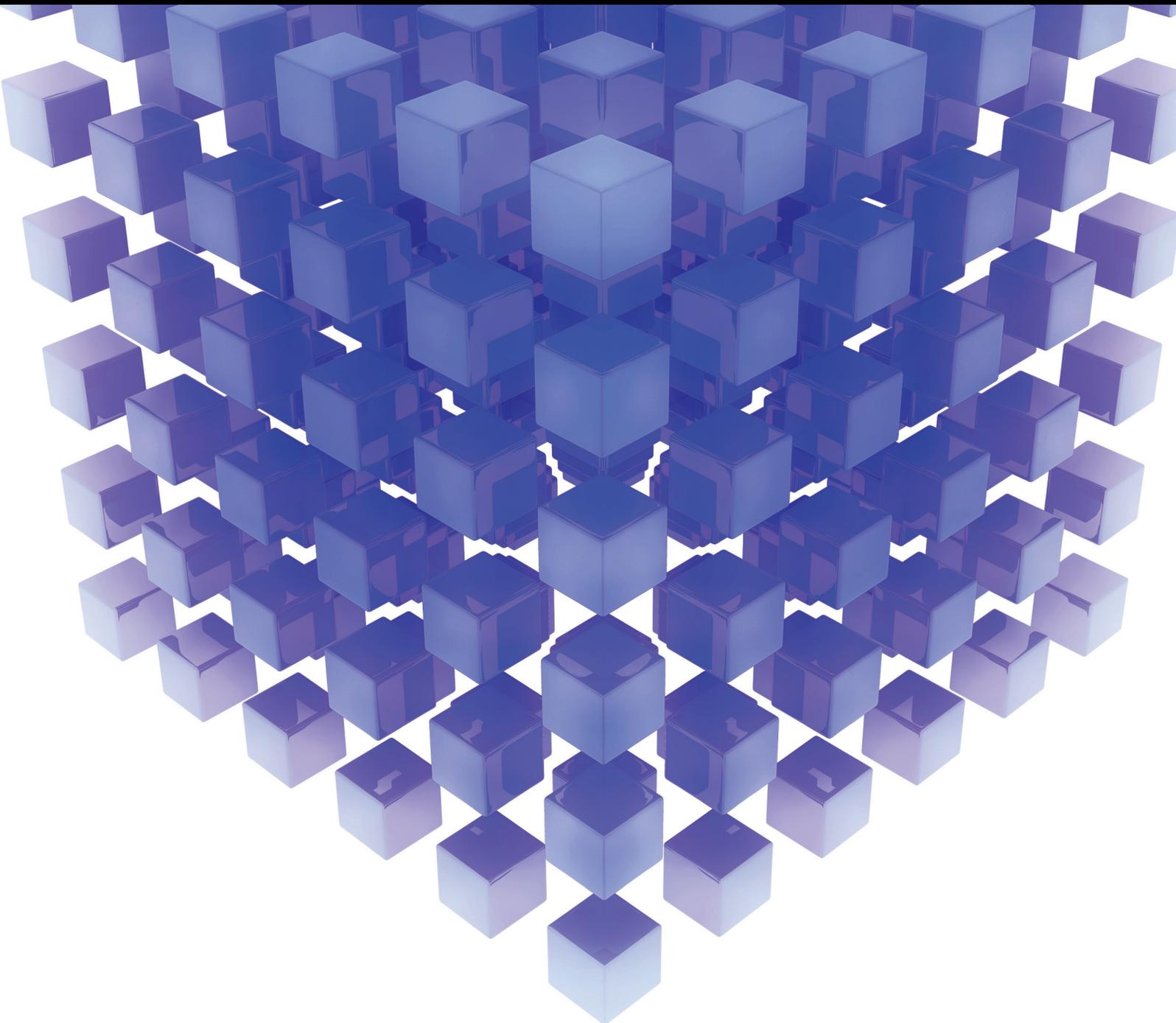


Mathematical Problems in Engineering

Propagation Phenomena and Transitions in Complex Systems 2014

Guest Editors: Ezzat G. Bakhoun, Haipeng Peng, and Florin Pop





**Propagation Phenomena and Transitions
in Complex Systems 2014**

Mathematical Problems in Engineering

**Propagation Phenomena and Transitions
in Complex Systems 2014**

Guest Editors: Ezzat G. Bakhoun, Haipeng Peng,
and Florin Pop



Copyright © 2015 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in “Mathematical Problems in Engineering.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Mohamed Abd El Aziz, Egypt
Farid Abed-Meraim, France
Silvia Abrahão, Spain
Paolo Addresso, Italy
Claudia Adduce, Italy
Ramesh Agarwal, USA
Juan C. Agüero, Australia
Ricardo Aguilar-López, Mexico
Tarek Ahmed-Ali, France
Hamid Akbarzadeh, Canada
Muhammad N. Akram, Norway
Mohammad-Reza Alam, USA
Salvatore Alfonzetti, Italy
Francisco Alhama, Spain
Juan A. Almendral, Spain
Saiied Aminossadati, Australia
Lionel Amodeo, France
Igor Andrianov, Germany
Sebastian Anita, Romania
Renata Archetti, Italy
Felice Arena, Italy
Sabri Arik, Turkey
Fumihiko Ashida, Japan
Hassan Askari, Canada
Mohsen Asle Zaeem, USA
Francesco Aymerich, Italy
Seungik Baek, USA
Khaled Bahlali, France
Laurent Bako, France
Stefan Balint, Romania
Alfonso Banos, Spain
Roberto Baratti, Italy
Martino Bardi, Italy
Azeddine Beghdadi, France
Abdel-Hakim Bendada, Canada
Ivano Benedetti, Italy
Elena Benvenuti, Italy
Jamal Berakdar, Germany
Enrique Berjano, Spain
Jean-Charles Beugnot, France
Simone Bianco, Italy
David Bigaud, France
Jonathan N. Blakely, USA
Paul Bogdan, USA
Daniela Boso, Italy

Abdel-Ouahab Boudraa, France
Francesco Braghin, Italy
Michael J. Brennan, UK
Gunther Brenner, Germany
Maurizio Brocchini, Italy
Julien Bruchon, France
Javier Buldú, Spain
Tito Busani, USA
Pierfrancesco Cacciola, UK
Salvatore Caddemi, Italy
Jose E. Capilla, Spain
Ana Carpio, Spain
Miguel E. Cerrolaza, Spain
Mohammed Chadli, France
Gregory Chagnon, France
Ching-Ter Chang, Taiwan
Michael J. Chappell, UK
Kacem Chehdi, France
Xinkai Chen, Japan
Chunlin Chen, China
Francisco Chicano, Spain
Hung-Yuan Chung, Taiwan
Joaquim Ciurana, Spain
John D. Clayton, USA
Carlo Cosentino, Italy
Paolo Crippa, Italy
Erik Cuevas, Mexico
Peter Dabnichki, Australia
Luca D'Acerno, Italy
Weizhong Dai, USA
P. Damodaran, USA
Farhang Daneshmand, Canada
Fabio De Angelis, Italy
Stefano de Miranda, Italy
Filippo de Monte, Italy
Xavier Delorme, France
Luca Deseri, USA
Yannis Dimakopoulos, Greece
Zhengtao Ding, UK
Ralph B. Dinwiddie, USA
Mohamed Djemai, France
Alexandre B. Dolgui, France
George S. Dulikravich, USA
Bogdan Dumitrescu, Finland
Horst Ecker, Austria

Karen Egiazarian, Finland
Ahmed El Hajjaji, France
Fouad Erchiqui, Canada
Anders Eriksson, Sweden
Giovanni Falsone, Italy
Hua Fan, China
Yann Favennec, France
Giuseppe Fedele, Italy
Roberto Fedele, Italy
Jacques Ferland, Canada
Jose R. Fernandez, Spain
S. Flapper, The Netherlands
Thierry Floquet, France
Eric Florentin, France
Francesco Franco, Italy
Tomonari Furukawa, USA
Mohamed Gadala, Canada
Matteo Gaeta, Italy
Zoran Gajic, USA
Ciprian G. Gal, USA
Ugo Galvanetto, Italy
Akemi Gálvez, Spain
Rita Gamberini, Italy
Maria Gandarias, Spain
Arman Ganji, Canada
Zhong-Ke Gao, China
Xin-Lin Gao, USA
Giovanni Garcea, Italy
Fernando Garcia, Spain
Laura Gardini, Italy
Alessandro Gasparetto, Italy
Vincenzo Gattulli, Italy
Jürgen Geiser, Germany
Oleg V. Gendelman, Israel
Mergen H. Ghayesh, Australia
Anna M. Gil-Lafuente, Spain
Hector Gómez, Spain
Rama S. R. Gorla, USA
Oded Gottlieb, Israel
Antoine Grall, France
Jason Gu, Canada
Quang Phuc Ha, Australia
Ofer Hadar, Israel
Masoud Hajarian, Iran
Frédéric Hamelin, France

Zhen-Lai Han, China
Thomas Hanne, Switzerland
Takashi Hasuike, Japan
Xiao-Qiao He, China
María I. Herreros, Spain
Vincent Hilaire, France
Eckhard Hitzer, Japan
J. Horacek, Czech Republic
Muneo Hori, Japan
András Horváth, Italy
Gordon Huang, Canada
Sajid Hussain, Canada
Asier Ibeas, Spain
Giacomo Innocenti, Italy
Emilio Insfran, Spain
Nazrul Islam, USA
Payman Jalali, Finland
Reza Jazar, Australia
Khalide Jbilou, France
Linni Jian, China
Zhongping Jiang, USA
Bin Jiang, China
Ningde Jin, China
Grand R. Joldes, Australia
Joaquim Joao Judice, Portugal
Tadeusz Kaczorek, Poland
Tamas Kalmar-Nagy, Hungary
Tomasz Kapitaniak, Poland
Haranath Kar, India
K. Karamanos, Belgium
C. M. Khalique, South Africa
Nam-Il Kim, Korea
Do Wan Kim, Korea
Oleg Kirillov, Germany
Manfred Krafczyk, Germany
Frederic Kratz, France
Jurgen Kurths, Germany
Kyandoghere Kyamakya, Austria
Davide La Torre, Italy
Risto Lahdelma, Finland
Hak-Keung Lam, UK
Antonino Laudani, Italy
Aime' Lay-Ekuakille, Italy
Marek Lefik, Poland
Yaguo Lei, China
Thibault Lemaire, France
Stefano Lenci, Italy
Roman Lewandowski, Poland

Qing Q. Liang, Australia
Panos Liatsis, UK
Wanquan Liu, Australia
Yan-Jun Liu, China
Peter Liu, Taiwan
Peide Liu, China
Jean J. Loiseau, France
Paolo Lonetti, Italy
Luis M. López-Ochoa, Spain
V. C. Loukopoulos, Greece
Valentin Lychagin, Norway
Fazal Mahomed, South Africa
Yassir T. Makkawi, UK
N. Manamanni, France
Didier Maquin, France
Paolo Maria Mariano, Italy
Benoit Marx, France
Gefhrard A. Maugin, France
Driss Mehdi, France
Roderick Melnik, Canada
Pasquale Memmolo, Italy
Xiangyu Meng, Canada
Jose Merodio, Spain
Luciano Mescia, Italy
Laurent Mevel, France
Yuri V. Mikhlin, Ukraine
Aki Mikkola, Finland
Hiroyuki Mino, Japan
Pablo Mira, Spain
Vito Mocella, Italy
Roberto Montanini, Italy
Gisele Mophou, France
Rafael Morales, Spain
Aziz Moukrim, France
Emiliano Mucchi, Italy
Domenico Mundo, Italy
Jose J. Muñoz, Spain
Giuseppe Muscolino, Italy
Marco Mussetta, Italy
Hakim Naceur, France
Hassane Naji, France
Dong Ngoduy, UK
Tatsushi Nishi, Japan
Ben T. Nohara, Japan
Mohammed Nouari, France
Mustapha Nourelfath, Canada
Sotiris K. Ntouyas, Greece
Roger Ohayon, France

Mitsuhiro Okayasu, Japan
Eva Onaindia, Spain
Javier Ortega-Garcia, Spain
A. Ortega-Moñux, Spain
Naohisa Otsuka, Japan
Erika Ottaviano, Italy
Alkiviadis Paipetis, Greece
Alessandro Palmeri, UK
Anna Pandolfi, Italy
Elena Panteley, France
Manuel Pastor, Spain
Pubudu Pathirana, Australia
Francesco Pellicano, Italy
Haipeng Peng, China
Mingshu Peng, China
Zhike Peng, China
Marzio Pennisi, Italy
Matjaz Perc, Slovenia
Francesco Pesavento, Italy
Maria do R. Pinho, Portugal
Antonina Pirrotta, Italy
Vicent Pla, Spain
Javier Plaza, Spain
Jean-Christophe Ponsart, France
Mauro Pontani, Italy
Stanislav Potapenko, Canada
Sergio Preidikman, USA
Christopher Pretty, New Zealand
Carsten Proppe, Germany
Luca Pugi, Italy
Yuming Qin, China
Dane Quinn, USA
Jose Ragot, France
K. R. Rajagopal, USA
Gianluca Ranzi, Australia
Sivaguru Ravindran, USA
Alessandro Reali, Italy
Giuseppe Rega, Italy
Oscar Reinoso, Spain
Nidhal Rezg, France
Ricardo Riaza, Spain
Gerasimos Rigatos, Greece
José Rodellar, Spain
Rosana Rodriguez-Lopez, Spain
Ignacio Rojas, Spain
Carla Roque, Portugal
Aline Roumy, France
Debasish Roy, India

Rubén Ruiz García, Spain
Antonio Ruiz-Cortes, Spain
Ivan D. Rukhlenko, Australia
Mazen Saad, France
Kishin Sadarangani, Spain
Mehrdad Saif, Canada
Miguel A. Salido, Spain
Roque J. Saltarén, Spain
Francisco J. Salvador, Spain
Alessandro Salvini, Italy
Maura Sandri, Italy
Miguel A. F. Sanjuan, Spain
Juan F. San-Juan, Spain
Roberta Santoro, Italy
Ilmar F. Santos, Denmark
José A. Sanz-Herrera, Spain
Nickolas S. Sapidis, Greece
E. J. Sapountzakis, Greece
Themistoklis P. Sapsis, USA
Andrey V. Savkin, Australia
Valery Sbitnev, Russia
Thomas Schuster, Germany
Mohammed Seaid, UK
Lotfi Senhadji, France
Joan Serra-Sagrasta, Spain
Leonid Shaikhet, Ukraine
Hassan M. Shanechi, USA
Sanjay K. Sharma, India
Bo Shen, Germany
Babak Shotorban, USA
Zhan Shu, UK
Dan Simon, USA
Luciano Simoni, Italy
Christos H. Skiadas, Greece

Michael Small, Australia
Francesco Soldovieri, Italy
Raffaele Solimene, Italy
Ruben Specogna, Italy
Sri Sridharan, USA
Ivanka Stamova, USA
Yakov Strelniker, Israel
Sergey A. Suslov, Australia
Thomas Svensson, Sweden
Andrzej Swierniak, Poland
Yang Tang, Germany
Sergio Teggi, Italy
Roger Temam, USA
Alexander Timokha, Norway
Rafael Toledo, Spain
Gisella Tomasini, Italy
Francesco Tornabene, Italy
Antonio Tornambe, Italy
Fernando Torres, Spain
Fabio Tramontana, Italy
Sébastien Tremblay, Canada
Irina N. Trendafilova, UK
George Tsiatas, Greece
Antonios Tsovdos, UK
Vladimir Turetsky, Israel
Mustafa Tutar, Spain
Efstratios Tzirtzilakis, Greece
Filippo Ubertini, Italy
Francesco Ubertini, Italy
Hassan Ugail, UK
Giuseppe Vairo, Italy
Kuppalapalle Vajravelu, USA
Robertt A. Valente, Portugal
Raoul van Loon, UK

Pandian Vasant, Malaysia
M. E. Vázquez-Méndez, Spain
Josep Vehi, Spain
Kalyana C. Veluvolu, Korea
Fons J. Verbeek, The Netherlands
Franck J. Vernerey, USA
Georgios Veronis, USA
Anna Vila, Spain
Rafael J. Villanueva, Spain
Uchchukwu E. Vincent, UK
Mirko Viroli, Italy
Michael Vynnycky, Sweden
Shuming Wang, Singapore
Junwu Wang, China
Yan-Wu Wang, China
Yongqi Wang, Germany
J. Witteveen, The Netherlands
Yuqiang Wu, China
Dash Desheng Wu, Canada
Xuejun Xie, China
Guangming Xie, China
Gen Qi Xu, China
Hang Xu, China
Xinggong Yan, UK
Luis J. Yebra, Spain
Peng-Yeng Yin, Taiwan
Ibrahim Zeid, USA
Qingling Zhang, China
Huaguang Zhang, China
Jian Guo Zhou, UK
Quanxin Zhu, China
Mustapha Zidi, France
Alessandro Zona, Italy

Contents

Propagation Phenomena and Transitions in Complex Systems 2014, Ezzat G. Bakhoun, Haipeng Peng, and Florin Pop

Volume 2015, Article ID 439507, 2 pages

An Effective Error Correction Scheme for Arithmetic Coding, Qiuzhen Lin, Kwok-Wo Wong, Ming Li, and Jianyong Chen

Volume 2015, Article ID 861093, 10 pages

Topology Identification of Coupling Map Lattice under Sparsity Condition, Jiangni Yu, Lixiang Li, and Yixian Yang

Volume 2015, Article ID 303454, 6 pages

Cryptanalysis and Improvement of the Robust and Blind Watermarking Scheme for Dual Color Image, Hai Nan, Bin Fang, Weibin Yang, Jiye Qian, Ming Li, Yi Liu, and Yushu Zhang

Volume 2015, Article ID 526174, 10 pages

Alternating Coordinate-Momentum Representation for Quantum States Based on Bopp Operators for Modelling Long-Distance Coherence Aspects, Ezzat G. Bakhoun and Cristian Toma

Volume 2015, Article ID 818019, 7 pages

A Parallel Community Structure Mining Method in Big Social Networks, Songchang Jin, Philip S. Yu, Shudong Li, and Shuqiang Yang

Volume 2015, Article ID 934301, 13 pages

On the Propagation of Longitudinal Stress Waves in Solids and Fluids by Unifying the Navier-Lame and Navier-Stokes Equations, Ahmad Barzkar and Hojatollah Adibi

Volume 2015, Article ID 789238, 9 pages

A Secure and Effective Anonymous Integrity Checking Protocol for Data Storage in Multicloud,

Lingwei Song, Dawei Zhao, Xuebing Chen, Chenlei Cao, and Xinxin Niu

Volume 2015, Article ID 614375, 8 pages

Stochastic Fractional Heat Equations Driven by Fractional Noises, Xichao Sun and Ming Li

Volume 2015, Article ID 421705, 16 pages

A New Approach for Flexible Molecular Docking Based on Swarm Intelligence, Yi Fu, Xiaojun Wu, Zhiguo Chen, Jun Sun, Ji Zhao, and Wenbo Xu

Volume 2015, Article ID 540186, 10 pages

Obtaining Cross Modal Similarity Metric with Deep Neural Architecture, Ruifan Li, Fangxiang Feng, Xiaojie Wang, Peng Lu, and Bohan Li

Volume 2015, Article ID 293176, 9 pages

Prevention and Trust Evaluation Scheme Based on Interpersonal Relationships for Large-Scale Peer-To-Peer Networks, Lixiang Li, Jürgen Kurths, Yixian Yang, and Guole Liu

Volume 2014, Article ID 189213, 11 pages

Topological Embedding Feature Based Resource Allocation in Network Virtualization, Hongyan Cui, Shaohua Tang, Fangfang Sun, Yue Xu, and Xiaoli Yang

Volume 2014, Article ID 271493, 10 pages

Editorial

Propagation Phenomena and Transitions in Complex Systems 2014

Ezzat G. Bakhoum,¹ Haipeng Peng,² and Florin Pop³

¹*University of West Florida, 11000 University Parkway, Pensacola, FL 32514, USA*

²*Beijing University of Posts and Telecommunications, Beijing, China*

³*Politehnica University, Bucharest, Romania*

Correspondence should be addressed to Ezzat G. Bakhoum; ebakhoum@uwf.edu

Received 2 April 2015; Accepted 2 April 2015

Copyright © 2015 Ezzat G. Bakhoum et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The challenge in advanced engineering applications based on efficient mathematical models for propagation and transition phenomena can be noticed nowadays in many research fields. Fractal theory and special mathematical functions are used not only for the design of nanostructures but also for studying propagation in complex artificial networks. Differential geometry is adapted for solving nonlinear partial differential equations with great number of variables for modelling propagation and transitions for different type of electromagnetic, acoustic, and optic waves. Commutative and/or additive consequences of quantum physics are used extensively in the design of long range transmission systems. Advanced mathematical tools connected to wavelets are recommended for biological phenomena. All these advanced engineering subjects require efficient mathematical models adapted for nonlinear propagation phenomena and for complex systems, when specific limitations are involved (very long distance propagation, fractal aspects and transitions in nanostructures, complex systems with great number of variables, and infinite spatiotemporal extension of material media). Using advanced mathematical tools for modeling propagation and transition phenomena, this special issue presents high qualitative and innovative developments for efficient mathematical approaches of propagation phenomena and transitions in complex systems. Significant results were obtained for propagation of waves in advanced materials dynamics of complex systems, efficient signal and image analysis based on fundamental mathematical and physical laws, and transitions in complex networks.

This special issue involves 12 original papers selected by the editors so as to present the most significant results in the above mentioned topics. These papers are organized as follows.

Three papers on analytical methods based on swarm intelligence and data mining are as follows: (i) "A New Approach for Flexible Molecular Docking Based on Swarm Intelligence," by Y. Fu et al.; (ii) "A Parallel Community Structure Mining Method in Big Social Networks," by S. Jin et al.; and (iii) "Prevention and Trust Evaluation Scheme Based on Interpersonal Relationships for Large-Scale Peer-to-Peer Networks," by L. Li et al.

Three papers on watermarking, feature embedding, and data integrity checking are as follows: (i) "A Secure and Effective Anonymous Integrity Checking Protocol for Data Storage in Multicloud," by L. Song et al.; (ii) "Cryptanalysis and Improvement of the Robust and Blind Watermarking Scheme for Dual Color Image," by H. Nan et al.; and (iii) "Topological Embedding Feature Based Resource Allocation in Network Virtualization," by H. Cui et al.

Three papers on coherence, mapping, and similarity modeling are as follows: (i) "Obtaining Cross Modal Similarity Metric with Deep Neural Architecture," by R. Li et al.; (ii) "Alternating Coordinate-Momentum Representation for Quantum States Based on Bopp Operators for Modelling Long-Distance Coherence Aspects," by E. G. Bakhoum and C. Toma; and (iii) "Topology Identification of Coupling Map Lattice under Sparsity Condition," by J. Yu et al.

Two papers on aspects of wave propagation are as follows:

- (i) “On the Propagation of Longitudinal Stress Waves in Solids and Fluids by Unifying the Navier-Lame and Navier-Stokes Equations,” by A. Barzkar and H. Adibi, and (ii) “Stochastic Fractional Heat Equations Driven by Fractional Noises,” by X. Sun and M. Li.

One paper on error correction schemes is as follows:

- (i) “An Effective Error Correction Scheme for Arithmetic Coding,” by Q. Lin et al.

Ezzat G. Bakhoun
Haipeng Peng
Florin Pop

Research Article

An Effective Error Correction Scheme for Arithmetic Coding

Qiuzhen Lin,^{1,2} Kwok-Wo Wong,² Ming Li,^{3,4} and Jianyong Chen¹

¹College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

²Department of Electronic Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong 999077, Hong Kong

³Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China

⁴School of Information Science and Technology, East China Normal University, Shanghai 200241, China

Correspondence should be addressed to Jianyong Chen; cjyok2000@hotmail.com

Received 9 September 2014; Accepted 28 November 2014

Academic Editor: Ezzat G. Bakhoun

Copyright © 2015 Qiuzhen Lin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose an effective error correction technique for arithmetic coding with forbidden symbol. By predicting the occurrence of the subsequent forbidden symbols, the forbidden region is actually expanded and theoretically, a better error correction performance can be achieved. Moreover, a generalized stack algorithm is exploited to detect the forbidden symbol beforehand. The proposed approach is combined with the *maximum a posteriori* (MAP) metric to keep the highly probable decoding paths in the stack. Simulation results justify that our scheme performs better than the existing MAP methods on the error correction performance, especially at a low coding rate.

1. Introduction

Traditionally, channel coding is performed after source coding to protect the compressed bit stream sent over a noisy channel. For example, an image file is first compressed by using Discrete Cosine Transformation or arithmetic coding [1–3] with high coding efficiency. Then, the compressed sequence is further protected by Turbo code [4] or Hamming code [5] against channel noise. This traditional separate scheme lacks the cooperation between source and channel coding processes and may not result in the optimal performance. Recent studies have revealed that the joint operation of them leads to some advantages when compared with the traditional separately operated approach [6–9]. As certain implicit redundancy still exists in the bit streams when the encoder cannot ideally decorrelate the source symbols, it can be utilized in the joint scheme to improve the overall error correcting performance. Thus, it is possible for the joint scheme to outperform the separate approach [10].

