2015 Third International Conference on Advanced Cloud and Big Data (CBD 2015)

Yangzhou, China
30 October – 1 November 2015

IEEE Catalog Number: CFP1503Y-POD
ISBN: 978-1-4673-8538-1

Copyright © 2015 by the Institute of Electrical and Electronic Engineers, Inc
All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

******This publication is a representation of what appears in the IEEE Digital Libraries. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number: CFP1503Y-POD
ISBN (Print-On-Demand): 978-1-4673-8538-1
ISBN (Online): 978-1-4673-8537-4

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

2015 Third International Conference on Advanced Cloud and Big Data

CBD 2015

Table of Contents

Welcome Message from the General Co-Chairs....x

Organizing Committee....xi

Program Committee....xii

Keynote 1: Big Data Can Open New Era of Healthcare by Sungyoung Lee....xiv

Keynote 2: A Cloud-Based System for Facilitating Collaborative and Micro Learning in Mobile MOOC Learning by Jun Shen....xvi

Keynote 3: Interplay between Machine Learning and Artificial Intelligence by Yixin Chen....xviii

---

Session 1: Social Computing and Service Computing I

Many Is Better Than All: Efficient Selfish Load Balancing in Mobile Crowdsourcing Systems 1
Qingyu Li, Panlong Yang, Shaojie Tang, Chaocan Xiang, and Fan Li

Discovering Topic Evolution Topology in a Microblog Corpus 7
Bin Zhao, Wei Xu, Genlin Ji, and Chao Tan

Social Recommendation Algorithm Based on the Context of Time and Tags 15
Xinhua Wang and Lingyuan Dou

A Study on the Retweeting Behaviour of Marketing Microblogs with High Retweets in Sina Weibo 20
Zizhu Zhang, Bing Li, Weiliang Zhao, and Jian Yang

A Topic Community-Based Method for Friend Recommendation in Online Social Networks via Joint Nonnegative Matrix Factorization 28
Chaobo He, Hanchao Li, Xiang Fei, Yong Tang, and Jia Zhu

Position Prediction Social-Relationship-Based on Multi-order Markov Model 36
Yue Zou and Sanfeng Zhang

Session 2: Big Data and Applications I

NIS: A New Index Scheme for Flash File System 44
Yang Ou, Xiaoquan Wu, Nong Xiao, Fang Liu, and Wei Chen

Parallel Area Navigation Enhanced Information Extraction Algorithm Based on Massive Historical GNSS Data and MapReduce 52
Dengao Li, Gang Wu, Jumin Zhao, Wenhui Niu, and Shuai Li

Replication Management Framework for HDFS Based on Prediction Technique 58
Dinh-Mao Bui, Thien Huynh-The, Sungyoung Lee, Bin Li, and Jin Wang

Memory-Based Data Storing Technologies on Hadoop Distribution File System 64
Aibo Song, Jinlin Tu, Jinghua Zhao, and Xuejiao Qian

VPCH: A Consistent Hashing Algorithm for Better Load Balancing in a Hadoop Environment 69
Qi Liu, Weidong Cai, Jian Shen, Baowei Wang, Zhangjie Fu, and Nigel Linge

An Automatic Discovery Framework of Cross-Source Data Inconsistency for Web Big Data 73
Sha Yang, Wei Yu, Yahui Hu, Kai Wang, Jun Wang, and Shijun Li

Session 3: Cloud Computing and Applications I

An Energy-Aware QoS Enhanced Method for Service Computing across Clouds and Data Centers 80
Wanchun Dou, Xiaolong Xu, Shunmei Meng, and Shui Yu

A Randomized Heuristic for Stochastic Workflow Scheduling on Heterogeneous Systems 88
Wei Zheng, Bugingo Emmanuel, and Chen Wang

A Power-Aware Virtual Machine Mapper Using Firefly Optimization 96
Shoubao Su, Yu Su, Fei Shao, and Haifeng Guo

Online Scheduling of Multiple Deadline-Constrained Workflow Applications in Distributed Systems 104
Wei Zheng, Chao Xu, and Wen Bao

An Improved Adaptive Workflow Scheduling Algorithm in Cloud Environments 112
Yinjuan Zhang and Yun Li

A QoS Guarantee Framework for Cloud Services Based on Bayesian Prediction 117
Wei Chen, Jiming Chen, Jine Tang, and Liangmin Wang

Session 4: Social Computing and Service Computing II

Traffic-Aware Data Placement for Online Social Networks 125
Jingya Zhou, Jianxi Fan, Jin Wang, and Juncheng Jia

Cloud-Based Infrastructure for Mobile Testing as a Service 133
Chuanqi Tao, Jerry Gao, and Bixin Li

Using Suffix Tree to Detect Communities from Bipartite Network 141
Dai Caiyan, Chen Ling, and Chen Bolun

