensitive user data that may be present on a wearable device and develop a method to demonstrate that they can be exfiltrated by an adversary. They then present a technique that allows an adversary to exfiltrate data from smartwatches. Using this technique, the study concludes that the smartwatch stores a relatively large amount of sensitive user data, including SMS messages, contact information, and biomedical data, and does not effectively protect this user data from physical exfiltration.

The fifth paper *Brain Data processing with massively parallel computing technology: challenges and opportunities* by Dan Chen, Yangyang Hu, Chang Cai, Ke Zeng, and Xiaoli Li presents a survey of challenges and opportunities in processing massive brain data using parallel computing technologies. They identify general-purpose computing on the graphics processing unit technology as an ideal environment for processing brain data, which benefits from the tremendous computing power of modern graphics processing units in massively parallel architecture that is frequently an order of magnitude larger than the modern multi-core CPUs. This article first recaps significant speed-ups of existing algorithms aided by the graphics processing unit and demonstrates a series of successful approaches to processing electroencephalogram data in various dimensions and scales in a massively parallel manner. The findings of their survey indicate massively parallel computing technology manifests great potentials in addressing the grand challenges of brain big data processing.

The sixth paper *CloudEyes: Cloud-based malware detection with reversible sketch for resource-constrained internet of things (IoT) devices* by Hao Sun, Xiaofeng Wang, Rajkumar Buyya, and Jinshu Su addresses the issues of a malware attack on IoT. The paper proposes a cloud-based anti-malware system, called CloudEyes, which provides efficient and trusted security services for resource-constrained IoT devices. CloudEyes guarantees both the data privacy and low-cost communications using a novel signature detection mechanism based on the reversible sketch structure and a lightweight scanning agent that utilizes the digest of signature fragments to dramatically reduce the range of accurate matching. The performance evaluation of CloudEyes demonstrates that the mechanisms in CloudEyes are effective and practical, and the system can outperform other existing systems with less time and communication consumption.

The seventh paper *Ahab: A cloud-based distributed big data analytics framework for the Internet of Things* by Michael Vögler, Johannes M. Schleicher, Christian Inzinger, and Schahram Dustdar present a generic, scalable, and fault-tolerant data processing framework based on the cloud that allows operators to perform online and offline analyses on gathered data to better understand and optimize the behavior of the available smart city infrastructure. Ahab addresses the challenges in efficient collection, processing, and storage of this information in the smart city domain. Ahab is designed for easy integration of new data sources, provides an extensible API to perform custom analysis tasks, and a domain-specific language to define adaptation rules based on analysis results. The paper demonstrates the feasibility of the proposed approach using an example application for autonomous intersection management in smart city environments.

The eighth paper *XHAMI – extended HDFS and MapReduce interface for Big Data image processing applications in cloud computing environments* by Raghavendra Kune, Pramod Kumar Konugurthi, Arun Agarwal, Raghavendra Rao Chillarige, and Rajkumar Buyya addresses the challenges in using Hadoop Distributed File System (HDFS) and MapReduce for scientific applications such as image mining and knowledge data mining. The authors present extended HDFS and

DOI: 10.1002/spe

MapReduce interface (XHAMI) that proposes a two-phase extension to the HDFS and MapReduce programming model. The extended XHAMI interfaces are presented as APIs and its effectiveness demonstrated through an image processing case study. XHAMI works without any changes for the existing MapReduce models and can be utilized by many applications where there is a requirement of overlapped data.

The ninth paper *Resource requests prediction in the cloud computing environment with a deep belief network* by Weishan Zhang, Pengcheng Duan, Laurence T. Yang, Feng Xia, Zhongwei Li, Qinghua Lu, Wenjuan Gong, and Su Yang addresses the classic problem of accurately predicting resource requests to achieve optimal job scheduling and load balancing for cloud computing. In particular, they focus on providing a high level of accuracy using a deep belief network-based approach. The proposed approach achieves high accuracy with mean square error of $[10^{-6}, 10^{-5}]$, approximately 72% reduction compared with the traditional autoregressive integrated moving average predictor, and has better prediction accuracy compared with the state-of-art fractal modeling approach.

The final paper *Adaptable secure communication for the Cloud of Things* by Valter Vasic, Aleksander Antonic, Kresimir Pripužic, Miljenko Mikuc, and Ivana Podnar Zarko addresses the challenges in providing secure, energy-efficient, and reliable CoT services. They present an adaptable model for secure communication in CoT environments that defines six secure communication operations to enable CoT entities to autonomously and dynamically agree on the security protocol and cryptographic keys used for communication. The proposed solution is verified by a prototype implementation, which takes into account the capabilities of communicating devices. Experimental evaluation demonstrates the viability of the proposed solution in real-world deployments.

Guest Editors

Rajiv Ranjan
China University of Geosciences, Wuhan, China
Data61, CSIRO, Canberra, Australia
E-mail: *rranjans@gmail.com*

Lizhe Wang
China University of Geosciences, Wuhan, China

Prem Prakash Jayaraman
Swinburne University of Technology, Melbourne, Australia

Karan Mitra
Lulea University of Technology, Lulea, Sweden

Dimitrios Georgakopoulos
Swinburne University of Technology, Melbourne, Australia

DOI: 10.1002/spe