Early works on joint source-channel coding were devoted to the study of error resilience in variable length codes (VLC). In particular, most of which were focused on the resynchronization ability of Huffman code [7–9]. The corresponding

hard and soft decoding schemes based on maximum likelihood (ML) or MAP metrics are well-studied for a binary symmetric channel (BSC) with additive white Gaussian noise (AWGN). As arithmetic coding (AC) represents a source symbol using a fractional number of bits, it leads to a better compression efficiency and achieves the optimal entropy coding. However, the high compression ratio makes the codeword more sensitive to channel noise and is difficult to be resynchronized. Therefore, there is a growing interest in improving the robustness of AC against channel noise.

In [11], a forbidden symbol introduced by a reduction in the coding interval is adopted to detect the transmission error continuously. These errors can be detected when the forbidden region is visited. This continuous nature in error detection is exploited to improve the overall performance of the communication system [12]. It provides a tradeoff between the extra redundancy and the delay in detecting an error since its occurrence. Instead of the forbidden symbol, the insertion of markers in some particular positions of the input sequence plays the role of synchronization between the encoder and the decoder [13]. The markers which do not appear in the expected positions indicate transmission errors. Three strategies for the selection of the markers were studied

in [13]. A better compression ratio can be achieved using an adaptive [14] or an artificial marker scheme [15]. The adaptive marker scheme selects the most frequent source symbol as the marker symbol while the artificial marker scheme creates an artificial marker with an arbitrary probability. Making use of the error detection capacity of AC, error correction is performed by sequential decoding, which successively removes the erroneous decoding paths. In [16], depth-first and breadth-first decoding algorithms were proposed with binary branching based on a null zone. The decoding paths are discarded due to the error detection capacity of the forbidden symbol. All the decoding paths with the lowest Hamming distance from the received sequence are preserved in a list.

In [17], a MAP criterion based on the context-based AC was proposed with the insertion of synchronization markers, where the symbol clock and the bit clock models were analyzed. The iterative decoding of error resilient AC concatenated with a convolutional code is adopted and its error correcting capability is validated with the transmission of images over an AWGN channel. A novel MAP decoding approach based on the forbidden symbol was proposed in [10], with a high flexibility in adjusting the coding rate. Sequential decoding algorithms, such as stack algorithm and M -algorithm, are adopted and the proposed system outperforms the separate approach based on convolutional codes in terms of error correcting capability. It is serially concatenated with channel codes and iterative decoding is employed to further improve the overall performance [18, 19]. Chaos phenomenon, which generally exists in complex systems [20, 21], is also observed during the iterative decoding procedures. Thus, chaos control techniques can be adopted to further enhance the error correction performance [22–24]. A sequential MAP estimation for CABAC coder was proposed in [25], which employs an improved sequential decoding technique to determine the tradeoff between complexity and efficiency. In [26], a look-ahead technique for AC decoder was proposed to allow quick error detection. Considering the improvement in the implementation efficiency, AC can be modeled as a finite-state machine corresponding to a variable-length trellis code. The trellis code based on AC was proposed in [27, 28], where a list Viterbi decoding algorithm is applied on the corresponding trellis code and a cyclic redundancy check code is employed for detecting small Hamming-distance errors. The free distance of the corresponding AC-based VLC and its theoretical error correction performance were investigated in [29, 30]. Besides that, the practical implementations on this joint source-channel coding scheme were studied in [31, 32] for high coding speed.

The error detecting capability of AC was analyzed in our previous paper [15]. Here we extend our previous work to tackle the problem of error correction in AC. An effective error correction technique utilizing the forbidden symbol is proposed, which predicts the occurrence of the subsequent forbidden symbols. With our approach, the forbidden region is theoretically expanded and so a better error correction performance is achieved. Furthermore, a generalized stack

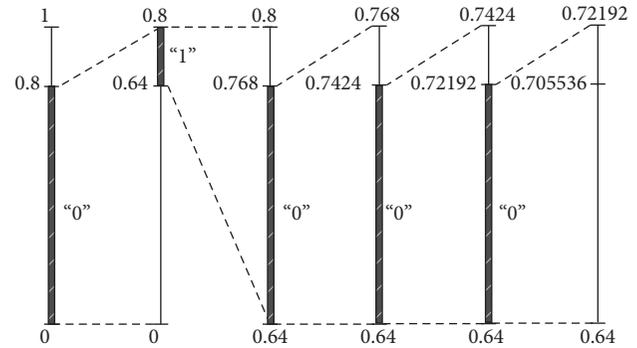


FIGURE 1: The arithmetic coding steps for encoding the sequence “01000.”

algorithm (SA) extending 2^k branches from the best node is also studied for the detection of the forbidden symbol beforehand. The MAP metric [10] is integrated with our approach to preserve the most probable decoding paths in the stack. The idea of our approach was briefly presented in [33], which mainly focuses on the forecasting of the forbidden symbols. Here, the procedures of AC with forecasted forbidden symbols are described in detail. More analyses and simulation results are provided to justify that the proposed scheme outperforms the look-ahead scheme [26] and the original MAP scheme [10] on the error correction performance, especially at a low coding rate.

The rest of this paper is organized as follows. The background of AC is reviewed in Section 2. The proposed scheme is described in Section 3, where the estimation of the subsequent forbidden symbols and the generalized SA are introduced. Simulation results are presented in Section 4 to show the improvement of our scheme. Finally, conclusions are drawn in Section 5.

2. Background of Arithmetic Coding

Arithmetic coding is an iterative operation, which recursively assigns the coding interval to a sequence of source symbols. In general, a prior source model is required, which initializes the coding interval according to the occurrence probabilities of the source symbols. Considering the binary case that the occurrence probabilities of “0” and “1” are correspondingly 0.8 and 0.2, the coding units $[0, 0.8)$ and $[0.8, 1)$ are then assigned to the symbols “0” and “1”, respectively. The arithmetic coding steps for encoding the source sequence “01000” are illustrated in Figure 1.

At last, the final coding unit $[0.64, 0.72192)$ is obtained, within which any real value can be selected and exported as the compressed bits. Theoretically, it is guaranteed to obtain a compressed sequence with the shortest length using $[-\log(0.72192 - 0.64)] = 4$ bits (1011) in this example. In the decoding process, the received codeword sequence (1011) is firstly put after the decimal point to make it 0.1011 which is within the range $[0, 1)$. Then the representation 0.1011 is converted to the decimal value 0.6875. As it falls into the intervals $[0, 0.8]$, $[0.64, 0.8]$, $[0.64, 0.768]$, $[0.64, 0.7424]$,

Function AC_Encoder**Input:** s_k, c, l_k, u_k **Output:** \mathbf{b} Set $l_{k+1} = l_k + (u_k - l_k + 1)c(s_k)$ and $u_{k+1} = l_k + (u_k - l_k + 1)c(s_k + 1) - 1$

While(True)

 If $u_{k+1} < Half$ Set $l_{k+1} = 2 \times l_{k+1}$ and $u_{k+1} = 2 \times u_{k+1} + 1$ Emit a bit 0 and f_{k+1} bits 1 to \mathbf{b} Set $f_{k+1} = 0$ Else If $l_{k+1} \geq Half$ Set $l_{k+1} = 2 \times (l_{k+1} - Half)$ and $u_{k+1} = 2 \times (u_{k+1} - Half) + 1$ Emit a bit 1 and f_{k+1} bits 0 to \mathbf{b} Set $f_{k+1} = 0$ Else If $l_{k+1} \geq First_quarter$ and $u_{k+1} < Third_quarter$ Set $l_{k+1} = 2 \times (l_{k+1} - First_quarter)$ and $u_{k+1} = 2 \times (u_{k+1} - First_quarter) + 1$ Set $f_{k+1} = f_k + 1$

Else

Break;

ALGORITHM 1: Pseudocode of the encoder.

Function AC_Decoder**Input:** $b_k, c, l_k, u_k, v_l_k, v_u_k$ **Output:** \mathbf{s} If $b_k == 0$ Set $v_l_{k+1} = v_l_k$ and $v_u_{k+1} = ((v_u_k - v_l_k + 1)/2) - 1$

Else

 Set $v_l_{k+1} = (v_u_k - v_l_k + 1)/2$ and $v_u_{k+1} = v_u_k$ Set $V = l_k + (u_k - l_k + 1) \times c(1)$

While(True)

 If $V > v_u_{k+1}$ Emit source symbol "0" to \mathbf{s} Set $l_{k+1} = l_k$ and $u_{k+1} = V - 1$ Scale the intervals $[v_l_{k+1}, v_u_{k+1}]$ and $[l_{k+1}, u_{k+1}]$ as done in AC_Encoder Else If $V \leq v_l_{k+1}$ Emit source symbol "1" to \mathbf{s} Set $l_{k+1} = V$ and $u_{k+1} = u_k$ Scale the intervals $[v_l_{k+1}, v_u_{k+1}]$ and $[l_{k+1}, u_{k+1}]$ as done in AC_Encoder

Else

Break;

ALGORITHM 2: Pseudocode of the decoder.

and $[0.64, 0.72192]$, the decoder will sequentially export the symbols "0," "1," "0," "0," and "0." The decoded sequence is exactly the same as the source sequence since AC is a lossless source coding scheme.

A practical problem encountered in the implementation of AC is that the interval will continue to shrink in the iterative encoding steps. Thus, a high precision is needed to represent the very small real numbers encountered in the coding process. A solution to this problem is to use integer representation, where the coding interval can be rescaled when the most significant bits in the representation of the lower and upper bounds are the same. Suppose that the binary source sequence represented by $\mathbf{s} = \{s_1, s_2, \dots, s_N\}$ with $p(s = 0) = p_0$ and $p(s = 1) = p_1$, where $(p_0 + p_1 = 1)$, is encoded as a

variable-length codeword $\mathbf{b} = \{b_1, b_2, \dots, b_L\}$. In Algorithm 1, the pseudocode of the encoder is given, where l_k and u_k are, respectively, the lower and upper bounds for encoding the source symbol s_k . The vector \mathbf{c} represents the cumulative probabilities of the source model with $c(0) = 0$, $c(1) = p_0$ and $c(2) = 1$. The initial lower and upper bounds are set to 0 and $2^P - 1$, respectively, where P is the length of the register for storing the value of the bounds. The number of bits not emitted in the interval rescaling operations is recorded by f_k . The values of *First_quarter*, *Half* and *Third_quarter* are fixed and are set to 2^{P-2} , 2^{P-1} , and $3 \times 2^{P-2}$, respectively.

In Algorithm 2, the pseudocode of the corresponding sequential decoder is listed, which avoids the decoding delay. Once both the bounds, d_l_{k+1} and d_u_{k+1} , of the decoding

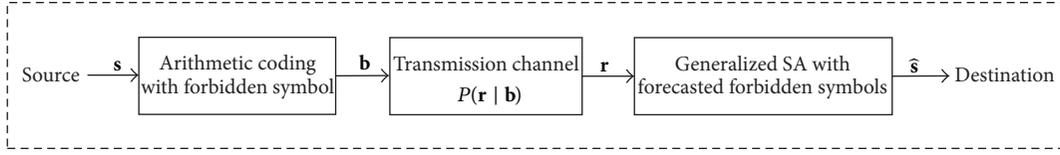


FIGURE 2: A block diagram of the transmission system.

interval for the compressed bit b_k are located in the encoding interval of a particular source symbol, the symbol can be decoded out. The encoding and decoding intervals are then rescaled as performed in the encoder. The lower and upper bounds of the decoding interval are also initialized to 0 and $2^P - 1$, respectively. The details of this kind of AC encoding and decoding can be found in [1, 34].

3. The Proposed Algorithm

Assume that the variable-length codeword $\mathbf{b} = \{b_1, b_2, \dots, b_L\}$ is transmitted over a channel with transition probability $P(\mathbf{r} | \mathbf{b})$. The receiver obtains the demodulated sequence $\mathbf{r} = \{r_1, r_2, \dots, r_L\}$, with which the recovered message $\hat{\mathbf{s}}$ is found using the generalized stack algorithm. A block diagram of this transmission system is depicted in Figure 2.

3.1. MAP Metric. In our scheme, the MAP metric [10] is employed for finding the most probable message $\hat{\mathbf{s}}$ from all possible sequences s , by maximizing the likelihood $P(\mathbf{s} | \mathbf{r})$, as expressed by

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}} P(\mathbf{s} = s | \mathbf{r}). \quad (1)$$

The Bayesian relationship states that

$$P(\mathbf{s} | \mathbf{r}) = \frac{P(\mathbf{r} | \mathbf{s}) P(\mathbf{s})}{P(\mathbf{r})} = \frac{P(\mathbf{r} | \mathbf{b}) P(\mathbf{s})}{P(\mathbf{r})}. \quad (2)$$

In the case of memoryless channels, it is straightforward to represent (2) in an additive form

$$m = \log P(\mathbf{s} | \mathbf{r}) = \log P(\mathbf{r} | \mathbf{b}) + \log P(\mathbf{s}) - \log P(\mathbf{r}). \quad (3)$$

For each bit of \mathbf{r} ,

$$m_k = \log P(r_k | b_k) + \log P(\mathbf{s}_k) - \log P(r_k), \quad k \in [1, L], \quad (4)$$

where the vector \mathbf{s}_k contains the decoded source symbols when the compressed bit b_k is shifted into the decoder. It should be noticed that \mathbf{s}_k can be empty when no source symbol is outputted from the decoder. There are three terms at the right-hand-side of (4). The first term $\log P(r_k | b_k)$ is the channel transition probability while the second term $\log P(\mathbf{s}_k)$ represents the *a priori* probabilities of the source symbols. The first two terms can be evaluated based on the channel and source models, respectively. The last term $\log P(\mathbf{r})$ is complicated, which needs to sum up all the $\log P(\mathbf{r} | \mathbf{b})P(\mathbf{b})$ terms, as follows:

$$\log P(\mathbf{r}) = \log \sum_{\mathbf{b}} P(\mathbf{r} | \mathbf{b}) P(\mathbf{b}). \quad (5)$$

As the full knowledge on the codeword \mathbf{b} with length L is required, it is impractical to evaluate (5) exactly. However, assuming that the codeword \mathbf{b} has equal probabilities of occurrence of “0” and “1,” this term can be approximated by

$$\begin{aligned} P(r_k) &= P(r_k | b_k = 0) P(b_k = 0) + P(r_k | b_k = 1) P(b_k = 1) \\ &= \frac{P(r_k | b_k = 0) + P(r_k | b_k = 1)}{2}. \end{aligned} \quad (6)$$

When hard decoding is adopted in an AWGN channel using binary phase-shift keying (BPSK) modulation with a signal-to-noise ratio (SNR) E_b/N_0 , the channel transition probability is

$$P(r_k | b_k) = \begin{cases} 1 - p, & \text{if } r_k = b_k \\ p, & \text{if } r_k \neq b_k, \end{cases} \quad (7)$$

where $p = (1/2) \operatorname{erfc} \sqrt{E_b/N_0}$. By (6), $P(r_k) = (p + 1 - p)/2 = 1/2$ in this case. Therefore,

$$m_k = \begin{cases} \log(1 - p) + \log P(\mathbf{s}_k) + \log 2, & \text{if } r_k = b_k \\ \log p + \log P(\mathbf{s}_k) + \log 2, & \text{if } r_k \neq b_k. \end{cases} \quad (8)$$

In the soft decoding process, each bit in \mathbf{b} is mapped to \mathbf{t} by $t_k = \sqrt{E_b}(2b_k - 1)$ before transmitted over the AWGN channel. The decoder receives the noisy signal $r_k = t_k + n_k$, where n_k is the additive white noise with standard deviation δ . Given the input sequence \mathbf{b} , the conditional probability of the received signal \mathbf{r} is

$$\begin{aligned} P(r_k | b_k) &= \frac{1}{\sqrt{2\pi}\delta^2} \exp\left(-\frac{(r_k - t_k)^2}{2\delta^2}\right) \\ &= \begin{cases} \frac{1}{\sqrt{2\pi}\delta^2} \exp\left(-\frac{(r_k + \sqrt{E_b})^2}{2\delta^2}\right), & \text{if } b_k = 0 \\ \frac{1}{\sqrt{2\pi}\delta^2} \exp\left(-\frac{(r_k - \sqrt{E_b})^2}{2\delta^2}\right), & \text{if } b_k = 1. \end{cases} \end{aligned} \quad (9)$$

Making use of (6), we have

$$\begin{aligned} P(r_k) &= \left(\frac{1}{\sqrt{2\pi}\delta^2} \exp\left(-\frac{(r_k + \sqrt{E_b})^2}{2\delta^2}\right) \right. \\ &\quad \left. + \frac{1}{\sqrt{2\pi}\delta^2} \exp\left(-\frac{(r_k - \sqrt{E_b})^2}{2\delta^2}\right) \right) \cdot 2^{-1}. \end{aligned} \quad (10)$$



FIGURE 3: Source symbol model with forbidden symbol.

Thus,

$$\begin{aligned} \log P(r_k) &= \log \left(\frac{1}{\sqrt{2\pi\delta^2}} \exp \left(-\frac{(r_k + \sqrt{E_b})^2}{2\delta^2} \right) \right) \\ &\quad \cdot \left(1 + \frac{\exp \left(-(r_k - \sqrt{E_b})^2 / 2\delta^2 \right)}{\exp \left(-(r_k + \sqrt{E_b})^2 / 2\delta^2 \right)} \right) \\ &\quad - \log 2 \\ &= -\frac{(r_k + \sqrt{E_b})^2}{2\delta^2} - \log \sqrt{2\pi\delta^2} \\ &\quad + \log \left(1 + \exp \left(\frac{2\sqrt{E_b}r_k}{\delta^2} \right) \right) - \log 2, \end{aligned} \quad (11)$$

$$m_k = \begin{cases} \log P(s_k) + \log 2 \\ \quad - \log \left[1 + \exp \left(\frac{2\sqrt{E_b}r_k}{\sigma^2} \right) \right], & \text{if } b_k = 0 \\ \log P(s_k) + \log 2 + \frac{2\sqrt{E_b}r_k}{\sigma^2} \\ \quad - \log \left[1 + \exp \left(\frac{2\sqrt{E_b}r_k}{\sigma^2} \right) \right], & \text{if } b_k = 1. \end{cases} \quad (12)$$

3.2. Forecasted Forbidden Symbols. In order to embed error detecting capacity into AC, a forbidden symbol μ with probability of occurrence ε is inserted in the source model, as shown in Figure 3. The probabilities of “0” and “1” are changed to $p_0(1-\varepsilon)$ and $p_1(1-\varepsilon)$, respectively. The overhead of this approach is a lower coding rate as the available coding space for AC shrinks. This accounts for $R_\mu = -\log_2(1-\varepsilon)$ additional bits for each source symbol. The expected length of the compressed sequence is $L = N(H + R_\mu)$ when the forbidden symbol is adopted, where $H = -p_0 \log p_0 - p_1 \log p_1$ is the memoryless source entropy rate. As the forbidden symbol is never encoded, the decoder can assure that some estimated bits are erroneous once it observes the forbidden symbol in the decoding process. Thus, the erroneous decoding path can be pruned. Theoretically, the number of symbols decoded before an error is detected is greater than n at a probability of $(1-\varepsilon)^n$. Therefore, as more source symbols after the erroneous bits are decoded, the error can be detected with a higher probability. Moreover, a large value of ε enables short error detection delay at the expense of compression efficiency.

Thanks to the iterative nature of the AC encoding process, the forbidden symbols after the currently encoded source symbol can actually be estimated beforehand. This is useful in detecting the errors at an earlier stage, so as to prune

the erroneous decoding tree quickly and to increase the chance for the correct decoding tree to remain in the stack. As shown in Figure 4(a), the second forbidden symbols in the original coding regions for “0” and “1” are forecasted, the lengths of which are $p_0(1-\varepsilon)\varepsilon$ and $p_1(1-\varepsilon)\varepsilon$, respectively. Similarly, the third forbidden symbols shown in Figure 4(b) are predicted with the corresponding lengths $p_0p_1(1-\varepsilon)^2\varepsilon$ and $p_1^2(1-\varepsilon)^2\varepsilon$. Theoretically, the lengths of all the successive forecasted forbidden symbols can be summed up as $p_0(1-\varepsilon)\varepsilon \sum_{i=0}^{n-1} p_1^i(1-\varepsilon)^i$ and $\varepsilon \sum_{i=0}^n p_1^i(1-\varepsilon)^i$ in the two coding units, where n is the number of the forecasted forbidden symbols. When n tends to ∞ , we have the length of forecasted forbidden regions $fs(1)$ and $fs(2)$ as follows:

$$\begin{aligned} fs(1) &= p_0(1-\varepsilon)\varepsilon \sum_{i=0}^{\infty} p_1^i(1-\varepsilon)^i = \frac{p_0(1-\varepsilon)\varepsilon}{1-p_1(1-\varepsilon)}, \\ fs(2) &= \varepsilon \sum_{i=0}^{\infty} p_1^i(1-\varepsilon)^i = \frac{\varepsilon}{1-p_1(1-\varepsilon)}. \end{aligned} \quad (13)$$

They are the theoretical limit for the length of successive forbidden symbols. As shown in Figure 4(c), the forecasted forbidden region is much larger than that in Figure 3, which obviously improves the error correcting capability.

On the other hand, the look-ahead technique [26] usually employed in AC decoders can detect forbidden symbol quickly by decoding the source symbol even when the decoding interval bounds $d_{J_{k+1}}$ and $d_{u_{k+1}}$ are located in the encoding intervals for a particular source symbol and the forbidden region by assuming that it is error-free. An example of which is given in Figure 5, where the source symbol “1” is decoded in Figure 5(a) and then the forbidden symbol can be detected in Figure 5(b). Compared with the look-ahead technique in AC decoders, our forecasted forbidden regions can effectively detect the possible errors that may be found by the look-ahead technique. For example, the forbidden symbol in Figure 5(b) can also be detected with our forecasted forbidden regions without the need to decode source symbol “1” in advance. Besides that, the forecasted forbidden regions in the middle of the source symbols “0” and “1” enable the adoption of generalized stack algorithm introduced in Section 3.3. The look-ahead technique can be adopted in our scheme to further enhance the overall correction performance. An example of this scenario is depicted in Figure 5(c) that the source symbol “1” is decoded in advance. The improvement of our scheme is further validated by the simulation results to be reported in Section 4.

With the look-ahead technique adopted in our scheme, it is possible that $d_{J_{k+1}} < 0$ and $d_{u_{k+1}} > 2^P - 1$ in the implementation of AC. The pseudocode of the modified decoder for error detection can be found in Algorithm 3.

3.3. Generalized Stack Algorithm. The generalized SA is a variation of SA, which is a metric first search algorithm. In the original SA, all the explored decoding paths with better metric are stored in an ordered stack with size M . The best decoding path, which has the maximum value given by (8) or

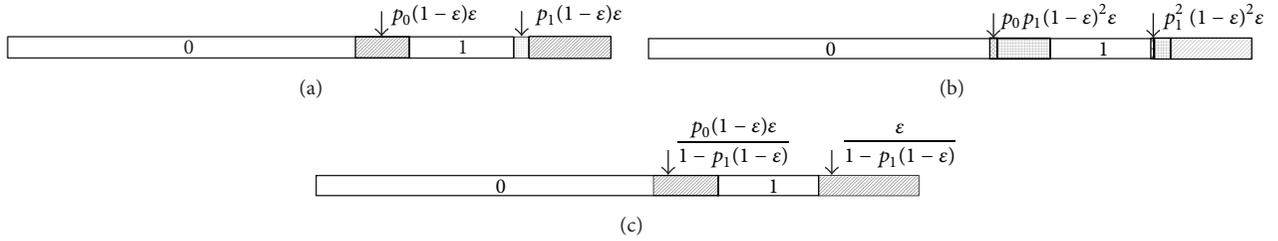


FIGURE 4: (a) The second forecasted forbidden symbol; (b) the third forecasted forbidden symbol; (c) theoretical length of all the successive forecasted forbidden symbols.

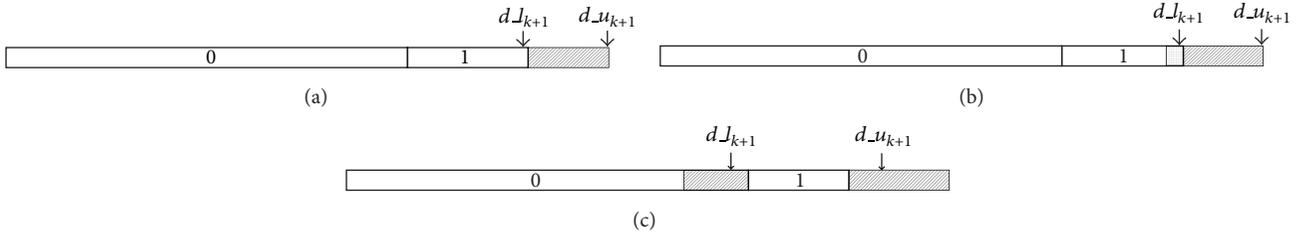


FIGURE 5: (a) Decode source symbol “1” using look-ahead technique; (b) detect forbidden symbol using look-ahead technique; (c) the use of look-ahead technique in our scheme.