Clustering on Uncertain Data Stream over Sliding Windows 148
Li Tu

Session 5: Big Data and Applications II

vGraph: Graph Virtualization towards Big Data 153
Maqbool Khan, Meng Liu, Wanchun Dou, and Shui Yu

Parallel Optimization of Pixel Purity Index Algorithm for Hyperspectral Unmixing Based on Spark 159
Jinping Gu, Zebin Wu, Yonglong Li, Yufeng Chen, Zhihui Wei, and Wubin Wang

Superblock: An Application-Aware Dynamic Partition Strategy for Large-Scale Graph 167
Junxue Zhang, Fang Dong, Dian Shen, Jiahui Jin, and Junzhou Luo

Analyzing Microarray Data with Classification and Clustering Methods 175
Shaohua Wan

A Fast Level Set Model for Intensity Inhomogeneity Correction in eHealth Analysis System 180
Huimin Lu, Junwu Zhu, Bin Li, Yujie Li, and Seiichi Serikawa

Information Filtering Based on Personalized Topology Information 184
Bolun Chen and Ling Chen

Session 6: Cloud Computing and Applications II

Non-cooperative Resource Allocation Scheme for Data Access in VANET Cloud Environment 190
Jun Tao, Ziyi Zhang, Fuqin Feng, Jian He, and Yifan Xu

PROXY: A High Cost-Performance Cloud Storage Gateway 197
Jibin Wang, Lili Yang, Hu Zhang, Zhaogang Xu, and Ying Guo

SSDP: A Slot-Sensitive Data Placement Strategy in Cloud Computing 205
Tian Tian, Peng Liu, and HuaXing Kuang

Evaluating the Performance of Erasure Codes 213
Ping Lu, Zhihao Huang, Shengmei Luo, Hui Li, and Jun Chen

Session 7: Cloud Computing and Security

A Semantic Hash Tree Based Verifiable Data Access Protocol on the Cloud 219
Fei Chen, Tao Xiang, Jianyong Chen, Wei Yu, Xinwen Fu, and Shengyu Zhang

A Novel Behavior-Based Tracking Attack for User Identification 227
Xiaodan Gu, Ming Yang, Jiaxuan Fei, Zhen Ling, and Junzhou Luo

A Personalized Extended (a, k)-Anonymity Model 234
Xiangwen Liu, Qingqing Xie, and Liangmin Wang

Block-Based Method and Its System for Outsourcing Data by Using Onions of Encryption 241
Shurong Ping, Haiqin Wu, and Liangmin Wang

A Novel User Model Based on Searchable Encryption Scheme 247
Youjing Lu, Qingqing Xie, and Liangmin Wang

Detection and Prevention of Code Injection Attacks on HTML5-Based Apps 254
Xi Xiao, Ruibo Yan, Runguo Ye, Qing Li, Sancheng Peng, andYong Jiang

Session 8: Cloud Computing and Networking

An Improved Energy-Efficient Clustering Algorithm Based on MECA and PEGASIS for WSNs 262
Jin Wang, Jiayi Cao, Yiquan Cao, Bin Li, and Sungyoung Lee

A Mobile-Sink Based Energy Efficiency Clustering Routing Algorithm for WSNs in Coal Mine 267
Jian Shen, Dengzhi Liu, Yongjun Ren, Sai Ji, Jin Wang, and Dongmin Choi

A Novel Method for Virtual Network Embedding with Incentive Convergence Mechanism 275
Cong Wang, Guohua Liu, and Ying Yuan

The Experimentation and Analysis of Adobe RTMFP Traffic Locality 282
Miaoxin Xu and Zhijie Han

Real-Time True-Color Synthesis of Remote Sensing Image with CUDA 286
Baolin Chen, Xiaoyu Li, Hang Lei, and Hang Wang

Session 9: Big Data Science and Applications

A New Closed Frequent Itemsets Mining Algorithm Based on GPU 291
Yun Li, Jie Xu, and Ling Chen

Using Feature-Interface Graph for Automatic Interface Recommendation: A Case Study 296
Wei Shi, Xiaobing Sun, Bin Li, Yucong Duan, and Xiangyue Liu

SLFE: A New Semi-supervised Local Feature Extraction Algorithm 304
Chao Tan, Genlin Ji, and Bin Zhao

Manifold Density Peaks Clustering Algorithm 311
Xiaohua Xu, Yongsheng Ju, Yali Liang, and Ping He

A Parallel Clustering Algorithm Based on Grid Index for Spatio-temporal Trajectories 319
Min Wang, Genlin Ji, Bin Zhao, and Mengmeng Tang

Analysis of Line Loss Based on Matlab and Oracle Database 327
Xue-sen Quan, Pei-hong Wang, Yong-sheng Hao, Ming Zhao, and Meng-yang Li

Author Index 332