(12), is usually stored at the top of the stack. It is extended to two branches after the current decoding node by estimating the subsequent decoding bits as “0” and “1,” respectively. Then the top node is removed and the two child nodes are inserted into the stack. Once the stack is full, the one with the worst metric will be discarded. In the generalized SA, 2^k branches instead of 2 branches are extended from the top node. As the coding region assigned to the forbidden symbol is small, it usually needs more bits to make sure that the decoder will visit the forbidden region or not. Thus, 2^k branches from the best node are able to result in a fast detection and the removal of erroneous decoding paths. This in turn means a higher probability to preserve the correct path in the stack. It is noted that the generalized SA is not applicable in the original MAP algorithm and the look-ahead technique as the underflow problem may happen when 2^k branches are extended from the very small decoding interval. However, as the forbidden regions in the middle of source symbols “0” and “1” are forecasted in our scheme, it guarantees that the decoding interval is not smaller than the length of the intermediate forbidden region. Therefore, the underflow problem can be avoided.

There are three conditions for discarding decoding paths in the generalized SA. The first condition is that the forbidden symbol is encountered. The second corresponds to the situation that the number of decoded symbols is equal to N but the number of decoded bits is smaller than L . The third case is that the number of decoded bits is equal to L but the number of decoded symbol is smaller than N . The generalized SA stops when the L decoded bits can exactly recover N source symbols or the stack is empty. A diagram illustrating the generalized SA is shown in Figure 6, with $k = 2$. Therefore, four child nodes are extended from the best nodes that are

identified with gray color. The one marked with X is deleted as it visits the forbidden region.

4. Simulations

In this section, the proposed scheme is compared with the original MAP scheme [10] and the look-ahead scheme [26]. Binary source symbols with $p_0 = 0.8667$ are randomly generated, which correspond to the memoryless source entropy $H = 0.567$. This entropy is the same as that of the simulation data used in [10]. Each packet consists of 2304 binary symbols. It is then encoded by arithmetic coding with forbidden symbol to generate the variable-length compressed sequence **b**. The packet length and the priori bit probability p_0 are sent to the decoder as side information. They are protected by a high-redundant channel code to guarantee their correctness. Each packet is terminated with an EOS (End of Sequence) symbol having probability 10^{-5} , which protects the last few bits of **b**. The stack size M is chosen as 256 for all the algorithms. The value of k is set to 8 in the generalized SA. All the simulations are run for 10^5 times over an AWGN channel with BPSK modulation. The number of forecasted forbidden symbols is selected as 4. As the original MAP scheme [10] has already been shown to have a better error correction performance than the traditional separated source and channel coding scheme, a comparison with the latter scheme is not repeated here. As the placement of the forbidden symbol can affect the error correction performance [26, 29], two placements are considered in our simulations, which are identified with source models A and B in Figures 7(a) and 7(c). The corresponding forecasted forbidden symbols in source models A and B are illustrated in Figures 7(b) and 7(d). Note that the performance of

```

Function AC_FS_Decoder
Input:  $b_k, c, l_k, u_k, d_{l_k}, d_{u_k}$ 
Output:  $s$ 
If  $b_k == 0$ 
    Set  $d_{l_{k+1}} = d_{l_k}$  and  $d_{u_{k+1}} = ((d_{u_k} - d_{l_k} + 1)/2) - 1$ 
Else
    Set  $d_{l_{k+1}} = (d_{u_k} - d_{l_k} + 1)/2$  and  $d_{u_{k+1}} = d_{u_k}$ 
While(True)
    Find two estimated forbidden regions in the encoding interval;
    If  $d_{l_{k+1}}$  and  $d_{u_{k+1}}$  are completely located in the forecasted forbidden regions
    or out of the encoding interval
    Then delete the decoding path and break;
    Set  $V = l_k + (u_k - l_k + 1) \times c(1)$  and  $V_{fs} = l_k + (u_k - l_k + 1) \times (c(1) - fs(1))$ 
    If  $V > d_{u_{k+1}}$ 
        Emit source symbol "0" to  $s$ 
        Set  $l_{k+1} = l_k$  and  $u_{k+1} = V - 1$ 
        Rescale the intervals  $[d_{l_{k+1}}, d_{u_{k+1}}]$  and  $[l_{k+1}, u_{k+1}]$  as done in AC_Encoder
    Else If  $V_{fs} \leq d_{l_{k+1}}$  and  $d_{u_{k+1}} \leq u_k$ 
        Emit source symbol "1" to  $s$ 
        Set  $l_{k+1} = V$  and  $u_{k+1} = u_k$ 
        Rescale the intervals  $[d_{l_{k+1}}, d_{u_{k+1}}]$  and  $[l_{k+1}, u_{k+1}]$  as done in AC_Encoder
    Else
        Break;
    
```

ALGORITHM 3: Pseudocode of the modified decoder for error detection using forecasted forbidden symbols.

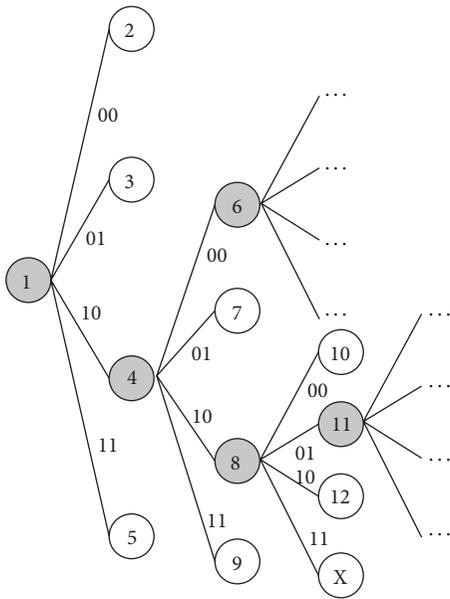


FIGURE 6: A diagram illustrating the generalized SA.

the compared schemes is evaluated by the packet error rate (PER).

The PERs of the proposed, the look-ahead, and the original MAP schemes with source models A and B at various channel SNRs are plotted in Figure 8. The value of ϵ is set to 0.185 in Figure 8, which corresponds to the coding rate of 2/3. Considering source model A, in which many forbidden symbols can be forecasted as indicated by Figure 7(b), it

contributes to the major improvement of our scheme and the look-ahead scheme when compared with the original MAP scheme. The simulation results plotted in Figure 8 validate that the look-ahead scheme performs much better than the original MAP scheme while ours achieves the best results. However, the results obtained with source model B are better than that with source model A in all algorithms. These observations show that the error correcting capacity of source model B is better than that of source model A. Although many forbidden symbols can be forecasted in source model A, it will cause a lot of forbidden symbols assigned in the upper bound of the coding interval and leads to weak error detection in the errors occurring in the lower bound of the coding interval. In summary, our scheme performs much better than the look-ahead scheme and the original MAP scheme at all SNRs for the two source models. Of course, the best results in soft and hard decoding are found by using our scheme with source model B, which achieves a coding gain of around 0.5 dB for hard decoding and 0.25 dB for soft decoding, when compared with the original MAP scheme. Moreover, the values of ϵ at 0.097 and 0.05 are also selected for source model B, which correspond to the coding rate of 4/5 and 8/9, respectively. The PERs of our, the look-ahead, and the original MAP schemes are plotted in Figures 9-10. As indicated in these two figures, the coding gain decreases when the value of ϵ becomes small. The graphs reveal that, for large ϵ , our scheme has a much better performance than the look-ahead and the original MAP schemes. In other words, it is especially effective at a low coding rate.

Figure 11 shows the error correction performance with source model B at various ϵ using soft and hard decoding. In this figure, soft decoding is applied in our, the look-ahead,

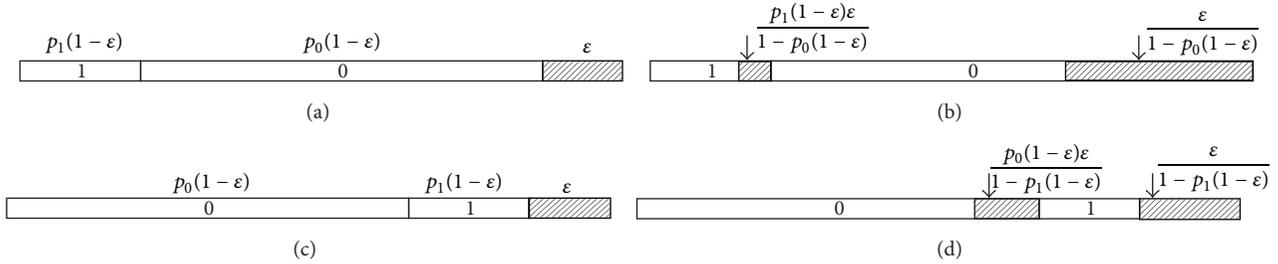


FIGURE 7: (a) Source model A; (b) source model A with forecasted forbidden symbols; (c) source symbol B; (d) source model B with forecasted forbidden symbols.

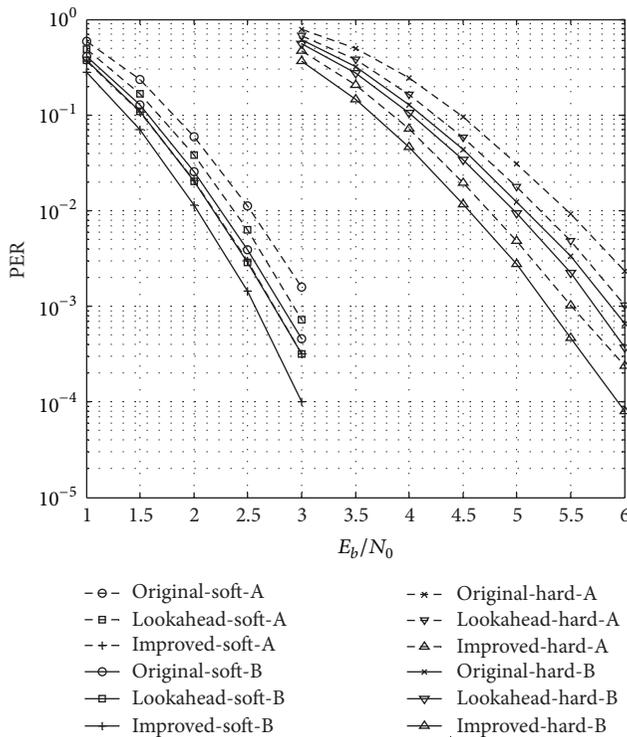


FIGURE 8: Error correction performance of our, the look-ahead, and the original MAP schemes at $\varepsilon = 0.185$.

and the original MAP schemes with the channel SNR fixed at 3.5 dB. The PERs are plotted against the ε value ranging from 0.04 to 0.16. With the increase of ε , the PERs of all schemes drop accordingly. This is reasonable as a large value of ε leads to more redundant bits for error detection, which are helpful in removing the erroneous decoding paths. When ε is large, the gain of our scheme over the look-ahead scheme and the original MAP scheme becomes apparent, which also indicates that our scheme performs much better at a low coding rate. Considering hard decoding in a channel with SNR 5.5 dB, the PERs of our, the look-ahead, and the original MAP schemes are also depicted in Figure 11 with various ε between 0.04 and 0.16. Results similar to those obtained using soft decoding are observed and they further confirm the superiority of our scheme at a low coding rate.

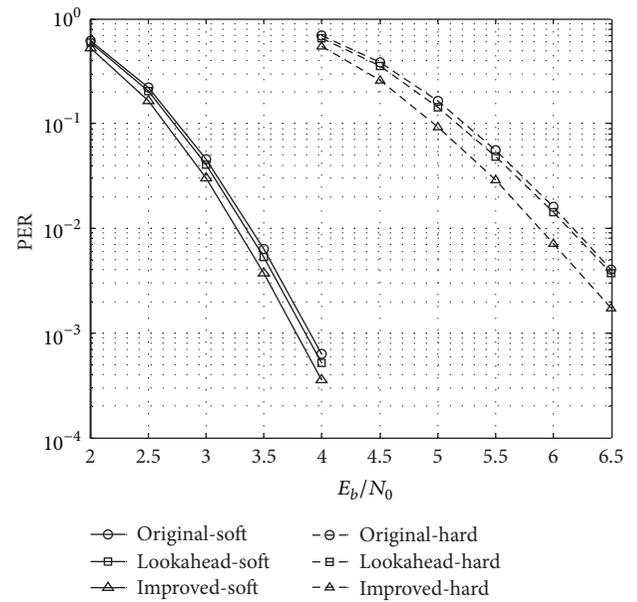


FIGURE 9: Error correction performance of our, the look-ahead, and the original MAP schemes at $\varepsilon = 0.097$.

5. Conclusions

We have proposed an effective error detection technique based on the forecasting of forbidden symbols, which widens the forbidden region by estimating the occurrence of the subsequent forbidden symbols. A generalized SA is also adopted to detect the forbidden symbol beforehand and to remove the erroneous decoding paths earlier. As a result, the chance of preserving the correct decoding path increases and the error correction performance is improved. Simulation results validate the superiority of our approach over the look-ahead and the original MAP schemes, especially at a low coding rate.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

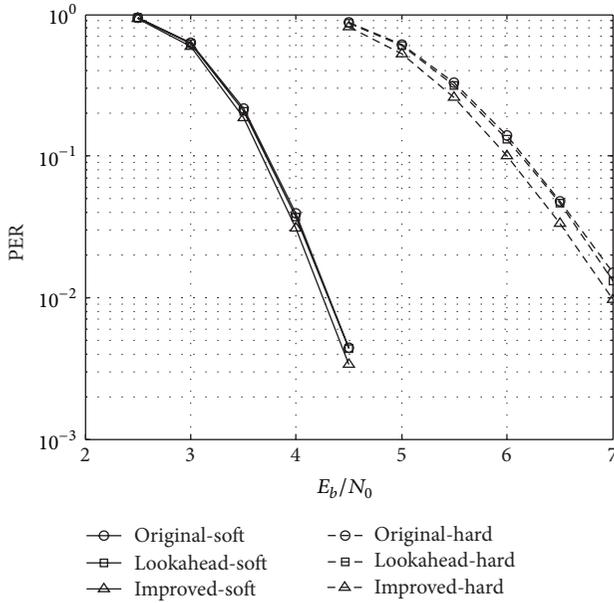


FIGURE 10: Error correction performance of our, the look-ahead, and the original MAP schemes at $\epsilon = 0.05$.

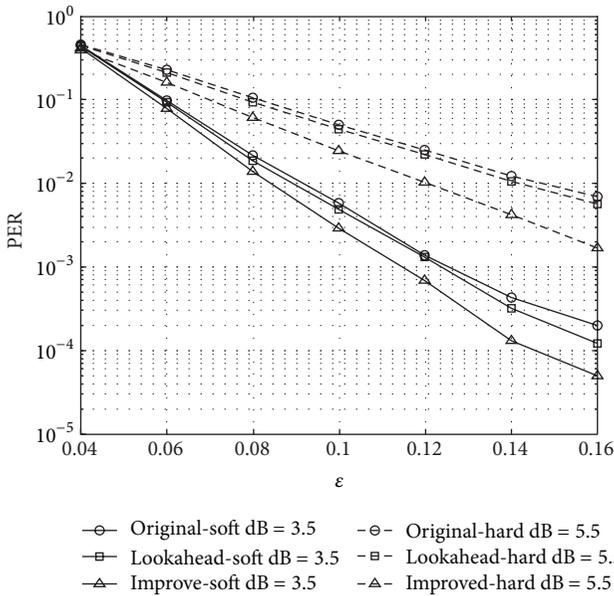


FIGURE 11: The PERs of our, the look-ahead, and the original MAP schemes at various ϵ , using soft decoding at channel SNR 3.5 dB and hard decoding at channel SNR 5.5 dB.

Acknowledgments

The work described in this paper was supported by a grant from CityU (Project no. 7004062), the National Natural Science Foundation of China under the Project Grant nos. 61402291, 61272402, 61070214, 60873264, and 61170283, National High-Technology Research and Development Program (“863” Program) of China under Grand 2013AA01A212, and Ministry of Education in the New Century Excellent

Talents Support Program under Grand NCET-12-0649. Ming Li also thanks the support by the Science and Technology Commission of Shanghai Municipality under research Grant no. 14DZ2260800.

References

- [1] I. H. Witten, R. M. Neal, and J. G. Cleary, “Arithmetic coding for data compression,” *Communications of the ACM*, vol. 30, no. 6, pp. 520–540, 1987.
- [2] H.-C. Hsin, T.-Y. Sung, Y.-S. Shieh, and C. Cattani, “Adaptive binary arithmetic coder-based image feature and segmentation in the compressed domain,” *Mathematical Problems in Engineering*, vol. 2012, Article ID 490840, 14 pages, 2012.
- [3] Y.-K. Chan, R.-C. Chen, P.-Y. Pai, and C.-C. Chang, “Lossless image compression based on multiple-tables arithmetic coding,” *Mathematical Problems in Engineering*, vol. 2009, Article ID 128317, 13 pages, 2009.
- [4] C. Koller, A. G. I. Amat, J. Kliewer, F. Vatta, K. S. Zigangirov, and D. J. Costello, “Analysis and design of tuned turbo codes,” *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4796–4813, 2012.
- [5] A. Sanchez-Macian, P. Reviriego, and J. A. Maestro, “Enhanced detection of double and triple adjacent errors in hamming codes through selective bit placement,” *IEEE Transactions on Device and Materials Reliability*, vol. 12, no. 2, pp. 357–362, 2012.
- [6] O. Y. Bursalioglu, G. Caire, and D. Divsalar, “Joint source-channel coding for deep-space image transmission using rateless codes,” *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3448–3461, 2013.
- [7] S. Chabbouh and C. Lamy, “A structure for fast synchronizing variable-length codes,” *IEEE Communications Letters*, vol. 6, no. 11, pp. 500–502, 2002.
- [8] M. Park and D. J. Miller, “Joint source-channel decoding for variable-length encoded data by exact and approximate MAP sequence estimation,” *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 1–6, 2000.
- [9] K. Sayood, H. H. Otu, and N. Demir, “Joint source/channel coding for variable length codes,” *IEEE Transactions on Communications*, vol. 48, no. 5, pp. 787–794, 2000.
- [10] M. Grangetto, P. Cosman, and G. Olmo, “Joint source/channel coding and MAP decoding of arithmetic codes,” *IEEE Transactions on Communications*, vol. 53, no. 6, pp. 1007–1016, 2005.
- [11] C. Boyd, J. G. Cleary, S. A. Irvine, I. Rinsma-Melchert, and I. H. Witten, “Integrating error detection into arithmetic coding,” *IEEE Transactions on Communications*, vol. 45, no. 1, pp. 1–3, 1997.
- [12] R. Anand, K. Ramchandran, and I. V. Kozintsev, “Continuous error detection (CED) for reliable communication,” *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1540–1549, 2001.
- [13] G. F. Elmasry, “Joint lossless-source and channel coding using automatic repeat request,” *IEEE Transactions on Communications*, vol. 47, no. 7, pp. 953–955, 1999.
- [14] S. Chokchaitam and P. Teekaput, “Embedded error detection in arithmetic coding using markers,” in *Proceedings of the International Conference on Information Technology: Coding Computing (ITCC '04)*, pp. 747–750, April 2004.
- [15] K.-W. Wong, Q. Z. Lin, and J. Y. Chen, “Error detection in arithmetic coding with artificial markers,” *Computers & Mathematics with Applications*, vol. 62, no. 1, pp. 359–366, 2011.

- [16] B. D. Pettijohn, M. W. Hoffman, and K. Sayood, "Joint source/channel coding using arithmetic codes," *IEEE Transactions on Communications*, vol. 49, no. 5, pp. 826–835, 2001.
- [17] T. Guionnet and C. Guillemot, "Soft decoding and synchronization of arithmetic codes: application to image transmission over noisy channels," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1599–1609, 2003.
- [18] M. Grangetto, B. Scanavino, G. Olmo, and S. Benedetto, "Iterative decoding of serially concatenated arithmetic and channel codes with JPEG 2000 applications," *IEEE Transactions on Image Processing*, vol. 16, no. 6, pp. 1557–1567, 2007.
- [19] S. Zaibi, A. Zribi, R. Pyndiah, and N. Aloui, "Joint source/channel iterative arithmetic decoding with JPEG 2000 image transmission application," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, article 114, 2012.
- [20] C. Toma, "Wavelets-computational aspects of stierian realistic approach to uncertainty principle in high energy physics: a transient approach," *Advances in High Energy Physics*, vol. 2013, Article ID 735452, 6 pages, 2013.
- [21] E. G. Bakhoum and C. Toma, "Modeling transitions in complex systems by multiplicative effect of temporal patterns extracted from signal flows," *Mathematical Problems in Engineering*, vol. 2012, Article ID 409856, 11 pages, 2012.
- [22] L. Kocarev, F. Lehmann, G. M. Maggio, B. Scanavino, Z. Tasev, and A. Vardy, "Nonlinear dynamics of iterative decoding systems: analysis and applications," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1366–1384, 2006.
- [23] L. X. Li, H. P. Peng, J. Kurths, Y. X. Yang, and H. J. Schellnhuber, "Chaos-order transition in foraging behavior of ants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 23, pp. 8392–8397, 2014.
- [24] L. Li, J. Xiao, H. Peng, Y. Yang, and Y. Chen, "Improving synchronous ability between complex networks," *Nonlinear Dynamics*, vol. 69, no. 3, pp. 1105–1110, 2012.
- [25] S. B. Jamaa, M. Kieffer, and P. Duhamel, "Improved sequential MAP estimation of CABAC encoded data with objective adjustment of the complexity/efficiency tradeoff," *IEEE Transactions on Communications*, vol. 57, no. 7, pp. 2014–2023, 2009.
- [26] T. Spiteri and V. Buttigieg, "Maximum a posteriori decoding of arithmetic codes in joint source-channel coding," *Communications in Computer and Information Science*, vol. 222, pp. 363–377, 2012.
- [27] D. S. Bi, M. W. Hoffman, and K. Sayood, "State machine interpretation of arithmetic codes for joint source and channel coding," in *Proceedings of the Data Compression Conference (DCC '06)*, pp. 143–152, March 2006.
- [28] H. Moradmand, A. Payandeh, and M. R. Aref, "Joint source-channel coding using finite state integer arithmetic codes," in *Proceedings of the IEEE International Conference on Electro/Information Technology (EIT '09)*, pp. 19–22, Windsor, Canada, June 2009.
- [29] S. Ben-Jamaa, C. Weidmann, and M. Kieffer, "Analytical tools for optimizing the error correction performance of arithmetic codes," *IEEE Transactions on Communications*, vol. 56, no. 9, pp. 1458–1468, 2008.
- [30] A. Diallo, C. Weidmann, and M. Kieffer, "Efficient computation and optimization of the free distance of variable-length finite-state joint source-channel codes," *IEEE Transactions on Communications*, vol. 59, no. 4, pp. 1043–1052, 2011.
- [31] S. Zezza, S. Nooshabadi, and G. Masera, "A 2.63 Mbit/s VLSI implementation of SISO arithmetic decoders for high performance joint source channel codes," *IEEE Transactions on Circuits and Systems. I. Regular Papers*, vol. 60, no. 4, pp. 951–964, 2013.
- [32] W. Zhang, F. Cen, and F. L. Zhu, "Research on practical implementation of binary arithmetic coding with forbidden symbol for error resilience," *Frontiers in Computer Education, Advances in Intelligent and Soft Computing*, vol. 133, pp. 791–798, 2012.
- [33] Q. Z. Lin and K. W. Wong, "Improving the error correction capability of arithmetic coding by forecasting forbidden symbols," in *Proceedings of IEEE International Symposium of Circuits and Systems (ISCAS '13)*, pp. 1139–1142, IEEE, Beijing, China, May 2013.
- [34] A. Said, "Hewlett-Packard Laboratories report," HPL 2004-76, HP Laboratories, Palo Alto, Calif, USA, 2004, <http://www.hpl.hp.com/techreports/2004/HPL-2004-76.pdf>.

Research Article

Topology Identification of Coupling Map Lattice under Sparsity Condition

Jiangni Yu,^{1,2} Lixiang Li,^{1,2} and Yixian Yang^{1,2}

¹Information Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

²National Engineering Laboratory for Disaster Backup and Recovery, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Lixiang Li; li.lixiang2006@163.com

Received 30 July 2014; Accepted 20 September 2014

Academic Editor: Florin Pop

Copyright © 2015 Jiangni Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Coupling map lattice is an efficient mathematical model for studying complex systems. This paper studies the topology identification of coupled map lattice (CML) under the sparsity condition. We convert the identification problem into the problem of solving the underdetermined linear equations. The ℓ_1 norm method is used to solve the underdetermined equations. The requirement of data characters and sampling times are discussed in detail. We find that the high entropy and small coupling coefficient data are suitable for the identification. When the measurement time is more than 2.86 times sparsity, the accuracy of identification can reach an acceptable level. And when the measurement time reaches 4 times sparsity, we can receive a fairly good accuracy.

1. Introduction

The coupled map lattice with nonlocally coupling chaotic characteristic is widely observed and highly involved in many fields, which ranges from complex network [1–3] to neural network, from biological system to ecological system, and from physics to computer science [4–6]. Driven by some practical applications which benefit from the better controlling of CML, a great deal of current research of CML has focused on dynamical analysis, control, and modeling. However, the behavior of CML is largely influenced by the topology of the network which generally is invisible to us. Thus, its identification usually becomes the promise of application.

Up to now, various researches focus on identifying the patterns of coupled map lattice models [7, 8], such as the study of the formation and evolution of spatiotemporal patterns based on a reference model [7], and the identification of CML based on the wavelet [8]. However, few researches have discussed the topology identification of CML. There are two problems of topology identification: how we should get the result with less measurement time (M) and what precondition the measured data need.

In this paper, as CML is discrete, for the facilitation of identification, we transform the CML equation so that the identification problem can be converted into a problem of solving the linear equation $y = \Phi x$. The ℓ_1 method is introduced to solve the topology identification of CML. In the process of identification, the relationship between the measurement time (M) and the sparsity (S) of x is taken into consideration. In the experiment, through the analysis of entropy, we discuss the coupling coefficient and the solvability of the equation $y = \Phi x$. We find that smaller coupling coefficient benefits the identification. Through further research, we study the influence of the measurement time on the identification precision; that is, when the measurement time $M \geq 2.86S$, we can achieve a decent identification result, and when $M \geq 4S$, the identification result is very good.

2. Analysis of Identification Process

Considering a CML with N elements is as follows:

$$z_{t+1}(i) = (1 - \varepsilon) f(z_t(i)) + \frac{\varepsilon}{2W} \sum_{j=1}^W c_{ij} F_j(t, j), \quad (1)$$

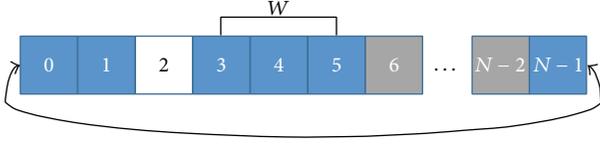


FIGURE 1: Periodic boundary condition of element $i = 2$ when neighbors $W = 3$.

where $z_t(i)$ is the state of each node, $i = 1, 2, \dots, N$, and t is the discrete time step. $\varepsilon \in (0, 1)$ is the strength of the coupling. $f(z) = 4z(1 - z)$ is the standard form of logistic nonlinear function. Also $F_i(t, j) = f(z_t(i - j)) + f(z_t(i + j))$, where $j = 1, 2, \dots, W$. $c_{ij} \in \mathbb{R}^{N \times (2 \times W + 1)}$ is the weighted topology connection from element i to its neighbor elements $i - j$ and $i + j$, and the weighted topology connection obeys the periodic boundary condition. The schematic diagram is shown as in Figure 1.

Normally, the weighted topology connections are composed by the known part K and the unknown part X . Our purpose is to identify this unknown part. In order to identify these unknown connections, the dynamical equation (1) can be transformed as the following equation:

$$\frac{2W}{\varepsilon} (z_{t+1}(i) - (1 - \varepsilon) f(z_t(i))) = \sum_{j=1}^W c_{ij} F_i(t, j). \quad (2)$$

Now, we expand the matrix $c \in \mathbb{R}^{N \times (2 \times W + 1)}$ to $C \in \mathbb{R}^{N \times N}$. In the expanded matrix C , if $j > W$, $C_{ij} = 0$; if $j \leq W$, $C_{ij} = c_{ij}$. Under these circumstances, (2) can be expressed as

$$\frac{2W}{\varepsilon} (z_{t+1}(i) - (1 - \varepsilon) f(z_t(i))) = \sum_{j=1}^N C_{ij} \cdot f(z_t(j)). \quad (3)$$

We set $R(t, i) = (N/\varepsilon)(z_{t+1}(i) - (1 - \varepsilon) f(z_t(i)))$. If we put $t = 1, 2, \dots, M$ iterations together, then we get M measurements. Its formula can be expressed as

$$\begin{pmatrix} R(1, 1) & \cdots & R(1, N) \\ \cdots & & \cdots \\ R(M, 1) & \cdots & R(M, N) \end{pmatrix} = \begin{pmatrix} f_1(1) & \cdots & f_1(N) \\ \cdots & & \cdots \\ f_M(1) & \cdots & f_M(N) \end{pmatrix} * \begin{pmatrix} C_{11} & \cdots & C_{N1} \\ \cdots & & \cdots \\ C_{1N} & \cdots & C_{NN} \end{pmatrix}. \quad (4)$$

Here, C can be divided into two parts, that is, the known part K and unknown part X , which is $C = K + X$. Thus, to get the solution X , we have

$$\begin{pmatrix} R_1(1) & \cdots & R_1(N) \\ \cdots & & \cdots \\ R_M(1) & \cdots & R_M(N) \end{pmatrix} = \begin{pmatrix} f_1(1) & \cdots & f_1(N) \\ \cdots & & \cdots \\ f_M(1) & \cdots & f_M(N) \end{pmatrix} * K \\ = \begin{pmatrix} f_1(1) & \cdots & f_1(N) \\ \cdots & & \cdots \\ f_M(1) & \cdots & f_M(N) \end{pmatrix} * X. \quad (5)$$

To simplify, the above equation set can be represented as

$$Y = \Phi * X, \quad (6)$$

where Y implies the left side of (5) and Φ implies the left matrix of the right side.

Obviously, the elements in Y and Φ are decided by ε , W , and the sampled value z . The size of Y and Φ are decided by M and N . Thus, (6) can be expressed as

$$(y_1, \dots, y_N) = \Phi * (x_1, \dots, x_N), \quad (7)$$

where $x_i \in \mathbb{R}^{N \times 1}$ and $y_i \in \mathbb{R}^{M \times 1}$ are the columns of X and Y , respectively. Without loss of generality, we use $x \in \mathbb{R}^{N \times 1}$ instead of x_i and use $y \in \mathbb{R}^{M \times 1}$ instead of y_i . The equation set finally can be described as

$$y = \Phi * x. \quad (8)$$

Eventually, we transform the topology identification problem into a problem of solving the equation set. The vectors y and Φ can be computed by the sampled value $z_t(i)$ (which is the state value of each element in the t th iteration). The character of solutions x depends on the matrix $\Phi \in \mathbb{R}^{M \times N}$. When there are at most S unknown connections of each element, which means the nonzero value of x will be less than S , in this case we defined x as S -sparse. In the following section, the influences of Φ and S will be taken into consideration in the identification problem.

3. The Analysis of Φ and U

Normally, when Φ is full rank, which means Φ is reversible, (5) has a unique solution, which implies that the measurement time M should be larger than the element N . However, when M is smaller than N , the identification of the unknown connections is an interesting problem. On account of the partial unknown connection, namely, x is S -sparse, the solution can be worked out by an underdetermined equation $y = \Phi x$ when $M < N$. The solving of $y = \Phi x$ is related to two aspects: the construction of Φ and the method to work out the solution x .

3.1. The Solving Condition for Φ . According to [9], the necessary and sufficient condition for the solution of the underdetermined problem is that, for any vector x sharing the same sparsity s , there exists a $\delta_s \in (0, 1)$ such that

$$(1 - \delta_s) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_s) \|x\|_2^2, \quad (9)$$

where $\delta_s \in (0, 1)$ implies that the columns of Φ should be linearly independent. The property is generally satisfying the law of random matrix such as i.i.d Gauss and Bernoulli. Recently, owing to the pseudorandom property of chaotic system, the chaos sequence is used for the construction of Φ . It was proved in [10] that when the sampling distance guarantees a certain value, the logistic system satisfies the RIP condition. In [11], a series of chaos systems, such as Lorenz system and Chuan's circuit, are used to produce Φ . Thus, in the model of this paper, it was promising to identify the unknown connections.

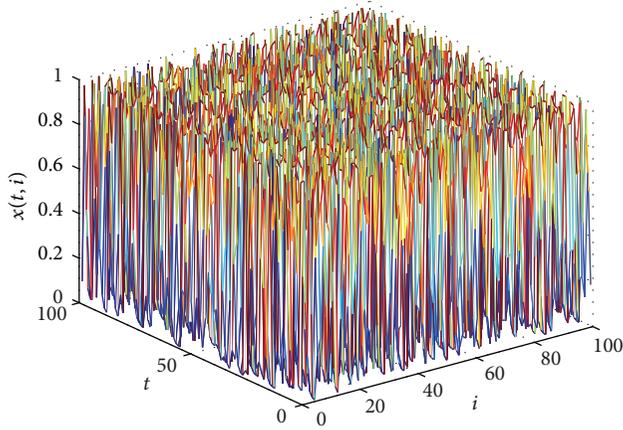


FIGURE 2: The state trajectory of CML.

3.2. The Solving Process of x . As illustrated above, the unknown vector is S -sparse; then the resolution problem can be transformed into an optimization problem which is to find the most sparse solution of (8). Thus the ℓ_0 norm optimization is taken into consideration. However, to find a $x = \operatorname{argmin} \|x\|_0$ to satisfy $y = \Phi x$ is a NP hard problem that requires an exhaustive enumeration of all $C(S, N)$ possible combinations for the locations of the nonzero entries in x . Then the application of ℓ_2 norm optimization is tried, which is to find a $x = \operatorname{argmin} \|x\|_2$ to satisfy $y = \Phi x$. It is also denied for the lack of precision. After that, ℓ_1 norm is considered; it is more efficient than ℓ_0 and more precise than ℓ_2 . The optimization method is expressed as follows:

$$x = \operatorname{argmin} \|x\|_1 \quad \text{s.t. } y = \Phi x. \quad (10)$$

The ℓ_1 optimization problem can be transformed to a linear programming known as basis pursuit (BP). As [12] illustrated, the matrix Φ can be shown to have the RIP, when the measurement time satisfies the condition as follows:

$$M > cS \times \log\left(\frac{N}{S}\right), \quad (11)$$

where c is a constant; we can exactly reconstruct S -sparse vector x via the BP method. Plenty of experimental results show that most of S -sparse signals can be worked out when $M \geq 4S$ [13]. In this paper, we will discuss the relation between M and S in this identification circumstance by experimental analyses.

4. Experiments and Discussions

In the simulation, we set $N = 100$, $W = 40$, $z_0(i) \in [0, 1]$, $i = 1, 2, \dots, N$, $\varepsilon = 0.01$. The chaotic state trajectory of (1) by 100 iterations is shown in Figure 2.

The character of random likeness and boundedness of chaotic system can be observed from Figure 2, and for a better perceptual intuition the state of the 50th element is shown in Figure 3.

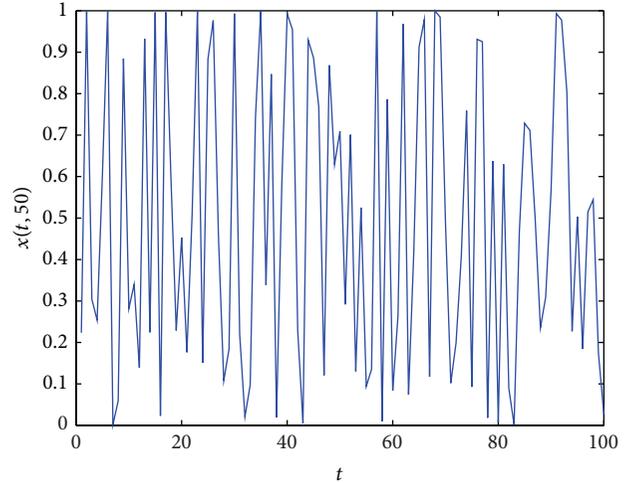
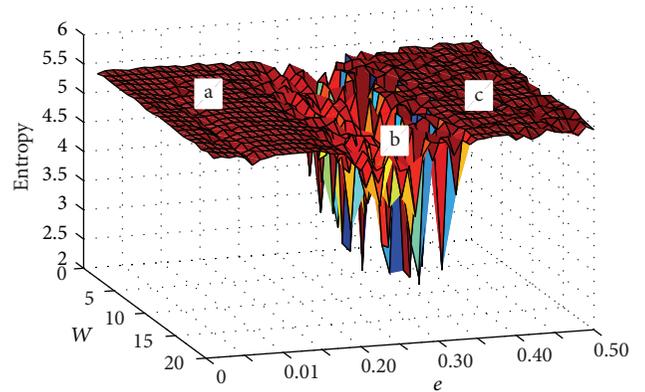


FIGURE 3: The state trajectory of the 50th element of CML.


 FIGURE 4: The diagram of entropy influenced by M and ε .

4.1. The Influence of ε on the Identification. The chaotic behavior is obvious when ε is low, while the system may exist in many states, such as chaos, periodicity, and synchronization. To observe the changing of the network state, we use the information entropy to describe the state as illustrated in [14], by which higher value of entropy implies better chaos character of each node and vice versa. The state of CML is influenced by the parameters ε and W . With the changing of these factors, the information entropy is shown as Figure 4.

Figure 4 shows the entropy with different ε and W ; it is obvious that the value of ε is the main influence factor to the entropy.

- (i) When $\varepsilon < 0.2$, the degree of coupling is low. It shows a high value of entropy, which means the high randomness of the network. CML system is under a nonsynchronous state, and the sampled sequence benefits the identification.
- (ii) When $0.2 < \varepsilon < 0.4$, it shows a low value of entropy, and the CML system is under a complex state between the nonsynchronous state and the synchronous one.

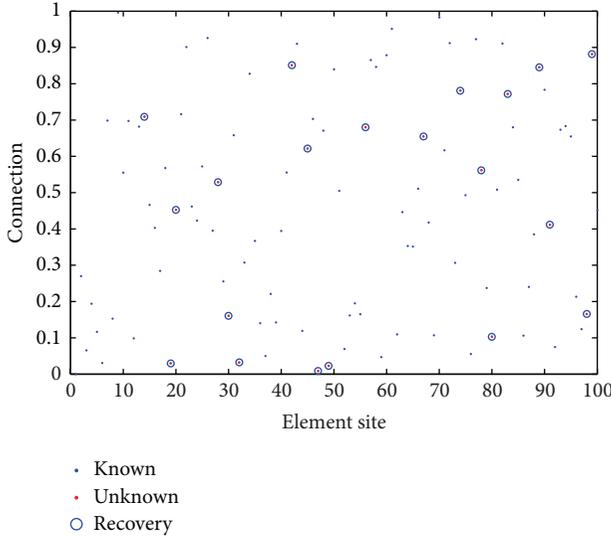


FIGURE 5: The identification result of unknown connections of the 50th element.

Some behavior such as periodicity, which indicates the randomness is weaker, may lead to a low entropy of the sequence. Thus, it is not the favorable condition for identification.

- (iii) When $0.4 < \varepsilon$, it arrives at a high value of entropy as chaos because of the synchronization of CML. Considering the synchronization of elements, the correlation between the columns of Φ is high and does not conform to the RIP condition. Therefore, despite the high entropy, it is not suitable for identification.

Thus, for the requirement of identification, smaller ε is better for the identification.

4.2. Identification Result of Unknown Connections. Here we set $\varepsilon = 0.01$, $N = 100$, $S = 20$, $M = 60$, $W = 50$. Figure 5 shows the connections of an element to the others achieve a successful identification with accuracy $\|x - \hat{x}\| = 1.8153e - 007$.

The optimization process is shown in Figure 6.

From Figure 6, the estimated value \hat{x} converges to the stable state, and the result is correct when $\|x - \hat{x}\|$ is closed to zero. The experiment is implemented in Matlab 2013 on Windows 8 Service Pack 1 64-bit operation system. We use a desktop which has a 4-core Intel(R) Core (TM) i3-2120 processor running at 3.30 GHz and 2 GB of RAM. It can be observed that with 4 iterations the result can be almost accurate, and it only takes 0.073322 seconds in our experimental simulation.

According to the above experiments, the identification process works perfectly when $\varepsilon = 0.01$. But does the algorithm work besides this value? As illustrated in Figure 4, higher coupling coefficient leads to a low entropy or a synchronized state as b and c. The algorithm works well when $\varepsilon < 0.1$; thus the premise for completing the identification

progress smoothly depends on the coupling coefficient of the network. The result just corresponds with the conclusion that smaller ε benefits the identification.

4.3. The Influence of Measurement Time and Sparsity. As we discussed above, to solve the underdetermined equation, the measurement time and the solution sparsity should fulfill a certain condition. For a more detailed understanding of these correlations, the three-dimensional diagram was drawn and is shown in Figure 7.

From Figure 7 we discover that with smaller measurement time and bigger sparsity the solution turns bad and vice versa. To be more detailed, we showed the aerial view in Figure 8.

For Figure 8, we can divide the areas into four parts a, b, c, and d by the distribution of solution effect. The red line $y = 0.35$ means the ratio $M/S = 2.86$ and the green line $y = 0.25$ means the ratio $M/S = 4$.

- (i) For area a, when $M < S$, it does not meet the solution conditions.
- (ii) For area b, when $S < M < 2.86S$, x almost cannot be calculated by Φ accurately.
- (iii) For area c, when $2.86S < M < 4S$, x can be calculated out with acceptable accuracy.
- (iv) For area d, when $M > 4S$, x can be calculated out exactly.

The area d corresponds with the traditional condition $M > 4S$ discussed above. However, from the experiment we learn that the area c is regarded as the acceptable one. Such result may have a guiding significance in the identification process.

5. Conclusion and Future Work

In this paper, we propose a new approach for the identification of CML. We work out the unknown connections vector through the ℓ_1 method. The requirement of sampling time and data characters are discussed in detail in this paper, and the identification result meets the precision of the experiment. Moreover, also some possibility of further researches exists.

On one side, assuming there is an undirected network, besides all the values of the nodes' state, we only know the total sparsity of all the connections (some nodes may have sparse connections, while some nodes' connections may be not sparse); how can we get the correct identification result? This problem may be solved by the following steps. Firstly, we should find the connections between two nodes with sparse unknown connections and estimate them. Secondly, owing to the symmetry characteristic of undirected network, with more and more unknown connections being identified, some of the connections of nodes which were not sparse become sparse; thus we should find these nodes and estimate the connections. Thirdly, continue the second step until we know that remaining nodes' connections cannot be estimated.

On the other side, comparing with traditional method of topology identification, ℓ_1 method can finish the process

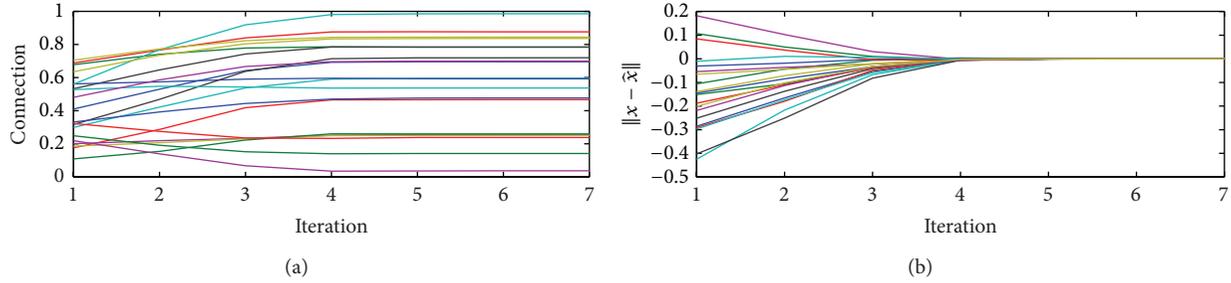


FIGURE 6: The optimization result of each iteration.

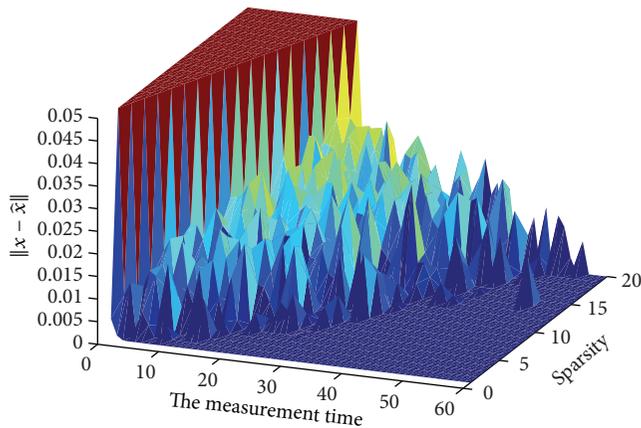


FIGURE 7: The correlations among the measurement time, sparsity, and the identification accuracy.

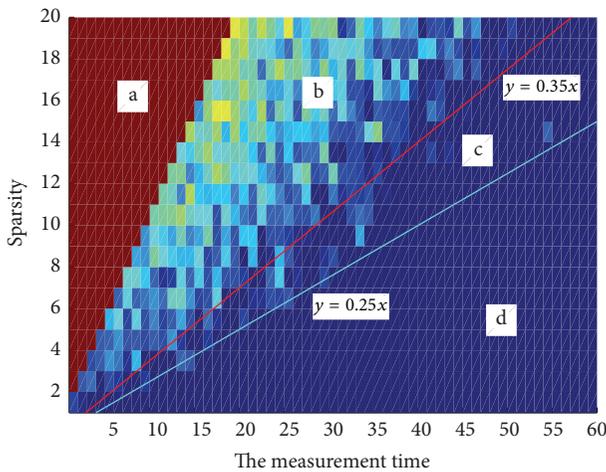


FIGURE 8: The aerial view of Figure 7, which is divided into 4 areas.

of identification with lesser sampling time, but it requires knowing the whole states of the network. Its applications may be limited in the network which are able to know all the states of the nodes. Thus, in the future work, the topology identification with the deficiency of a part of nodes' state value may be discussed.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (Grant nos. 61170269, 61121061), the Beijing Higher Education Young Elite Teacher Project (Grant no. YETP0449), the Asia Foresight Program under NSFC Grant (Grant no. 61161140320), and the Beijing Natural Science Foundation (Grant no. 4142016).

References

- [1] D. Zhao, L. Li, H. Peng, Q. Luo, and Y. Yang, "Multiple routes transmitted epidemics on multiplex networks," *Physics Letters. A*, vol. 378, no. 10, pp. 770–776, 2014.
- [2] D. Zhao, L. Li, S. Li, Y. Huo, and Y. Yang, "Identifying influential spreaders in interconnected networks," *Physica Scripta*, vol. 89, no. 1, Article ID 015203, 2014.
- [3] S. Li, Z. Jia, A. Li, L. Li, X. Liu, and Y. Yang, "Cascading dynamics of heterogenous scale-free networks with recovery mechanism," *Abstract and Applied Analysis*, vol. 2013, Article ID 453689, 13 pages, 2013.
- [4] M. C. Cross and P. C. Hohenberg, "Pattern formation outside of equilibrium," *Reviews of Modern Physics*, vol. 65, no. 3, pp. 851–1112, 1993.
- [5] V. Booth and T. Erneux, "Understanding propagation failure as a slow capture near a limit point," *SIAM Journal on Applied Mathematics*, vol. 55, no. 5, pp. 1372–1389, 1995.
- [6] R. V. Solé, J. Valls, and J. Bascompte, "Spiral waves, chaos and multiple attractors in lattice models of interacting populations," *Physics Letters A*, vol. 166, no. 2, pp. 123–128, 1992.
- [7] D. Coca and S. A. Billings, "Identification of coupled map lattice models of complex spatio-temporal patterns," *Physics Letters. A*, vol. 287, no. 1-2, pp. 65–73, 2001.
- [8] S. A. Billings, L. Z. Guo, and H. L. Wei, "Identification of coupled map lattice models for spatio-temporal patterns using wavelets," *International Journal of Systems Science*, vol. 37, no. 14, pp. 1021–1038, 2006.
- [9] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

- [10] L. Yu, J. P. Barbot, G. Zheng, and H. Sun, "Compressive sensing with chaotic sequence," *IEEE Signal Processing Letters*, vol. 17, no. 8, pp. 731–734, 2010.
- [11] V. Kafedziski and T. Stojanovski, "Compressive sampling with chaotic dynamical systems," in *Proceedings of the 19th Telecommunications Forum (TELFOR '11)*, pp. 695–698, November 2011.
- [12] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–124, 2007.
- [13] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [14] A. M. Hagerstrom, T. E. Murphy, R. Roy, P. Hövel, I. Omelchenko, and E. Schöll, "Experimental observation of chimeras in coupled-map lattices," *Nature Physics*, vol. 8, no. 9, pp. 658–661, 2012.

Research Article

Cryptanalysis and Improvement of the Robust and Blind Watermarking Scheme for Dual Color Image

Hai Nan,¹ Bin Fang,¹ Weibin Yang,² Jiye Qian,¹ Ming Li,³ Yi Liu,⁴ and Yushu Zhang¹

¹College of Computer Science, Chongqing University, Chongqing 400044, China

²College of Automation, Chongqing University, Chongqing 400044, China

³College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China

⁴PetroChina Chongqing Marketing Jiangnan Company, Chongqing 400060, China

Correspondence should be addressed to Bin Fang; fb@cqu.edu.cn

Received 2 July 2014; Accepted 2 September 2014

Academic Editor: Ezzat G. Bakhoum

Copyright © 2015 Hai Nan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With more color images being widely used on the Internet, the research on embedding color watermark image into color host image has been receiving more attention. Recently, Su et al. have proposed a robust and blind watermarking scheme for dual color image, in which the main innovation is the using of two-level DCT. However, it has been demonstrated in this paper that the original scheme in Su's study is not secure and can be attacked by our proposed method. In addition, some errors in the original scheme have been pointed out. Also, an improvement measure is presented to enhance the security of the original watermarking scheme. The proposed method has been confirmed by both theoretical analysis and experimental results.

1. Introduction

With the rapid development of the Internet, a great deal of digital data can be easily accessed, and the copyright of digital content against privacy and malicious manipulation becomes increasingly important. A variety of theories can be used for the security applications [1–3]. Watermarking [4–6] is a widely used technology for ensuring authenticity, copyright violation detection, and proof of ownership or distributorship of the content. The existing watermarking technologies always equip one of the two distinct properties: robustness and fragility. Robustness means that the embedded watermark in the host media cannot be easily modified or removed. Thus the robust watermarking can be used to protect ownership of the digital content [7]. Contrarily, the fragile watermarking is sensitive to any modification of the host media and can be used to authenticate the integrity of the protected content [8].

Watermarking technologies can be classified into two categories: blind detection watermarking technologies and nonblind detection watermarking technologies [9] according to whether it requires the original data when extracting

watermark. The blind detection means to extract watermark without the help of the original host image and the original watermark [10, 11], while the nonblind detection requires the original host image or the original watermark [12, 13]. Generally, the blind detection watermarking technologies are less robust than the nonblind detection ones, but they are more realistic in many applications and have been recently receiving more attention.

Su et al. [14] recently proposed a new robust and blind digital watermarking algorithm for dual color images based on the two-level DCT, which aimed to protect the copyright under JPEG compression attack effectively. In the original scheme, the authors embed extra information into the host image by modifying the selected AC coefficients of the result obtained by two-level DCT. The method is effective; nevertheless, it is not secure due to the fact that the embedded watermark can be replaced while the modification of the host image is imperceptible, and therefore the purpose of copyright protection cannot be achieved. In this paper, we propose a novel attack method to replace the watermark embedded in the protected image. Moreover, an improvement measure is presented to enhance the security of the original scheme.

The rest of the paper is organized as follows. Section 2 presents a brief introduction of Su et al.'s scheme. By correcting the errors and analyzing the weakness of the original scheme, the attack method is proposed in Section 3. The improvement measure is described in Section 4. Section 5 concludes this paper.

2. Review of Su et al.'s Scheme

In Su et al.'s scheme [14], the two-level DCT technology is used to achieve higher concentration of energy; therefore the watermark information can be embedded easily. Before watermark embedding, the color watermark should be first preprocessed to change into the binary sequence form. That is, three components red (R), green (G), and blue (B) of the original watermark are separated out and then are DCT transformed, respectively. After compress coding and random permutation, the final binary sequence to be embedded is obtained.

2.1. Watermark Embedding. The host RGB image should be first transformed into YIQ mode using

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (1)$$

Then, the luminance component Y of the block sized 8×8 is processed by two-level DCT. The obtained 4×4 embedding blocks (upper-left corner) are shown in Figure 1, where the bold characters, that is, nine AC coefficients (C3, C4, C7, C8, C11, C12, C14, C15, and C16), denote the selected positions for watermark embedding.

To embed the obtained binary sequence, that is, the preprocessed color watermark, into the selected positions, the selected coefficients should be modified according to

$$C^* = \begin{cases} A_k, & \text{if } x = 0, \operatorname{argmin} |A_k - C|, \\ B_k, & \text{if } x = 1, \operatorname{argmin} |B_k - C|, \end{cases} \quad (2)$$

where x denotes the watermark bit, C^* is the modified coefficient after embedding x in coefficient C , k is the bit number of the watermark, $A_k = \lfloor |C|/T \rfloor \times k + T/4$, $B_k = \lfloor |C|/T \rfloor \times k + 3 \times T/4$, $\lfloor \cdot \rfloor$ denotes the least nearest integer, and $|\cdot|$ is the absolute value operation. T is chosen to determine the embedding strength and the difference between C^* and C , in the worst condition, is $T/2$.

After modifying the coefficients, the inverse two-level DCT transform and inverse YIQ color transform should be performed to restore the original state of the host image. Thus, the watermark bits are embedded into the frequency domain of the host image based on two-level DCT.

2.2. Watermark Extraction. To extract the watermark from a watermarked host image, one should first acquire the luminance information through YIQ color transform and, then, use two-level DCT to get the modified coefficients C^{**} .

C1	C2	C3	C4
C5	C6	C7	C8
C9	C10	C11	C12
C13	C14	C15	C16

FIGURE 1: The selected positions for embedding watermark in each image block.

Three segments watermarks $x_{i,k}$, $i = 1, 2, 3$, of the original watermark can be extracted using

$$x_{i,k} = \begin{cases} 0, & \text{if } \operatorname{mod}(C^{**}, T) \leq \frac{T}{2} \\ 1, & \text{if } \operatorname{mod}(C^{**}, T) > \frac{T}{2}. \end{cases} \quad (3)$$

Then, according to the majority principle, the watermark bit x_k can be obtained from $x_{i,k}$. By inverse random permutation, decompression technology, and the inverse coding, the color watermark can be recovered.

3. The Cryptanalysis and Attack

3.1. Mistake in the Original Scheme. In the embedding process of the original scheme, there are some mistakes in (2); for example, suppose $C = 9$ and $T = 4$; then $A_k = 2 \times k + 1$ and $B_k = 2 \times k + 3$; according to (2), if the embedded x is 0, then $\min |A_k - C|$ is 0; here $k = 4$ and $C^* = A_k = C$. Similarly, if 1 is embedded, $k = 3$ and $C^* = B_k = C$. Thus, no matter which value is embedded, the modified coefficient C^* is not affected, and the watermark cannot be extracted exactly.

The appropriate approach is to change A_k and B_k to

$$A_k = k \times T + \frac{T}{4}, \quad B_k = k \times T + \frac{3 \times T}{4}, \quad (4)$$

where k is an integer. Then, the modified coefficient C^* is obtained from (2). The following steps demonstrate that the largest difference between C and C^* is $T/2$.

Suppose the embedded $x = 0$.

Let $q = \lfloor (C - T/4)/T \rfloor$; then

$$C - \frac{T}{4} = q \times T + r, \quad \text{where } r \in [0, T). \quad (5)$$

According to the range of r , it can be divided into two cases.

Case 1 ($r \in [0, T/2)$). Let $k = q$; we obtain

$$|(k - q) \times T - r| \in \left[0, \frac{T}{2}\right); \quad (6)$$

that is,

$$|k \times T - (q \times T + r)| \in \left[0, \frac{T}{2}\right). \quad (7)$$

According to (5),

$$\begin{aligned} \left|k \times T - \left(C - \frac{T}{4}\right)\right| &\in \left[0, \frac{T}{2}\right), \\ \left|\left(k \times T + \frac{T}{4}\right) - C\right| &\in \left[0, \frac{T}{2}\right); \end{aligned} \quad (8)$$

then

$$|A_k - C| \in \left[0, \frac{T}{2}\right). \quad (9)$$

Case 2 ($r \in [T/2, T)$). Let $k = q + 1$; we obtain

$$|(k - q) \times T - r| = |T - r| \in \left(0, \frac{T}{2}\right]. \quad (10)$$

According to (6)–(9),

$$|A_k - C| \in \left(0, \frac{T}{2}\right]. \quad (11)$$

Based on the above two cases, we deduce $|A_k - C| \leq T/2$.

Suppose the embedded $x = 1$; similarly, we obtain $|B_k - C| \leq T/2$.

3.2. Weakness of the Original Scheme. According to Kerckhoffs' principle [17], when analyzing an encryption algorithm, an assumption is that the cryptanalyst knows exactly the design and working of the cryptosystem. Namely, cryptanalyst knows everything about the cryptosystem except for the secret keys.

In the original scheme, according to (3) and the known coefficient T , the scrambled watermark bits x_k can be extracted from watermarked image. At the same time, according to (2) and the known coefficient T , any arbitrary bits can be embedded into the watermarked image. So, we can erase or change the original watermark embedded in the image; thus the copyright protection would fail. The details of the attack scheme are introduced in the next section.

3.3. Method to Modify the Watermark. Suppose I_O is the original image and I_A is the watermarked image carrying the watermark W_A . The attacker has a different image, called I_B , carrying another watermark W_B . The attacker aims to replace W_A of I_A with W_B to change the copyright of I_O . The two attack steps are as follows.

Step 1. Follow the watermark extraction process proposed in Section 2 and obtain the permuted watermark bits x_k from image I_B with known T in (3).

Step 2. Embed the permuted watermark bits x_k into image I_A using (2).

Then, the watermark W_A embedded in the original image is replaced with W_B , which is possessed by the attacker.

The fidelity of the attacked image is analyzed below. Let C denote the two-level DCT coefficient of I_O and C_A^* and C_B^* denote the corresponding coefficients of I_A and I_B , respectively. Based on the statements in Section 3.1, $|C_A^* - C| \leq T/2$, $|C_B^* - C_A^*| \leq T/2$, and then $|C_B^* - C| \leq T$, which means the difference of the coefficients between the original image and the attacked image increases, and, in the worst condition, the difference will be double. Since the modification occurred in the high-frequency coefficients of the two-level DCT, the PSNR (peak signal-to-noise ratio) value of the modified image is acceptable, and the invisibility of the watermark can be ensured. The experiments in the next section also verify this proposal.

3.4. Experiments. PSNR is defined in (12), which may be used to evaluate perceptual distortion of the proposed scheme:

$$\text{PSNR} = \frac{\sum_{i=1}^3 \text{PSNR}_i}{3}, \quad (12)$$

where $i = \{1, 2, 3\}$, respectively, denotes the R, G, B components, and PSNR_i presents the PSNR value of i component:

$$\text{PSNR}_i = 10 \lg \frac{M \times N \times \max\{[H(x, y, i)]^2\}}{\sum_{x=1}^M \sum_{y=1}^N [H(x, y, i) - H'(x, y, i)]^2}, \quad (13)$$

where $H(x, y, i)$ and $H'(x, y, i)$ are the pixel values location at (x, y) of i component of the original host image and the watermarked image. Generally, the larger the PSNR value is, the more invisible the watermark is.

Figure 2(a) shows the original watermarked image of size 512×512 . The original watermark of size 64×64 embedded in Figure 2(a) is given in Figure 2(b). The attacker's image whose size is the same as that of the original image is shown in Figure 2(c). Figure 2(d) displays the watermark embedded in the attacker's image using the original scheme with different secret key, the size of which is also $64 * 64$. Figure 2(e) which seems to be the same as Figure 2(a) is the attacked image using the proposed method, and the PSNR value is near to that of Figure 2(a). Figure 2(f) shows the extracted watermark from Figure 2(e). Since Figure 2(f) is totally different from the original watermark that is shown in Figure 2(b), the copyright protection of the original image is invalid.

4. Improvement Measure

The main drawback of Su et al.'s scheme [14] is that one can easily obtain and replace the scrambled watermark from the host image by the known T . Moreover, the scheme is not dependent on the plain image, which would increase the possibility of attacks. In order to overcome the above drawbacks and to improve security, a secret key must be introduced into the coefficients modification step. We propose the improvement measure here.



FIGURE 2: Attack of the original scheme. (a) The original watermarked image and (b) original watermark; (c) the attacker's watermarked image and (d) the watermark in (c); (e) the attacked image of (a); (f) the extracted watermark from (e).

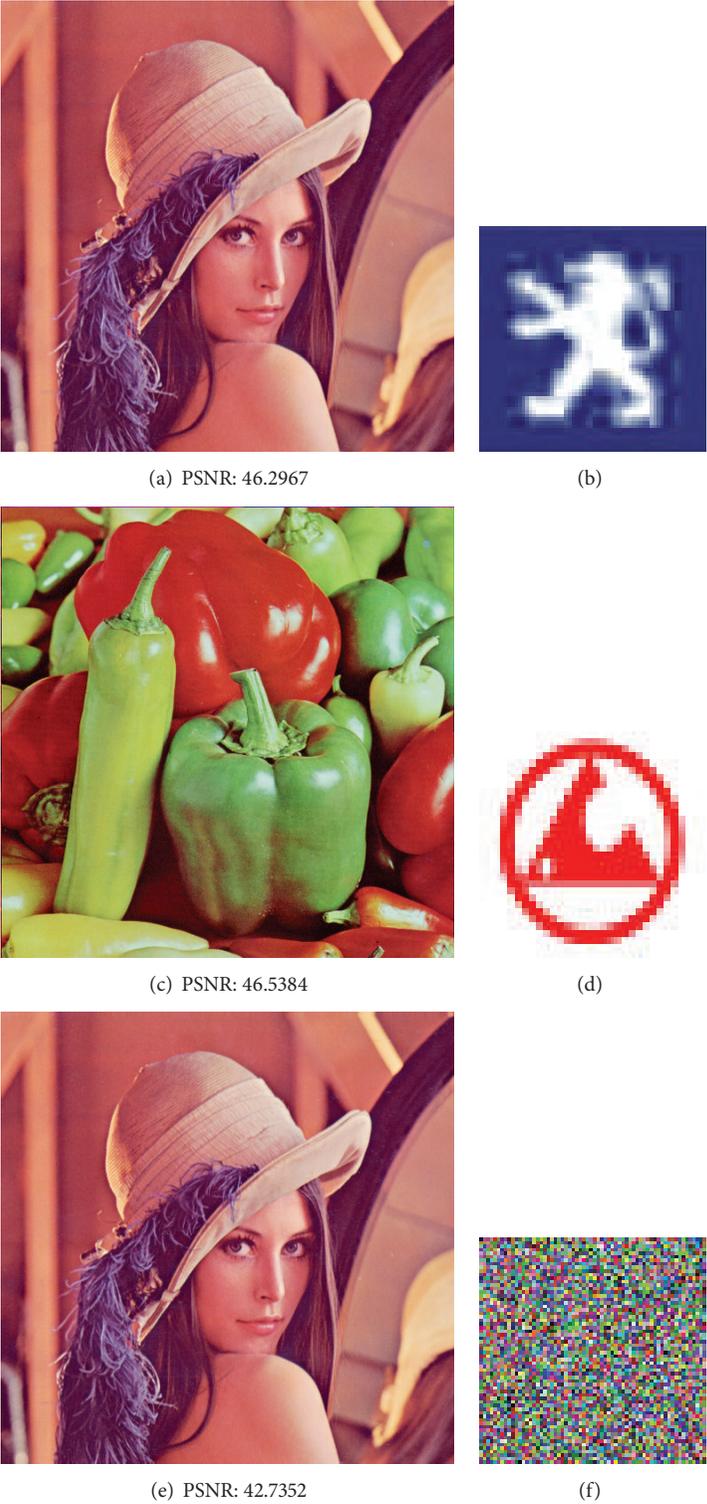


FIGURE 3: Attack of the improved scheme. (a) The original watermarked image and (b) original watermark; (c) the attacker's watermarked image and (d) the watermark in (c); (e) the attacked image of (a); (f) the extracted watermark from (e).

TABLE 1: Performance comparison between the different methods in terms of JPEG compression (CR, compression ratio).

	CR (%)	10	20	30	40	50	60
Reference [15]	PSNR (dB)	35.4	34.8	NA	NA	NA	NA
	NC	1.00	0.29	NA	NA	NA	NA
Reference [16]	PSNR (dB)	35.4	34.8	32.6	31.1	NA	NA
	NC	0.99	0.50	0.12	0.07	NA	NA
Original	PSNR (dB)	33.6428	32.2108	31.8814	31.2290	30.8597	30.4745
	NC	1.0000	1.0000	1.0000	0.9996	0.9792	0.9769
Improved	PSNR (dB)	33.6429	32.2107	31.8816	31.2288	30.8597	30.4744
	NC	1.0000	1.0000	1.0000	0.9995	0.9792	0.9768

4.1. Watermark Embedding

Step 1. The host RGB image should be transformed into YIQ mode using (1) in Section 2.

Step 2. The luminance component Y of the block sized 8×8 is processed by two-level DCT. The obtained 4×4 embedding blocks (upper-left corner) are shown in Figure 1, where the bold characters, that is, nine AC coefficients (**C3**, **C4**, **C7**, **C8**, **C11**, **C12**, **C14**, **C15**, and **C16**), denote the selected positions for watermark embedding.

Step 3. Set the initial parameter α and initial value q_0 , which can be considered as the watermarking key, to iterate the chaotic system (14) N_W times, where N_W is the size of scrambled watermark bits:

$$\begin{aligned} q_{i+1} &= \alpha q_i (1 - q_i), \\ i &= 0, 1, \dots, N_W - 1, \quad q_i \in (0, 1). \end{aligned} \quad (14)$$

With parameter $\alpha \in (3.5699456, 4]$, the system (14) is in chaotic state [18].

Step 4. Let x be the watermark bit and q be the corresponding pseudorandom numbers obtained from (14). For every x , obtain the integral number Z according to

$$Z = \text{floor}(q \times 10^{14}) \bmod 7 + 1, \quad (15)$$

where $\text{floor}(a)$ returns the largest integer smaller than a and Z is a series of integers ranging from 1 to 7. Then, the remnant AC coefficients (C1, C2, C5, C6, C9, C10, and C13) in each block based on two-level DCT are selected according to the values of Z to obtain the C' .

The selected coefficient C' is dependent on the plain host image, so different host image generates different coefficient C' , and any modification on the watermarked image will affect the coefficient C' and then will affect the watermark consequently (the explanations are given below), even if the watermark is obtained before.

Step 5. Obtain a bit x' from the selected AC coefficient C' using

$$x' = \begin{cases} 0, & \text{if } \text{mod}(C', T) \leq \frac{T}{2}, \\ 1, & \text{if } \text{mod}(C', T) > \frac{T}{2}. \end{cases} \quad (16)$$

Step 6. The coefficient C^* after watermark embedding can be obtained by the rule stated in

$$C^* = \begin{cases} A_k, & \text{if } x \oplus x' = 0, \text{ argmin } |A_k - C| \\ B_k, & \text{if } x \oplus x' = 1, \text{ argmin } |B_k - C|, \end{cases} \quad (17)$$

where $A_k = k \times T + T/4$, $B_k = k \times T + 3 \times T/4$, k is an integer, and $|a|$ is the absolute value of a . Based on (7), C^* can be easily computed and the difference between C^* and C is very small. In the worst condition, the largest difference between C and C^* is $T/2$.

Step 7. After modifying the coefficients, the inverse two-level DCT transform and inverse YIQ color transform should be performed to restore the original state of the host image. Thus, the watermark bits are embedded into the frequency domain of the host image based on two-level DCT.

4.2. Watermark Extraction

Step 1. Process the watermarked image I_W by the way of Steps 1–5 in the watermark embedding process to get a series of x' using (16).

Step 2. Extract the watermark according to the DCT coefficients based on the following rules. In (18), C^{**} is the DCT coefficient at the low-frequency position of watermarked image and $x_{i,k}$, $i = 1, 2, 3$, are three segments watermarks of the original watermark:

$$x_{i,k} = \begin{cases} x', & \text{if } \text{mod}(C^{**}, T) \leq \frac{T}{2} \\ \overline{x'}, & \text{if } \text{mod}(C^{**}, T) > \frac{T}{2}, \end{cases} \quad (18)$$

where $\overline{x'}$ is the flipped x' .

TABLE 2: Evaluation of the robustness of the watermarking scheme.

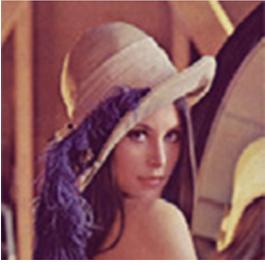
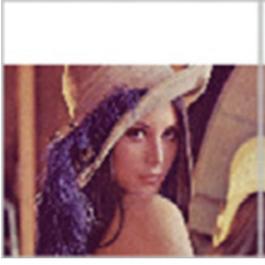
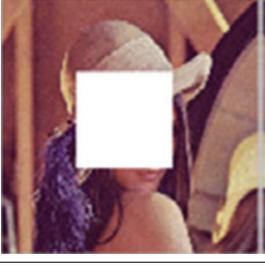
Attacks	Attacked image	Improved scheme		NC	Original scheme	
		PSNR (dB)	Extracted watermark		PSNR (dB)	NC
Brightness + 2		33.1662		0.9999	33.1664	1.0000
Sharpening		33.4241		0.983	33.424	0.9831
Mosaic 3 * 3		30.2472		0.9518	30.247	0.9519
Median filtering		28.8052		0.9093	28.8053	0.9095
Cropping 1		11.5489		0.9999	11.5491	0.9999
Cropping 2		14.0279		0.937	14.0276	0.9369

TABLE 2: Continued.

Attacks	Attacked image	Improved scheme		Original scheme		
		PSNR (dB)	Extracted watermark	NC	PSNR (dB)	NC
Rotating + 5°		14.683		0.7691	14.681	0.7687
Low pass filtering		28.4824		0.9991	28.4825	0.9991
Adding noise 4%		28.083		0.9971	28.0835	0.9971
JPEG 50%		30.8597		0.9791	30.8597	0.9792

Then, according to the majority principle, the watermark bit x_k can be obtained from $x_{i,k}$. By inverse random permutation, decompression technology, and the inverse coding, the color watermark can be recovered.

4.3. Experiments. Since the selected coefficient from the seven remnant AC coefficients (C1, C2, C5, C6, C9, C10, and C13) in each block based on two-level DCT is uncertain, the attacker cannot deduce x' without the secret key mentioned in the above section; therefore the embedded watermark cannot be replaced using the proposed attack method.

In addition, the modifications of the host image during the watermark embedding procedure of the improved scheme are essentially the same as those of the original scheme; that is, only nine AC coefficients (C3, C4, C7, C8,

C11, C12, C14, C15, and C16) in each block are selected to embed the watermark, and the largest difference between the coefficients before and after modification is $T/2$. Therefore, the performance of the improved scheme is unchanged. Figure 3 shows an example of the attack to the improved scheme. It can be seen that the PSNRs of Figures 3(a)–3(d) are similar to those of Figures 2(a)–2(d), respectively. Figure 3(e) is the attacked image using the improved method. Figure 3(f) shows the extracted watermark from Figure 3(e). Since the two keys, that is, the initial value q_0 and α of the chaotic map (14), are different between the original watermarked image and the attacker's watermarked image, the extracted watermark from the attacked image is confused, and the copyright protection of the original image will be successful.

In order to measure the robustness of the watermark, we use the normalized correlation (NC) between the original

watermark W and the extracted watermark W' , which is shown as follows:

$$\begin{aligned} \text{NC} &= \sum_{i=1}^3 \sum_{x=1}^P \sum_{y=1}^Q (W(x, y, i) \times W'(x, y, i)) \\ &\times \left(\sqrt{\sum_{i=1}^3 \sum_{x=1}^P \sum_{y=1}^Q [W(x, y, i)]^2} \right. \\ &\left. \times \sqrt{\sum_{i=1}^3 \sum_{x=1}^P \sum_{y=1}^Q [W'(x, y, i)]^2} \right)^{-1}. \end{aligned} \quad (19)$$

JPEG compression attack is one of the common attacks that must be verified in watermarking algorithm. In this experiment, the different watermarked images are lossy-compressed with different compression factors ranging from 10 to 100. To prove the robustness of the improved scheme in terms of JPEG compression, we compare our scheme with the earlier works [15, 16] that are based on one-level DCT and also with the original scheme of Su et al. The performance comparison is given in Table 1. Note that “NA” means the corresponding experimental datum is unavailable. As shown in Table 1, the watermark embedded by the proposed scheme can be almost fully extracted from the watermarked image when the compression ratio is 60%. However, under the same conditions, the watermarks embedded by [15, 16] cannot survive after attacked by JPEG compression with compression ratios 20% and 30%, respectively. It demonstrates that the improved and the original scheme have almost the same strong robustness against JPEG compression.

In addition to quantitative measurement, Table 2 gives the evaluation of the robustness of the watermarking scheme in terms of NC [19], in which the watermark is extracted from the watermarked image (as shown in Figure 3(a)) which is under different attacks such as JPEG compression, filtering, and noise addition. The experimental results show that the robustness of the improved watermark scheme keeps consistent with that of the original scheme.

5. Conclusions

This paper attacks Su et al.’s scheme applied for protecting the copyright of digital images. The watermark embedded in the protected image using the original scheme can be replaced by the attacker with another watermark for the purpose of copyright changing, while the modification of the original image is imperceptible. To overcome this defect, the chaotic map is introduced, and the performance of the improvement measure is the same as before. Analysis and experiment show that the proposed method is secure and effective.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Key Basic Research Program of China (973 Program 2013CB329103 of 2013CB329100), the National Natural Science Foundations of China (NSFC-61173129), the Specialized Research Fund for the Doctoral Program of Higher Education of China (no. 20120191110026), the Fundamental Research Funds for the Central University (CDJXS1182240 and CDJXS1180004), China Postdoctoral Science Foundation (2014M550456), and Chongqing Postdoctoral Special Funding Project (Xm2014087).

References

- [1] S. H. Huddleston and D. E. Brown, “Using discrete event simulation to evaluate time series forecasting methods for security applications,” in *Proceedings of the Winter Simulation Conference (WSC '13)*, pp. 2772–2783, Washington, DC, USA, December 2013.
- [2] M. Li, “Fractal time series—a tutorial review,” *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [3] M. Li, W.-S. Chen, and L. Han, “Correlation matching method for the weak stationarity test of LRD traffic,” *Telecommunication Systems*, vol. 43, no. 3-4, pp. 181–195, 2010.
- [4] F. Y. Shih and S. Y. T. Wu, “Combinational image watermarking in the spatial and frequency domains,” *Pattern Recognition*, vol. 36, no. 4, pp. 969–975, 2002.
- [5] X. Tong, Y. Liu, M. Zhang, and Y. Chen, “A novel chaos-based fragile watermarking for image tampering detection and self-recovery,” *Signal Processing: Image Communication*, vol. 28, no. 3, pp. 301–308, 2013.
- [6] L. Luo, Z. Chen, M. Chen, X. Zeng, and Z. Xiong, “Reversible image watermarking using interpolation technique,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 187–193, 2010.
- [7] H. S. Malvar and D. A. Florêncio, “Improved spread spectrum: a new modulation technique for robust watermarking,” *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 898–905, 2003.
- [8] P. W. Wong and N. Memon, “Secret and public key image watermarking schemes for image authentication and ownership verification,” *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1593–1601, 2001.
- [9] H. Luo, F.-X. Yu, Z.-L. Huang, and Z.-M. Lu, “Blind image watermarking based on discrete fractional random transform and subsampling,” *Optik*, vol. 122, no. 4, pp. 311–316, 2011.
- [10] X. Y. Luo, D. S. Wang, P. Wang, and F. L. Liu, “A review on blind detection for image steganography,” *Signal Processing*, vol. 88, no. 9, pp. 2138–2157, 2008.
- [11] C. G. Thorat and B. D. Jadhav, “A blind digital watermark technique for color image based on integer wavelet transform and SIFT,” *Procedia Computer Science*, vol. 2, pp. 236–241, 2010.
- [12] T. K. Tsui, X.-P. Zhang, and D. Androutsos, “Color image watermarking using multidimensional fourier transforms,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 16–28, 2008.
- [13] S. Liao, “Dual color images watermarking algorithm based on symmetric balanced multiwavelet,” in *Proceedings of the International Symposium on Intelligent Information Technology*

Application Workshop (IITA '08), pp. 439–442, Shanghai, China, December 2008.

- [14] Q. Su, Y. Niu, X. Liu, and T. Yao, “A novel blind digital watermarking algorithm for embedding color image into color image,” *Optik*, vol. 124, no. 18, pp. 3254–3259, 2013.
- [15] S. D. Lin and C.-F. Chen, “A robust dct-based watermarking for copyright protection,” *IEEE Transactions on Consumer Electronics*, vol. 46, no. 3, pp. 415–421, 2000.
- [16] S. D. Lin, S. C. Shie, and C. F. Chen, “A DCT-based image watermarking with threshold embedding,” *International Journal of Computers and Applications*, vol. 25, no. 2, pp. 130–135, 2003.
- [17] A. Kerckhoffs, “La cryptographie militaire,” *Journal Des Sciences Militaires*, vol. 4, pp. 161–191, 1883.
- [18] K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos*, Springer, Berlin, Germany, 1997.
- [19] T.-C. Hsu, W.-S. Hsieh, J. Y. Chiang, and T.-S. Su, “New watermark-removal method based on Eigen-image energy,” *IET Information Security*, vol. 5, no. 1, pp. 43–50, 2011.

Research Article

Alternating Coordinate-Momentum Representation for Quantum States Based on Bopp Operators for Modelling Long-Distance Coherence Aspects

Ezzat G. Bakhom¹ and Cristian Toma²

¹Department of Electrical and Computer Engineering, University of West Florida, 11000 University Parkway, Pensacola, FL 32514, USA

²Faculty of Applied Sciences, Politehnica University, 315 Splaiul Independentei, 060042 Bucharest, Romania

Correspondence should be addressed to Cristian Toma; cgtoma@physics.pub.ro

Received 6 August 2014; Accepted 22 August 2014

Academic Editor: Florin Pop

Copyright © 2015 E. G. Bakhom and C. Toma. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study presents an alternating coordinate-momentum representation for propagation and transition of associated wave function, based on Bopp operators and on a certain symbolic determinant corresponding to a set of two linear equations with null free terms. It is shown that this alternating representation can justify in a good manner the patterns created through reflection/refraction of waves on nonperfectly smooth interfaces and phase correspondence of diffracted beams without the need of supplementary support functions. Correlations with Lorentz transformation of wave functions by interaction with a certain material medium (the space-time origin of a wave-train being adjusted) are also presented, and supplementary aspects regarding the use of electromagnetic scalar and vector potentials for modelling evolution within this alternating representation are added.

1. Introduction

As it was shown in [1], the study of light wave propagation phenomena at the interface between two different media (reflection/refraction) based on wavefronts generated by multiple centers of reflection/refraction situated on this interface requires an explanation regarding further propagation along directions normal to these wavefronts. For an oblique incidence of a plane wave, a certain center of reflection/refraction will be the first one which emits wavelets with the of speed of light specific to that material medium. Until it interacts with the wavefront generated by another center of reflection/refraction, we should consider that parts of the incident wave are radiated along all spatial directions. The procedure can continue by analyzing the interaction of each newly created wavefront with subsequent wavelets, implying that each time a supplementary part of the received wave will be transmitted on spatial directions which differ to the main axis of reflection/refraction.

Moreover, the assumption regarding the constant phase shift (π for electric field E , e.g.) for reflected/refracted wave in any surface point is also questionable, since the interface

is far from being perfectly smooth and perfectly conductive. A certain transient time interval for creating the electrostatic equilibrium is always required, so as the electric field E to vanish on this surface. As a consequence, local phase shifts for reflected/refracted wave can not be avoided within this local mathematical model. According to the linear wave equation (with constant coefficients), these could generate multiple in-phase local waves propagating along spatial directions which differ for the main reflection/refraction axis. Thus parts of reflected/refracted wave would be transmitted in a large solid angle and directionality would be lost within a very short length interval.

According to solid state theory, a very good argument regarding the perfectly smooth approximation for reflection/refraction theory and for the constant phase shift consists in the fact that photons usually interact with collectivised electrons of the solid crystalline lattice before being reemitted. These electrons could be considered as moving tangent to the interface since sudden changes of trajectory could generate significant electromagnetic field (accelerations being involved). This picture is supported also by quantum physics, since the associated-wave function for

the collectivised electrons is represented in position for large space intervals, the influence of local nonuniformities being decreased. Thus can be drawn a tangent line with a great radius of curvature, resulting in a better directionality for reflected/refracted beams and the same phase shift for them.

However, this possible explanation does not take into account the constant phase shift between the incident and the reemitted beams from points situated far apart from each other, when the associated-wave function for collectivised electrons can not be considered as being the same anymore. There is no perfect crystal lattice, so the quantum functions corresponding to collectivised electrons or to vibrations (phonons) are spatially limited. Moreover, when supplementary diffraction phenomena are involved; some points becoming secondary wave-sources (reemitting light beams) are situated on nonadjacent surfaces (see the edges of a diffraction grating) and they can not be correlated by any surface quantum wave functions.

In [1], an internal report published later in scientific journals in 2013 suggested that a kind of support waves is generated on this interface. They act upon reflecting/refracting points and correlate the phase of reemitted waves by interface points situated at great distance. These support waves could also correlate nonadjacent space intervals wherefrom secondary light beams are emitted through diffraction. So as to be effective, this possible intuitive and computational model requires a high speed for these propagating support waves in order to regroup the wavefronts into light beams with certain directionality within an extremely short time interval. The use of momentum space representation for the light wave is also recommended for describing phase correlation and coherence phenomena.

The wave-vector (momentum) values generated through these support functions represent the base for wave propagation on next time intervals. Based on wave equation

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = \frac{1}{v^2} \frac{\partial^2 \phi}{\partial t^2} \quad (1)$$

a set of possible high speed propagating functions as

$$\phi(x, y, z, t) = ax + by + cx + dt + e, \quad (2)$$

where quantities a, b, c, d, e are constant values, were proposed as support functions. This corresponds with a time-dependent plane equation, a plane which moves in the three-dimensional space. The value of ϕ at space-time origin (when $x = y = z = t = 0$) can be encountered for $t > 0$ on a set of planes situated at distance

$$D = \left| \frac{0a + 0b + 0c + (dt)}{\sqrt{a^2 + b^2 + c^2}} \right| = \left| \frac{dt}{\sqrt{a^2 + b^2 + c^2}} \right| \quad (3)$$

considered from space origin (according to basic analytical geometry). As a consequence, the plane on which ϕ equals e (passing through origin at zero time moment) moves in space with velocity

$$w = \frac{D}{t} = \left| \frac{d}{\sqrt{a^2 + b^2 + c^2}} \right|. \quad (4)$$

There being no restriction regarding the choice for the set constant values a, b, c , and d results in that (unlike standard solutions of wave equation) these plane equations can be represented as constant values propagating in space-time with any velocity (the speed w is not correlated with the wave equation anymore).

However, the arbitrary choice of these constants is not consistent with basic laws and principles of both classical and quantum theory of light waves (being similar to the attempt of defining some hidden variables). An attempt for justifying different actions connected to coordinate and momentum representations by presenting coordinate and momentum operators as acting in different spaces (reciprocal Fourier spaces) has been first performed in [2], with the requirement of considering light speed c as limit speed for transmitting interaction at quantum level, for variables corresponding with the *same* quantum wave function (later this attempt was defined as the Sterian realistic approach, being connected to Schwinger principle in extended studies published in 2013). Yet this implies the use of a supplementary *inner* wave for transmitting information inside a certain wave (the light wave which is analysed). This is the same as using hidden variables (in this case hidden waves).

However, in [3] (an internal report published later in 2013) it was shown that uncertainty principle in quantum physics allows a kind of instantaneous propagation of interaction. The analysis is based on creation/annihilation phenomena. The annihilation of the particle implies the instant annihilation of all parts of this wave-train, irrespective to the distance to this interaction point. It is true that this seems to contradict the relativity postulates (no speed can surpass the light speed) but in fact any speed can be noticed just by an emission and a reception of a certain signal or particle (not just by a single emission or a single annihilation). Since the particle ceases to exist, its final momentum is zero. There being no uncertainty, it results that

$$\Delta p_x = \Delta p_y = \Delta p_z = 0 \quad (5)$$

and, according to uncertainty principle,

$$\begin{aligned} \Delta x &\geq \frac{\hbar}{\Delta p_x} \rightarrow \infty, \\ \Delta y &\geq \frac{\hbar}{\Delta p_y} \rightarrow \infty, \\ \Delta z &\geq \frac{\hbar}{\Delta p_z} \rightarrow \infty. \end{aligned} \quad (6)$$

Thus this kind of annihilation phenomenon could be considered as being infinitely extended. The uncertainty for momentum determines the uncertainty for space intervals. However, in [3] it was also shown that this aspect can not be automatically extended for transmitting fast information necessary for phase correlation or possible regrouping of light beams. Usually deflections (including any possible reflections/refractions or diffractions) are considered to alter just the wave-vector (the momentum) for parts of the spatially extended wave function, without any change regarding the

coherence (phase) of these beams. These phenomena can not be put in correspondence with any creation/annihilation of particles (despite the fact that the wave-vector is altered for parts of the associated wave); otherwise any reflection/refraction or diffraction of a part of the incident wave would generate an annihilation of the *entire* associated wave-function in any other point of space.

As a consequence, an improved intuitive and computational model able to include quantum aspects generated by uncertainty principle regarding a high speed for transmitting interaction and spatial-temporal coherence achieved for deflected parts of the same wave function still has to be determined. For this purpose, this study will show that a certain alternating representation for the wave function can justify in a good manner the patterns created through reflection/refraction of waves on nonperfectly smooth interfaces and phase correspondence of diffracted beams without the need of supplementary support functions. Moreover, this alternating coordinate-momentum representation is suggested by Lorentz transformation of wave trains within special relativity, since the space-time origin for this transformation can not be considered as corresponding to the first interaction with the observer's material medium (as presented in [4]). A very low value for the received signal at zero moment could easily correspond to a noise from the environment, so parts of the received wave-train around this moment could be considered to belong to different wave-functions, not being transformed as a whole (the synchronization moment would differ).

The problem of a superposition of different light (electromagnetic) signals is also connected to the case when a charged particle (an electron for example) moves in an electrostatic field. According to quantum mechanics, within Schrodinger's equation some supplementary terms should be added for momentum \vec{p} and energy E as

$$\begin{aligned}\vec{p} &\longrightarrow \vec{p} - q\vec{A}, \\ E &\longrightarrow E - qV\end{aligned}\quad (7)$$

in scaled measurement units (the light speed c at the denominator of second term in previous equation was set to unity for simplicity).

However, if aspects from quantum field theory are added (second quantization implying creation/annihilation of particles), then problems regarding a phase-loss phenomena are generated. Any local computed probability of interaction with virtual photons (resulted from electrostatic field decomposition) requires a previous transient time (so as physical quantities as wavelength and angular frequency to be defined on a certain time interval) and thus the timelength of the wave-train associated with the final electron tends to zero, a vanishing phenomenon for the initial electron in contradiction with experimental facts. As a consequence, the electromagnetic field (characterized by electric and magnetic fields \vec{E} and \vec{B} or by scalar and vector potentials V and \vec{A}) should be preliminary determined in a certain area before the suggested alternating coordinate-momentum representation to be taken into consideration.

2. Long-Distance Coherence Phenomena Based on Bopp Operators

Starting from standard coordinate and momentum representations in quantum mechanics, it was shown that there is not a unique way to represent a quantum state as a wave function $\Psi = \Psi(\Gamma)$, where Γ represents a point in the classical phase space q, p (position and momentum). Using an integral transformation as

$$\Psi(\Gamma) = \int K(\Gamma; q') \Psi(q') dq', \quad (8)$$

where $K(\Gamma; q')$ corresponds to an integral kernel, another representation can be obtained from the standard coordinate representation $\Psi(q)$. An example is the kernel

$$K_{CS} = (\lambda^2 \pi)^{-1/4} \exp\left(-\frac{(q - q')^2}{2\lambda^2} - \frac{ip}{\hbar}(q' - q)\right) \quad (9)$$

(λ being the natural length scale defined by the mass and frequency of the oscillator) able to generate the Glauber coherent states (eigenstates of the annihilation operator). The Gaussian factor is required by normalization. This kernel K_{CS} was used for constructing phase space representation of quantum mechanics of Torres-Vega and Frederick [5, 6]. They have shown that if $\Psi(q, t)$ satisfies the standard Schrodinger equation

$$i\hbar \frac{\partial}{\partial t} \Psi(q, t) = \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial q^2} + U(q) \right] \Psi(q, t) \quad (10)$$

then $\Psi(\Gamma)$ obtained through the integral transformation based on K_{CS} satisfies a kind of Schrodinger's equation in phase space as

$$i\hbar \frac{\partial}{\partial t} \Psi(\Gamma, t) = \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial q^2} + U\left(q + i\hbar \frac{\partial}{\partial p}\right) \right] \Psi(\Gamma, t). \quad (11)$$

Moreover, if a gauge transformation as

$$\Psi'(q, p) = \exp\left(\frac{-ipq}{2\hbar}\right) \Psi(q, p) \quad (12)$$

is performed, the symmetric formula

$$i\hbar \frac{\partial}{\partial t} \Psi'(\Gamma, t) = \left[\frac{1}{2m} \hat{P}^2 + U(\hat{Q}) \right] \Psi'(\Gamma, t), \quad (13)$$

where

$$\begin{aligned}\hat{Q} &= \frac{q}{2} + i\hbar \frac{\partial}{\partial p}, \\ \hat{P} &= \frac{p}{2} - i\hbar \frac{\partial}{\partial q}\end{aligned}\quad (14)$$

satisfying $[\hat{Q}, \hat{P}] = i\hbar$ according to canonical commutation relations.

Being convenient to scale phase space variables $(q, p) \rightarrow (2q, 2p)$, it results in

$$\begin{aligned}\widehat{Q} &= q + \frac{i\hbar}{2} \frac{\partial}{\partial p}, \\ \widehat{P} &= p - \frac{i\hbar}{2} \frac{\partial}{\partial q}.\end{aligned}\quad (15)$$

These are the so-called Bopp operators. Their utility is connected to quantum jumps, due to the additional term added to standard quantum operators q and p . However, we must point the fact that these Bopp operators are very useful for presenting fast long-range coherence phenomena suggested by operators q (position) and p (momentum) in quantum physics without any creation/annihilation phenomena to be involved (as considered in [3]). For space areas with the same energy (connected to frequency) and/or momentum (connected to wave-vector) it results in the possibility of fast long-range phenomena for transmitting information for parts of the *same* wave function.

This aspect can be proved by the action of operator \widehat{Q} upon a certain function ϕ with the *same* wave vector in an extended spatial area. According to previous relations, it results in

$$dp \rightarrow 0 \implies \frac{\partial}{\partial p} \phi \rightarrow \infty \implies \widehat{Q}\phi \rightarrow \infty. \quad (16)$$

Thus the action of operator \widehat{Q} (used instead of coordinate operator q in standard quantum relations) is extended to long-range distances (theoretically to infinite) within the equations for the same wave function (possibly extended to nonadjacent space areas due to deflecting phenomena previously mentioned). This allows fast-speed correlations for parts of the same wave function.

For example, the wave front of a certain monochromatic planar light beam propagating along x_i axis presents a null variation of the wave vector along the x_j axis normal to x_i . This corresponds to the same wave vector k_j , implying the same momentum p_j along j axis also. It results in $\Delta p_j = 0$, which means that $\widehat{Q}_j \rightarrow \infty$ when a change in the quantum function Ψ occurs along i axis due to a possible reflection/deflection (in the limit case, a nonzero difference $\Delta\Psi$ is divided by an approximately zero difference $\Delta p_j = 0$). Thus a fast long-range interaction can be noticed between all points situated on the planar wave front when a sudden change occurs for one of them, the long-distance phase correlations being justified. For the case of electromagnetic interaction, these changes are connected to the influence of electromagnetic fields (usually represented by scalar potential V and vector potential \vec{A}) in a certain area. This local action is long-distance transmitted by means of this Bopp operator \widehat{Q} .

In a similar manner, Bopp operators can be used for justifying regrouping aspects within the *same* wave packet corresponding to an associated wave-function. Within quantum physics, the motion of a particle is described as an *almost* monochromatic wave function, these means as a superposition of waves with angular frequency ω and wave vector k

situated within some limited intervals. The maximum amplitude for these sets of waves corresponds with a dominant component of angular frequency ω_0 and wave vector k_0 . The other components are considered to possess lower amplitudes symmetrically disposed as decreasing functions around these central values in this energy-momentum space.

The motion of this wave packet is described by the so-called group velocity, defined as

$$v_g = \frac{d\omega}{dk} \quad (17)$$

and (according to correspondence angular frequency-energy and wave vector-momentum) as

$$v_g = \frac{dE}{dp}. \quad (18)$$

The derivatives with respect to k or p are considered for the central values of angular frequency $\omega_0 \leftrightarrow$ energy E_0 and wave vector $k_0 \leftrightarrow$ momentum p_0 .

Since the function *amplitude versus momentum* presents a maximum value for $p = p_0$, it results that dp equals zero around this central value. As a consequence, the term $\partial\Psi/\partial p$ within Bopp operator \widehat{Q} (corresponding to spatial coordinate q) tends to infinite as in the previous case, for any nonzero change of quantum wave function Ψ (a quantum jump). A long-distance action of this dominant component upon the other lower amplitude components within this superposition can be considered, with possible effects upon regrouping aspects of quantum wave functions (spatial dispersion due to different velocities being not allowed).

This aspect is also consistent with Lorentz transformation of wave trains within special relativity, since the space-time origin for this transformation can not be considered as corresponding to the first interaction with the observer's material medium (as presented in [4]). It can be easily considered that a very low value for the received signal at an initial time moment represents a noise from the environment. Thus parts of the received wave-train around this initial moment could be considered to belong to different wave functions at the very beginning, not being transformed as a whole (the synchronization moment would differ). As a consequence, a further interaction between these *already* transformed low-amplitude parts and the central high-amplitude part of the same wave function is required, so as the space-time coordinates to be transformed as a whole for the *entire* wave-function. The synchronization moment (the zero moment for the Lorentz transformation) should be the same for all these parts (considered along time axis) and it is most convenient to define this time origin as the moment when this high-amplitude central part (where momentum p is almost constant and thus Δp is approximately equal to zero on an extended time interval) interacts with the observer's material medium. An extended temporal interval with almost constant momentum p could generate later regrouping phenomena based on the same Bopp operator \widehat{Q} which tends to infinity when Δp is almost zero for a wave function altered through interaction in a very short time interval.

It is true that this aspect would have been suggested by the Heisenberg uncertainty principle also, according to

$$\Delta q \Delta p \geq \frac{\hbar}{2}. \quad (19)$$

However, in the most general case this principle is connected to measurement aspects and does not imply inner dynamics for the quantum wave function. On the contrary, the use of these Bopp operators is related to certain inner dynamics of the wave function, aspects regarding the range of internal interaction corresponding to it being covered. Moreover, the operators generated by variables q and p are mixed together in an almost symmetrical form.

It can be noticed that p and q operators can be encountered both under: (i) multiplicative form—the operator written at the numerator of a multiplicative factor acting to the right/left upon a ket or bra state vector, and (ii) differential form—the operator written at the denominator of a differential expression acting to the right/left upon a ket or bra state vector (according to Dirac notations). Unlike multiplicative operators, the differential operators can not be considered as acting almost instantaneously in any intuitive or computational (discretized) model. A kind of memory for certain variables on previous time intervals is required. This aspect of any discretized computational model intending to solve a differential equation can be easily extended to quantum models based on Bopp operators, where discontinuities (quantum jumps) could be involved.

Moreover, the symmetrical form of Bopp operators as related to q, p operators suggests that these coordinate-momentum operators q and p should be used *at the same time* in a quantum equation of evolution. Yet this could correspond to an increased number of variables/operators used in the *same* equation. So a method for analyzing equations based on Bopp operators is required. As will be shown in next paragraph, a possible choice is represented by the use of symbolic determinants for solving a set of linear equations for the case when the free terms tend to zero (the limit being an undetermined set of equations with possible nontrivial solutions). As the source of these quantum jumps, terms corresponding to external electromagnetic fields should be inserted.

3. A Symbolic Determinant for Alternating Coordinate-Momentum Representation

Let us consider that two *real* functions should satisfy (in the limit case) a homogenous linear system of equations with null free terms as

$$\begin{aligned} \hat{a}\phi + \hat{b}\eta &= 0, \\ -\hat{b}\phi + \hat{a}\eta &= 0, \end{aligned} \quad (20)$$

where \hat{a} and \hat{b} are not constant coefficients, but linear operators. Trying to solve this system by reduction method, the first step consists in amplifying first equation by operator

\hat{b} to the left and the second equation by operator \hat{a} also to the left. It results in

$$\begin{aligned} \hat{b}\hat{a}\phi + \hat{b}^2\eta &= 0, \\ -\hat{a}\hat{b}\phi + \hat{a}^2\eta &= 0. \end{aligned} \quad (21)$$

Next step consists in performing the sum of these two equations. It results in

$$[\hat{b}\hat{a} - \hat{a}\hat{b}]\phi + [\hat{b}^2 + \hat{a}^2]\eta = 0. \quad (22)$$

From now on, this equation will be referred to as *Complete Equation*. If \hat{a} and \hat{b} are commutative operators, then first term vanishes in this complete equation, so the second term will be

$$[\hat{b}^2 + \hat{a}^2]\eta = 0. \quad (23)$$

This equation is similar to the wave equation written in the quadridimensional space, where operators \hat{b} and \hat{a} correspond to partial differentiation with respect to a certain space coordinate (x for example) and to ict , respectively (t representing time and c corresponding to light speed in vacuum). It should be noticed that these differentiation operators satisfy the commutative relation $\hat{b}\hat{a} - \hat{a}\hat{b} = 0$ and so the complete equation will correspond to a simple propagation equation.

However, if \hat{b} and \hat{a} are noncommutative operators, this means

$$\hat{b}\hat{a} - \hat{a}\hat{b} = 1 \quad (24)$$

in scaled measurement units; then the first term in complete equation previously presented does not vanish. Moreover, this first term in complete equation can be moved to right-hand side (with opposite sign) and becomes a kind of source (a free term) for the second term, as

$$[\hat{b}^2 + \hat{a}^2]\eta = [\hat{a}\hat{b} - \hat{b}\hat{a}]\phi. \quad (25)$$

The left-hand side operator acting upon function η is the symbolic determinant \hat{S} for the linear set of equations.

A similar equation (with a similar left-hand side quadratic expression and a right-hand side nonzero value) can be encountered for the hamiltonian function of the quantum harmonic oscillator. For this case operators \hat{b} and \hat{a} correspond usually to a coordinate operator and to differentiation with respect to this coordinate (the quantum momentum operator), so this hamiltonian \hat{H} equals

$$\hat{H}\eta = \left(\frac{\partial^2}{\partial x^2} + x^2 \right) \phi \quad (26)$$

in scaled measurement units (the constants in front of partial differentiation operators and in front of coordinate operators were set to unity for simplicity).

At first sight this seems to be pure algebra or at least symbolic algebra for representing the action of operators in quantum mechanics. No inner dynamics for a wave function

could be noticed. However, a decomposition of operators \hat{a} and \hat{b} according to

$$\begin{aligned}\hat{a} &= \hat{a}_i - \hat{a}_f, \\ \hat{b} &= \hat{b}_i - \hat{b}_f\end{aligned}\quad (27)$$

can be performed (the minus sign and the indices will be justified later). By substitution, it results in

$$\begin{aligned}(\hat{a}_i - \hat{a}_f)\phi + (\hat{b}_i - \hat{b}_f)\eta &= 0, \\ (-\hat{b}_i + \hat{b}_f)\phi + (\hat{a}_i - \hat{a}_f)\eta &= 0\end{aligned}\quad (28)$$

which can be written also as

$$\begin{aligned}\hat{a}_i\phi + \hat{b}_i\eta &= \hat{a}_f\phi + \hat{b}_f\eta, \\ -\hat{b}_i\phi + \hat{a}_i\eta &= -\hat{b}_f\phi + \hat{a}_f\eta.\end{aligned}\quad (29)$$

The right-hand side of these two previous equations (where index f can be encountered for \hat{a} and \hat{b} operators) could be considered as being determined by the left-hand side of them. It looks like the operators \hat{a}_i and \hat{b}_i act upon a pair of functions ϕ and η so as to determine the right-hand side of these two previous equations. Since this right-hand side depends upon the action of operators \hat{a}_f and \hat{b}_f upon the same functions ϕ and η considered at a bit later time moment, the value of the two functions ϕ and η at the later time can be further determined.

Certain *inner* dynamics of the wave function work so as to obtain a minimum difference between functions ϕ and η at the initial time moment and the same functions ϕ and η considered at this later time moment. If the time difference between the initial moment of time (considered in left-hand side) and the final moment of time (the later moment of time, considered in right-hand side) is extremely small (orders of magnitude less than the infinite small time intervals required by the use of differential operators as $\partial/\partial q$ or $\partial/\partial p$ for quantum states), it will be not necessary to use different notations or supplementary indices for functions ϕ and η in left- and right-hand sides of these two previous equations (indices i and f , with the meaning of initial and final, respectively, being kept just for operator decomposition).

By moving right-hand side terms to the left in both previous equations the set of linear equations with null free term presented at the beginning of this paragraph will be obtained. This shows that the symbolic determinant similar to cuadridimensional propagation equation (the second part of the complete equation) and the commutative or noncommutative rules (the first part of the complete equation) could be generated by an inner dynamics of quantum functions, a quick path from some initial functions ϕ_i, η_i to certain final functions ϕ_f, η_f being involved.

For determining the operators \hat{a}_i, \hat{a}_f and \hat{b}_i, \hat{b}_f to be placed hand and right-hand side on the previous set of equations the standard decomposition of hamiltonian operator in a sum of the kinetic energy operator (usually a differential

operator) and the potential energy operator (usually a multiplicative operator) could be used. This suggests the use of operators as

$$\begin{aligned}\hat{a}_i &= \frac{\partial}{\partial x}, & \hat{b}_i &= \frac{\partial}{\partial(ict)}, \\ \hat{a}_f &= -\frac{q}{\hbar c}V, & \hat{b}_f &= \frac{iq}{\hbar c}\vec{A}\end{aligned}\quad (30)$$

for the case of a particle with charge q situated within a macroscopic electromagnetic field represented by the scalar potential V and vector potential \vec{A} . Moving all terms left-hand side (so as to obtain the pair of linear equations with null free terms), the symbolic determinant \hat{S} previously mentioned (acting upon function η within second term of complete equation) will be

$$\hat{S} = \hat{b}^2 + \hat{a}^2 = \left[\frac{\partial}{\partial x} - \frac{iq\vec{A}}{\hbar c} \right]^2 + \left[\frac{\partial}{\partial(ict)} + \frac{qV}{\hbar c} \right]^2. \quad (31)$$

According to Klein-Gordon equation for the relativistic wave-function of an electron in an external electromagnetic field, the result for the action of this symbolic determinant \hat{S} upon η function should be equal to

$$\hat{S}\eta = \frac{(mc)^2}{\hbar^2}\eta, \quad (32)$$

where η is the associated wave function for the electron. It can be noticed that the action of the symbolic determinant \hat{S} upon this wave function equals the wave function multiplied by a constant. Thus the use of scalar and vector potentials as operators acting upon the wave functions at a later time moment within this alternating coordinate-momentum representation is justified.

As a consequence, the electromagnetic field (characterized by scalar and vector potentials V and \vec{A}) should be preliminary determined in a certain area before the suggested alternating coordinate-momentum representation to be taken into consideration. No more virtual photons are necessary for computing trajectories of charged particles within an external electromagnetic field, and phase-loss phenomena (see [5]) are avoided.

4. Conclusions

This study has presented an alternating coordinate-momentum representation for quantum states recommended for modelling propagation and transition of associated wave functions, based on Bopp operators and on a certain symbolic determinant corresponding to a set of two linear equations with null free terms for deriving basic equations in quantum mechanics. Unlike computational aspects (see [7] where just encoding aspects are involved for a certain data set), it tries to extend input-output data correlations at possible inner dynamics of wave function. Thus a certain background dynamics could be added to the theory of light deflection based on conformal fields (see [8]). The

long-distance interaction suggested by this dynamics based on Bopp operators is not frequency limited (as standard gaussian noise, see [9]).

It is shown that this alternating representation can justify in a good manner the patterns created through reflection/refraction of waves on nonperfectly smooth interfaces and phase correspondence of diffracted beams without the need of supplementary support functions. Correlations with Lorentz transformation of wave functions by interaction with a certain material medium (the space-time origin of a wave-train being adjusted) are also presented, and supplementary aspects regarding the use of electromagnetic scalar and vector potentials for modelling evolution within this alternating representation are added.

Computational aspects regarding decaying modes (as presented in [10]) or a greater number of particles involved will be analysed in further studies, based on scheduling processes (as in [11]) and distributing environment techniques (as in [12]).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] E. G. Bakhoun and C. Toma, "Transient aspects of wave propagation connected with spatial coherence," *Mathematical Problems in Engineering*, vol. 2013, Article ID 691257, 5 pages, 2013.
- [2] P. E. Sterian, "On the connection between Heisenberg's uncertainty relations and the evolution quantum equations," *Buletinul Institutului Politehnic București*, vol. 46-47, pp. 43-48, 1984 (Romanian).
- [3] C. Toma, "Wavelets-computational aspects of sterian-realistic approach to uncertainty principle in high energy physics—a transient approach," Internal Report, Politehnica University Bucharest, Physics Department, 2013.
- [4] C. Toma, "The advantages of presenting special relativity using modern concepts," *Balkan Physics Letters*, vol. 5, pp. 2334-2337, 1997.
- [5] G. Torres-Vega and J. H. Frederick, "Quantum mechanics in phase space: new approaches to the correspondence principle," *The Journal of Chemical Physics*, vol. 93, no. 12, pp. 8862-8874, 1990.
- [6] G. Torres-Vega and J. H. Frederick, "A quantum mechanical representation in phase space," *The Journal of Chemical Physics*, vol. 98, no. 4, pp. 3103-3120, 1993.
- [7] S. V. Chekanov, K. Strand, and P. Van Gemmeren, "ProMC: input-output data format for HEP applications using varint encoding," *Computer Physics Communications*, vol. 185, pp. 2629-2635, 2014.
- [8] C. Cattani, M. Scalia, E. Laserra, I. Bochicchio, and K. K. Nandi, "Correct light deflection in Weyl conformal gravity," *Physical Review D: Particles, Fields, Gravitation, and Cosmology*, vol. 87, Article ID 047503, 2013.
- [9] M. Li and W. Zhao, "On bandlimitedness and lag-limitedness of fractional Gaussian noise," *Physica A*, vol. 392, no. 9, pp. 1955-1961, 2013.
- [10] M. Dima, "Exclusive reconstruction of semi-leptonic B-hadron decay modes," *Journal of High Energy Physics*, vol. 2011, article 8, 2011.
- [11] V. Cristea and F. Pop, "Optimization of scheduling process in grid environments," *Politehnica University Bucharest (UPB) Scientific Bulletin, Series C: Electrical Engineering*, vol. 71, no. 4, pp. 35-48, 2009.
- [12] C. Dobre, F. Pop, and V. Cristea, "New trends in large scale distributed systems simulation," *Journal of Algorithms & Computational Technology*, vol. 5, no. 2, pp. 221-257, 2011.

Research Article

A Parallel Community Structure Mining Method in Big Social Networks

Songchang Jin,^{1,2} Philip S. Yu,² Shudong Li,¹ and Shuqiang Yang¹

¹College of Computer, National University of Defense Technology, Changsha, Hunan 410073, China

²Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA

Correspondence should be addressed to Songchang Jin; jinsongchang87@gmail.com

Received 5 July 2014; Accepted 2 August 2014

Academic Editor: Haipeng Peng

Copyright © 2015 Songchang Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Community structure plays a key role in analyzing network features and helping people to dig out valuable hidden information. However, how to discover the hidden community structures is one of the biggest challenges in social network analysis, especially when the network size swells to a high level. Infomap is a top-class algorithm in nonoverlapping community structure detection. However, it is designed for single processor. When tackling large networks, its limited scalability makes it less effective in fully utilizing server resources. In this paper, based on infomap, we develop a scalable parallel nonoverlapping community detection method, Pinfomr (parallel Infomap with MapReduce), which utilizes the MapReduce framework to solve the two problems. Experiments on artificial networks and real datasets show that our parallel method has satisfying performance and scalability.

1. Introduction

A few common properties in many complex networks have been discovered: small-world property, scale-free feature, and community structure pattern [1–4]. Community structure is playing a key role in the formation and function of these networks [5]. However, it is one grave challenge in complex systems [6].

Current social networks have jumped to millions even billions of nodes [7]. Take Facebook for example, its monthly active user has reached 1.16 billion [8]. However, due to computational costs, traditional community discovery algorithms are willing, but unable to tackle such huge complex networks. So, it is necessary to implement a fast and scalable approach to detect communities in big social networks.

Network partitioning is NP-complete [9]. Partitioning a network into approximately equal sized components while minimizing the number of edges between different components is extremely important in parallel computing [10]. For example, parallelizing many applications involves the problem of assigning data or processes evenly to processors, while minimizing the communication traffic. However, when the network size reaches a certain level, direct segmentation

on the original networks is not realistic, and there exist deficiencies of convergence rate of traditional algorithms.

Nowadays, mainstream servers are configured with high performance hardware. Empirical studies [11] have showed that infomap [12] is a top-class standalone algorithm for nonoverlapping community detection. However, due to the limitations of technological level, processing capability of single core has encountered a bottleneck and the scalability of infomap is suffered as a consequence, that is, because it only utilizes one core or processor of the server. Besides, computing resource waste is an additional product of infomap running on multiprocessor server. How to improve the scalability of infomap and make full use of servers is an awkward subject.

Information science is shifting from computing-intensive to data-intensive [13] with the advent of the era of big data. Some novel parallel computing frameworks shine, in which MapReduce [14] is one of the best. In this paper, based on our previous work [15], we present a new scalable parallel community detection method coalescing several existing excellent techniques, such as infomap, k -shell decomposition, multilevel network partitioning, and MapReduce. A high-level description of our approach is as follows. First, we divide the whole network into a number of partitions and

the number of partitions is far less than that of community structures. To speed up the process, we develop an enhanced multilevel partitioning method. Next, with MapReduce, we run parallel method to mine the community structures simultaneously within the partitions. Finally, we collect the community structures together to form a final result.

Main contributions of this paper are as follows: (1) we propose a new model to mine community structure in big social networks. (2) We integrate k -shell decomposition theory with multilevel k -way partitioning algorithm to deal with peripheral nodes. (3) We implement a scalable and parallel infomap to uncover community structures and to improve resource utilization rate.

The rest of this paper is organized as follows. Section 2 briefly reviews some concepts and background information. Section 3 provides problem statement and detailed description of the parallel community detection method. In Section 4, we conduct a couple of experiments to evaluate the performance of the method proposed in this paper. Finally, Section 5 provides some concluding remarks and outlines future research directions.

2. Preliminary Knowledge

2.1. Relevant Concepts. In this paper, we only study undirected networks, which can be mathematically described as G , consisting of node set V and edge set E ; n represents the number of nodes, $v_i \in V$ represents a node, and $d(v_i)$ means its degree; m represents the number of edges and $e_{i,j}$ is the edge between v_i and v_j , where $0 < i \neq j \leq n$.

Infomap is based on information-theory. So some information-theoretic concepts are briefly reviewed here. In information theory, the information contained in a distribution is called entropy. For a discrete random variable $X = \{x_1, x_2, \dots, x_n\}$ with a probability distribution $P(X)$, its entropy is

$$H(X) = - \sum_{x \in X} p(x) \log p(x). \quad (1)$$

Mutual information calibrates the shared information between two distributions, X and Y . We define $P(X, Y)$ as the joint probability of X and Y . $P_x(X)$ and $P_y(Y)$ are defined as marginal probability distribution of X and Y , respectively. Then, mutual information of X and Y is

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(x, y)}{p_x(x) p_y(y)}. \quad (2)$$

Normalized mutual information (NMI) is often used for evaluating clustering result, information retrieval, feature selection, and so forth. Value range of NMI is $[0, 1]$ and when X and Y are the same, $\text{NMI}(X; Y)$ equals 1.0. Consider

$$\text{NMI}(X; Y) = \frac{I(X; Y)}{\sqrt{H(X) H(Y)}}. \quad (3)$$

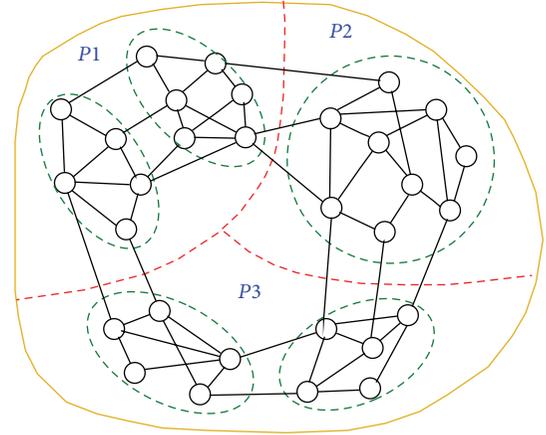


FIGURE 1: A network with 5 communities and 3 partitions.

2.2. K -Shell Decomposition Theory. K -shell decomposition is a well-established method for analyzing the structure of large-scale networks [16–18]. In particular, it provides a method for identifying hierarchies in a network. It is assumed that importance of a node is not related to its degree but its location. The process assigns an integer index, ks , to each node, representing its location within the successive layers (k -shells) in the network. The ks index is a robust measure and the node ranking is not significantly influenced in the case of incomplete information. The k -core of a network G is the maximum subnetwork of G whose degree is no less than k . The k -shell of G is the set of all nodes belonging to the k -core of G but not to the $(k + 1)$ -core.

Nodes are assigned to k -shells based on their remaining degree, which is obtained by successive pruning of nodes with degree smaller than the ks value of the current layer. The decomposition process starts by removing all nodes with degree $d = 1$. After that, some nodes may be left with one link. We then prune the system iteratively until there is no node left with $d = 1$ in the network. The removed nodes, along with the corresponding links, form a k shell with index $ks = 1$. In a similar fashion, we iteratively remove the next k -shell, where $ks = 2$, and continue to remove higher k shells until all nodes are removed. As a result, each node is associated with one ks index, and the network can be viewed as the collection of all k shells. ks value of a node can be very different from its degree. In Figure 2, we can see that v_9 has 7 neighbors with $ks(v_9) = 1$. Figure 5 is the result of Figure 2 of which peripheral nodes are processed.

2.3. Multilevel k -Way Partitioning Method. Partitioning the node set V of a network G into k disjoint subsets $\{V_1, V_2, \dots, V_k\}$ is called a k -way partitioning of V . Each subset and the edges within the subset constitute a *partition* of G . Figure 1 shows a simple network with 5 communities surrounded by the dotted circles and 3 partitions.

Definition 1 (effective edge lost ratio). An edge whose endpoints are in the same community, that is, intracommunity edge, is called an effective edge. If the endpoints of an effective

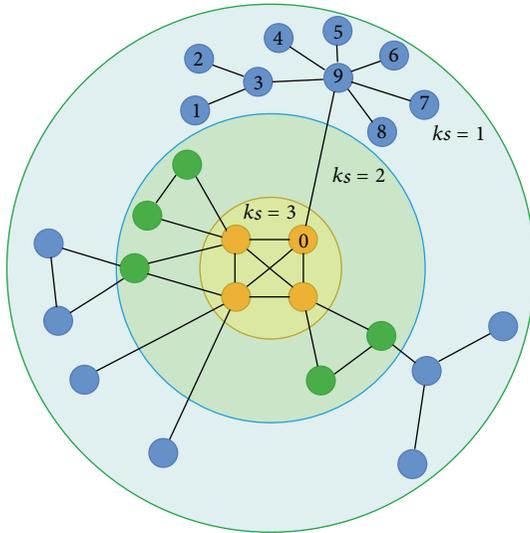


FIGURE 2: Schematic representation of k -shell decomposition.

edge are divided into different partitions, then we call it an effective edge lost. The effective edge lost ratio is the percentage of the effective edge lost divided by the total number of edges in the network.

In Figure 1, $e_{3,4}$ is an effective lost, whereas $e_{1,2}$ is not. It is apparent that effective edge lost plays a more important role in the community detection than the edges connecting nodes in different communities and being cut off by partition boundaries.

A number of high-quality and computationally efficient graph partitioning methods have been proposed and multilevel graph partitioning algorithms [9, 19, 20] are currently considered to be a start-of-the-art method and being extensively used. Here, we optimize the multilevel k -way partitioning method proposed by Abou-Rjeili and Karypis [21] to partition the power law networks.

From Figure 3, we can see that multilevel k -way partitioning method consists of coarsening phase, initial partitioning phase, and uncoarsening and refinement phase. Instead of trying to partition directly on the original graph G_0 , multilevel partitioning algorithms first obtain a sequence of successive approximations, such as G_1 , G_2 , and G_3 in the coarsening phase, of the original graph. Each of these approximations represents smaller than the size of the original graph. This process continues until a level of approximation is reached where the graph contains only a few hundreds of nodes. At this point, partitioning algorithms begin to compute partitions of that graph, corresponding to the five small partitions of G_3 in the initial partitioning phase, and since the graph is quite small, even simple algorithms are able to take it over and get reasonably good results, such as K-L [22]. And there is a parameter used to control the balance of partitions. The final step, uncoarsening and refinement phase, is to map the partitions of the smallest graph G_3 onto the original graph G_0 and to derive final partitions.

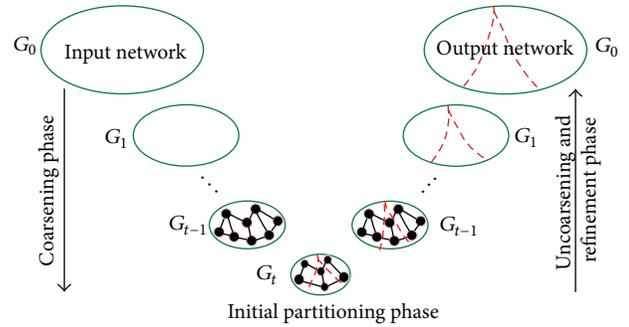


FIGURE 3: A schematic diagram of multilevel k -way partitioning method.

2.4. Infomap. In this paper, we continue our work on the information theoretic community detection model-infomap. First, we briefly review the model. It utilizes the duality between compressing a data set and detecting and extracting significant patterns or structures within the data, which is a statistical concept known as minimum description length statistics [23]. A random walk, represented as a Markov process, is used as a proxy for information flow. For a community-structured network, when a random walker enters a community, it tends to stay in it for a long time and the possibility of moving out into another community is low.

In an undirected network, the random walk has a state $x(t) \in V$ at time t , indicating where it is. Then in next step, $t + 1$, the walker will move to v_j chosen at random from neighbors of v_i . To describe the state of random walker, a 2-level description model with Huffman coding is proposed. The first level encodes the communities and the second level encodes the nodes within the communities. Then we can use “community ID + node ID” to uniquely describe a particular node in the network. Huffman codes are prefix-free coding scheme and are optimally efficient for symbol-by-symbol encoding. It saves space by assigning short codewords to common events or objects and long codewords to rare ones, just as Morse code uses short codes for common letters and longer codes for rare ones. So, the path of the random walker can be described as a coding sequence.

Figure 4 is an example for illustrating the 2-level description method. Assuming there are 2 communities divided like Figure 4(b), then the code sequence for the random walk in Figure 4(a) is “**0** III 00 10 III 010 10 011 110 00 10 **1011** 00 01 10 00 11 10 01 1.” The underlined word “**0**” in bold format means random walk starting from C1. The underlined word “**1011**” in bold format means random walk leaving from C1 and entering C2. The description length of the sequence will be 50 bits and about 2.63 bits per step. But, in Figure 4(c), we will need 57 bits and about 3.0 bits per step. Community division is obviously more reasonable in Figure 4(b) than in Figure 4(c), and the average description length in the former one is shorter than in the latter one. From the perspective of information theory, we know that smaller entropy corresponds to smaller uncertainty. Corresponding to the community detection, smaller entropy means smaller indistinctness or clearer community structure.

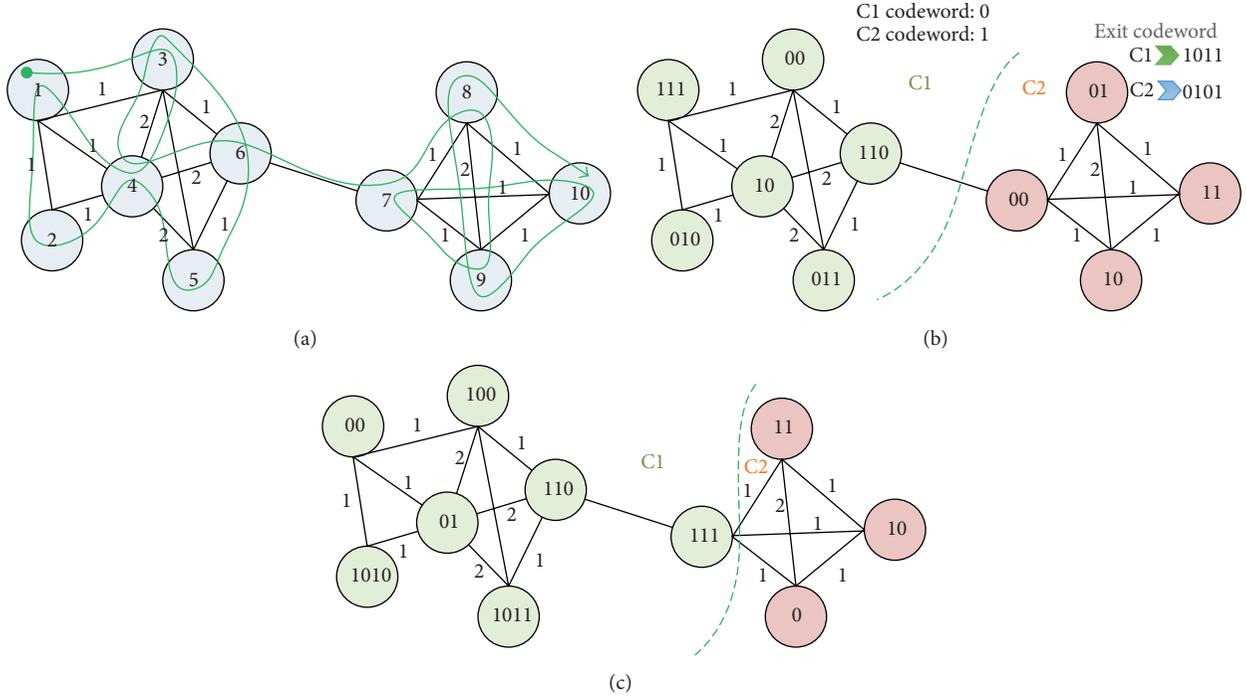


FIGURE 4: Random walk and 2-level Huffman coding on a network with two communities.

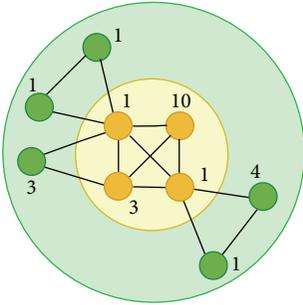


FIGURE 5: The result of Figure 2 without peripheral nodes. Integers indicate node weights.

3. Parallel Community Detection Method

3.1. Problem Statement. Assuming there is an optimal community division, M^* , in a community-structured network G . With M^* , the network G is divided into num^* communities, and the lower limit of average description code length is $L(M^*)$. According to the Shannon source coding theorem [24] and the Kraft's inequality [25], we know that, for any division pattern M , the average codeword length per source symbol, $L(M)$, for an optimal prefix-free code satisfies

$$H(X) \leq L(M^*) \leq L(M) \leq H(X) + 1. \quad (4)$$

Obviously, calculating an endless random walk on a network to get $L(M)$ is not realistic. Fortunately, when randomly walking on a network endlessly, we will get a steady

visit frequency for each node, and we can calculate that easily with many methods, such as PageRank. With the steady visit frequency distribution, we will be able to describe the state of the random walker easily. For $x(t+1) \in \{\text{neighbor}(x(t))\}$, the probability of $x(t+1)$ and $x(t)$ being in the same community is p_{within} and the probability of being in different communities is $q_{k,\text{out}}$, where $x(t+1)$ belongs to community k . Then the $L(M)$ can be described as

$$L(M) = q_{\text{out}} H(Q) + \sum_{i=1}^{\text{num}} p_{\text{within}}^i H(P^i), \quad (5)$$

where q_{out} means the probability of moving out from the current community and $q_{\text{out}} = \sum_{i=1}^{\text{num}} q_{i,\text{out}}$. $H(Q)$ is the average description length of nodes in all communities, and it can be expressed as

$$\begin{aligned} H(Q) &= \sum_{i=1}^{\text{num}} H_i(Q) \\ &= - \sum_{i=1}^{\text{num}} \frac{q_{i,\text{out}}}{\sum_{j=1}^{\text{num}} q_{j,\text{out}}} \log \frac{q_{i,\text{out}}}{\sum_{j=1}^{\text{num}} q_{j,\text{out}}}. \end{aligned} \quad (6)$$

With the probability p_a or $p(a)$ to visit the node v_a , p_{within}^i represents the probability of staying in the current community during the next step, and $p_{\text{within}}^i = \sum_{v_a \in C_i} p_a + q_{i,\text{out}}$. $H(P^i)$

```

(1) method Map(nid  $n$ , node  $v$ )
(2)  $p \leftarrow p_v / |v.adjacencyList| // (8)$ 
(3) emit(nid  $n$ ,  $v$ ) //pass along network structure
(4) for all nodeid  $m \in v.adjacencyList$  do
(5)   emit(nid  $m$ ,  $p$ ) //pass pagerank contribution to neighbors
(6) endfor
(7) method Reduce(nid  $m$ , [ $p_1, p_2, \dots$ ])
(8)  $v \leftarrow \phi$ 
(9) for all  $p \in counts[p_1, p_2, \dots]$  do
(10)  if isNode( $p$ ) then
(11)    $v \leftarrow p$  //recover local network structure
(12)  else
(13)    $s \leftarrow s + p$  //sums pagerank contributions
(14)  endif
(15) endfor
(16)  $p_v \leftarrow s$ 
(17) emit(nid  $m$ ,  $v$ )

```

ALGORITHM 1: Steady visiting probability vector calculation on MapReduce (VPC).

expresses the information entropy of the visiting probability of the nodes in the community C_i , which can be written as

$$\begin{aligned}
H(P^i) &= -\frac{q_{i,out}}{q_{i,out} + \sum_{v(b) \in C_i} p(b)} \\
&\times \log \frac{q_{i,out}}{q_{i,out} + \sum_{v(b) \in C_i} p(b)} \\
&- \sum_{v(a) \in C_i} \frac{p(a)}{q_{i,out} + \sum_{v(b) \in C_i} p(b)} \\
&\times \log \frac{p(a)}{q_{i,out} + \sum_{v(b) \in C_i} p(b)}. \tag{7}
\end{aligned}$$

For the NP-complete challenge, we cannot achieve the global optimal division pattern M^* by direct computing on a big social network. However, we can archive a set of local optima to approximate M^* by partitioning the network into small subnetworks (partitions) and tackling them independently with MapReduce. Then, the issue will become how to discover optimal division pattern M_i^* in partition P_i and get final $M^* = \bigcup_{i=1}^k \{M_i^*\}$. For P_i , it would be sufficient to calculate the $L(M_i)$ for different M_i s and pick up the one with shortest description length as M_i^* . Finally, we get a community set $C = \{C_1, C_2, \dots, C_k\}$, where $C_i = \{c_{i,1}, c_{i,2}, \dots, c_{i,k1}\}$ corresponds to P_i , $\bigcup_{i=1}^k C_i = V$ and $C_i \cap C_j = \emptyset$, where $0 < i \neq j \leq n$ and $|P| \ll |C|$.

3.2. Procedure of the Parallel Community Detection. For the convenience of illustration, we adopt Figure 1 to start this section and assume that the amount of partitions is far less than that of communities in big social networks. There are 3 stages in the parallel community detecting process.

In the first stage, we calculate the steady visiting probability of all nodes (shown in Algorithm 1). Here, we modify the traditional PageRank, which is used to deal with directed networks and run a iterative MapReduce-based version to get the global steady visit probability vector. In each iteration, visit probability of v_a is (since there is no teleport and link sink in undirected networks, we set $\tau = 0$)

$$p_a = \tau \times \frac{1}{n} + (1 - \tau) \times \sum_{v_i \in \text{neighbor}(v_a)} \frac{p_i}{d_i}. \tag{8}$$

Second, we use multilevel k -way partitioning method enhanced by k -shell decomposition method to divide a big social network into k approximately equal sized disjoint partitions (P_1, P_2 , and P_3 in Figure 1). Edges cut off by partition boundaries will be discarded. As networks studied here are sparse and with community structure, *edge cut (lost) ratio* will be low. However, partitioning method has a decisive influence on the final community detection effectiveness which will be explained with experiments in Section 4. A matching of a network is a set of edges and no two edges in it share a same node. To coarsen a network, a commonly used method is to collapse the node pairs forming the matching, such as random matching, heavy-edge matching, and maximum weighted matching. However, it shrinks at a slow rate and does not consider the relative importance of nodes in different locations. We all know that there is a large number of low-degree and low k s value nodes in power law networks, so we can turn this characteristic into revenue. Here, we use the k -shell decomposition to merge the peripheral nodes with high speed and more accurate performance during the coarsening phase (shown in Algorithm 2).

In the last stage, parallel community detection method is carried out on all partitions (such as the 3 partitions in Figure 1) and all partitions are tackled independently. When the parallel process finished, each partition generates

```

(1) set  $int\ id[n + 1] = \{0, 1, 2, \dots, n\}$ 
(2) while  $true$ , do
(3)   for all node  $v_i \in V$ , do
(4)     if  $d[i] == 1$ , then
(5)        $tmpV.add(v_i)$  //a list to store nodes with  $ks = 1$ 
(6)     endif
(7)   endfor
(8)   if  $tmpV.isEmpty()$ , then
(9)     break //process finished
(10)  endif
(11)  for all node  $v_j \in tmpV$ , do
(12)    if  $flag[j] != true$ , then // $v(j)$  hasn't been annexed
(13)       $k = getID(v_j.getDiffNeighbor())$  //get rid off  $v(j)$  itself
(14)       $id[j] = k$  //assign new id to  $v(j)$  annexed by  $v(k)$ 
(15)       $w_k = w_k + 1$ 
(16)       $updateNeighbor(v_k, j)$  //replace  $v(j)$  with  $v(k)$  in  $v(k)$ 's neighbor
(17)       $d[k] = d[k] - 1$ 
(18)       $d[j] = d[j] - 1 = 0$ 
(19)       $flag[j] = flag[k] = true$ 
(20)    endif
(21)  endfor
(22) endwhile

```

ALGORITHM 2: Amalgamation of peripheral nodes in coarsening phase.

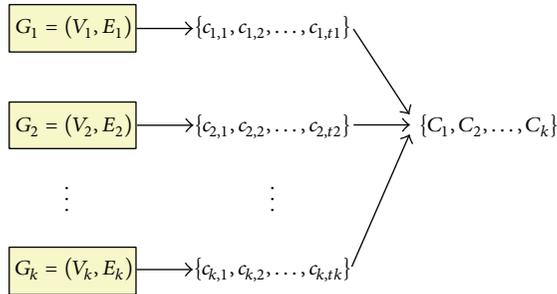


FIGURE 6: A schematic diagram of MapReduce process for community detection.

a community set (such as the 2 communities in P_1 in Figure 1). Combining all the community sets together will be the final result. Figure 6 gives a straightforward statement of this process and more detail is shown in Algorithm 3. At the beginning of this stage, we treat each node as a community and then use a bottom-up approach with greedy scheme to (1) find out community pairs that can minimize $\Delta L(M)$ and (2) merge them to form new communities.

4. Experiments and Analysis

In this section, we conduct several experiments and analyze the results. All experiments are running on the Hadoop-1.1.1 cluster of Antivision Software Ltd. The cluster consists

of 20 PowerEdge R320 servers (Intel Xeon CPU E5-1410 @2.80 GHz, memory 8 GB) with 64-bit NeoKylin Linux OS, and servers are connected by a Cisco 3750G-48TS-S switch. Data sets are shown in Table 1, including artificial networks and real networks.

All artificial networks used here are generated by LFR benchmark. In LFR, some parameters give us a direct control on network properties: network size (n), degree distribution ($\gamma, d_{max}, avg(d)$), and community structure (β, mix) [26]. γ and β are exponents for degree and community size distributions, which range between $[2, 3]$ and $[1, 2]$, respectively. Mix is the ratio of edges connecting nodes from different communities divided by collective edges of all communities. For the average and balance, we set $\gamma = 2.5$ and $\beta = 1.5$ for artificial networks.

4.1. Accuracy Experiments. In accuracy experiments, we compare our method, Pinformr, with two top-class methods, Louvain algorithm [27] and OSLOM algorithm [28], on different data sets and with different partition numbers. The data sets used are D_0, D_1 , and D_2 , and result is shown in Figure 7, where $|P|$ means partition number. The situation when $C = 1$ or $|C| = n$ is defined by us as no community structure and NMI in this case is set to 0, but the case when $|C|$ is close to n is discarded. Taking for instance Louvain in Figure 7(a), when $mix = 0.75, |C|/n = 0.373$ and we conclude that a community just contains 3 nodes averagely. From the design of LFR, we know that when mix value is too high, such as higher than 0.75, there will be no obvious community structures, and the network will not be a power law network but more like a random network which is not the focus here.

```

(1) Initialization-global variables
(2)  $nc = |V_t|$  //number of communities
(3)  $L_0 = 0$ 
(4) for  $i$  from 1 to  $|V_t|$ , do
(5)    $cid[i] = i$  //community ID of  $v_i$ 
(6)    $L_0 = L_0 + H(v_i)$ 
(7) endfor
(8) method Map( $node\ v, adjacencyList$ )
(9) while  $\delta < 0$ , do
(10) for  $i \in [1, |V_t|]$ , do
(11)    $lock[i] = false$  //able to merge in current iteration
(12) endfor
(13) for  $i \in [1, |V_t|]$ , do
(14)    $j = cid[random(|V_t|) + 1]$  //randomly select a node(community)
(15)    $curNeighbor[ ] = getAdjacencyList(c_j)$ 
(16)    $\delta = 0$  //decrease of  $L$ 
(17)    $maxCID = 0$  //community id which leads to minimum  $\delta$ 
(18)   for  $c_s$  in  $curNeighbor[ ]$ , do
(19)     if  $L_{s,k} - L_0 < \delta$  //if merge  $c_s$  and  $c_k$ 
(20)        $maxCID = s$ 
(21)        $\delta = L_{s,k} - L_0$ 
(22)     endif
(23)   endfor
(24)   if  $maxID \neq 0$  &  $\delta \neq 0$ , then
(25)      $lock[s] = lock[k] = true$ 
(26)      $nc = nc - 1$ 
(27)      $L_0 = L_0 - \delta$ 
(28)      $pid[s] = j$ 
(29)   endif
(30) endfor
(31) endwhile
(32) for  $i \in [1, nc]$ , do
(33)    $emit(i, getAdjacencyList(c_i))$ 
(34) endfor

```

ALGORITHM 3: The map procedure of community detection on a subnetwork G_t .TABLE 1: Data sets used in experiments (increment of mix is set to 0.05. $M = 10^6$).

Data set	n	m	avg(d)	γ	β	Size(C)	mix
LiveJournal	3,997,962	34,681,189	/	/	/	/	/
Youtube	1,134,890	2,987,624	/	/	/	/	/
Orkut	3,072,441	117,185,083	/	/	/	/	/
D0	0.1 M	/	45	2.5	1.5	[25, 120]	[0.1, 0.75]
D1	0.2 M	/	40	2.5	1.5	[25, 100]	[0.1, 0.75]
D2	0.4 M	/	40	2.5	1.5	[25, 100]	[0.1, 0.75]
D3	80,000	/	45	2.5	1.5	[25, 120]	[0.1, 0.75]
D4	0.2 M, 0.4 M, 0.8 M, 1.6 M	/	45	2.5	1.5	[25, 100]	0.45
D5	3.2 M, 6.4 M, 10 M	/	45	2.5	1.5	[25, 100]	0.45

Figure 7 indicates that Pinfomr is more stable and accurate than the others in uncovering community structures in power law networks. For running time, we can see that for the same data set, Louvain consumes the longest time and Pinfomr needs the shortest time. OSLOM requires a little more time than Pinfomr when mix parameter is not too big.

From Figure 7(c), we know that the NMI decrease as partition numbers increase, but the performance is excellent and stable when mix is less than 0.75, and NMI will maintain at about 1.0. Our results show that Pinfomr is able to achieve better results in a shorter period of time, although accompanied by some loss of performance.

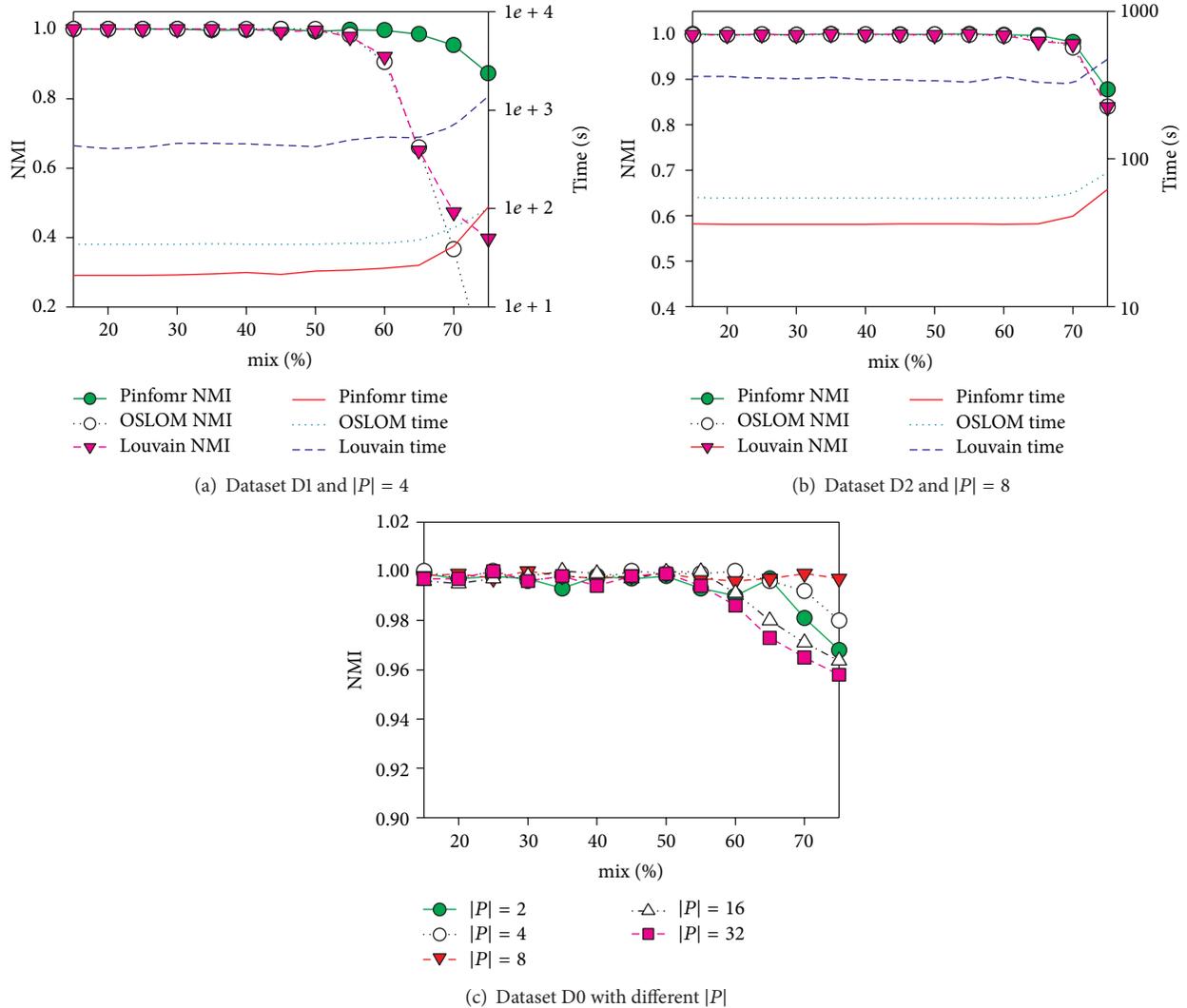


FIGURE 7: Accuracy and running time tests on different data sets.

4.2. Partitioning Experiment. In previous section, we have mentioned that the quality of partition will play a vital role in the final performance of parallel community detection. Therefore, we conduct experiment in this section to test the impact and effectiveness of different partitioning methods on Pinfomr.

We use two simple partitioning methods to compare with the improved multilevel k -way partitioning method. One is a sequential partitioning method dividing the network according to the storage order of the nodes and edges on the HDFS. The other one is a random matching partitioning method by randomly choosing nodes to generate a matching. For example, assuming that we bisect $G = (V, E)$ with $n = 20,000$ and $m = 300,000$ into $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, (1) when using sequential partitioning method, the first 10,000 nodes will be collected to form V_1 and links within V_1 will form E_1 . The other nodes are left for V_2 and links within V_2 form E_2 . (2) If we use the random matching method, we will randomly select 5,000 node pairs into V_1 and all links

within V_1 will form E_1 , and the process for G_2 is similar to G_1 . Dividing a connected network into subnetworks or partitions will cause edge loss. Excellent partitioning methods will always try to walk through the slits between communities and avoid cutting off the effective edges. Here, we use the data set $D1$ to test performance of different partitioning methods with $|P| = 2$ and $|P| = 4$. In Figures 8(a) and 8(b), we can see that multilevel k -way partitioning method is stable and results of Pinfomr on it are very close to the results of infomap and also very close to the real results. From Figures 8(c) and 8(d), we get that, for multilevel k -way partitioning method, total edge loss ratio increases linearly along with the increase of mix parameter. It is easy to understand that, from the meaning of the mix parameter, effective edge loss ratio always remains at low level before mix rising up to 0.70. Manifestations of sequential partitioning method and random partitioning method are also easy to explain. Distribution of edges of LFR generated networks is random and uniform, regardless of storage order. As a result, the total edge loss ratio will

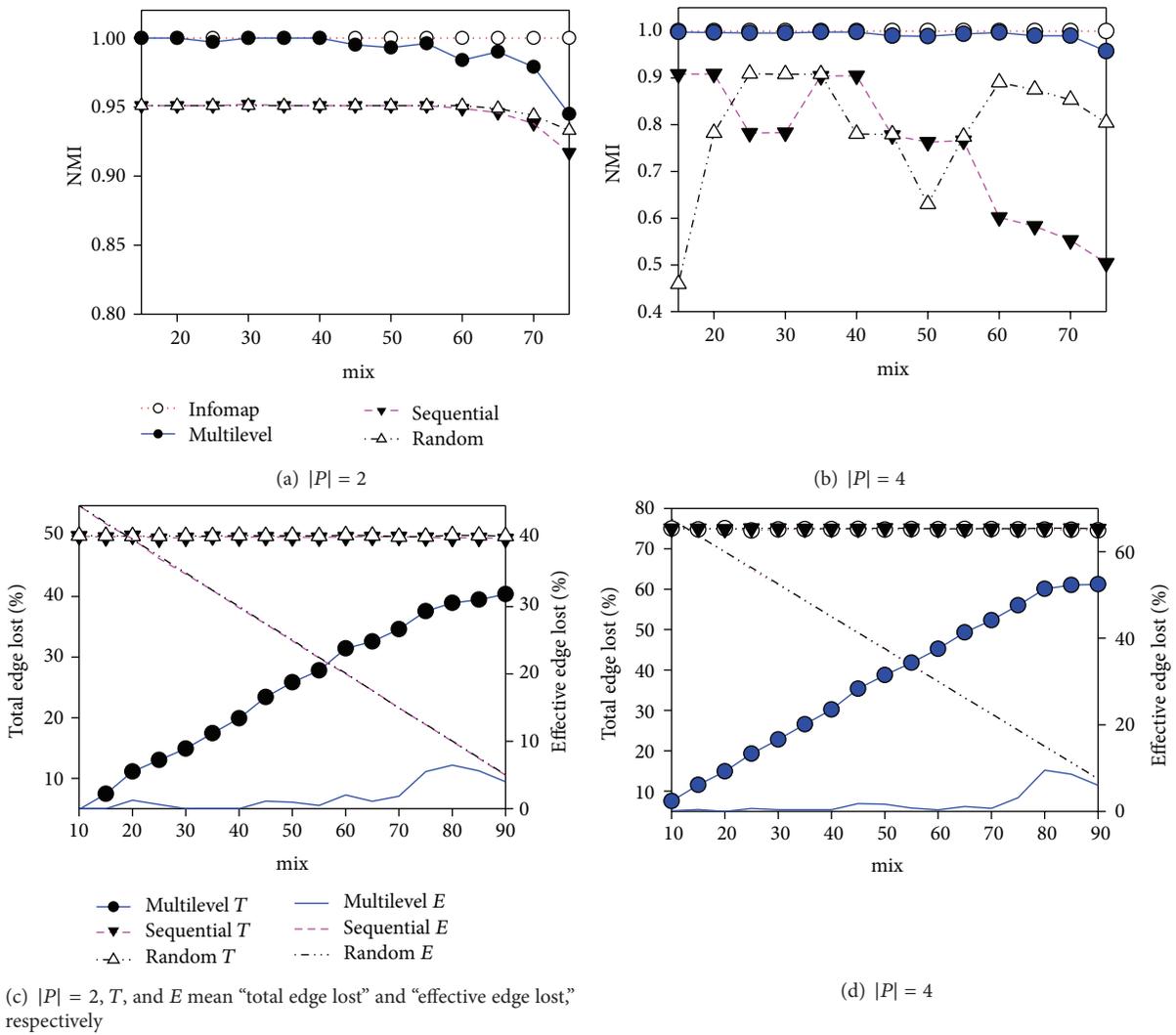


FIGURE 8: Performance and edge loss ratio of different partition methods on data set D3.

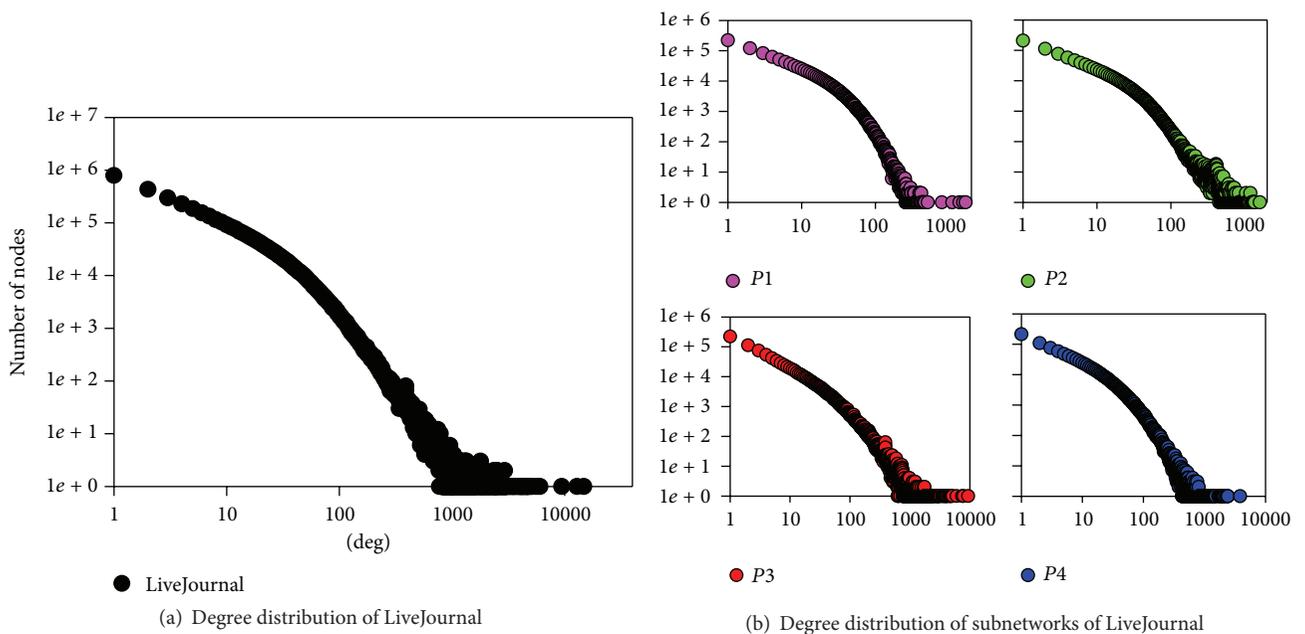


FIGURE 9: Degree distribution test on LiveJournal data set.

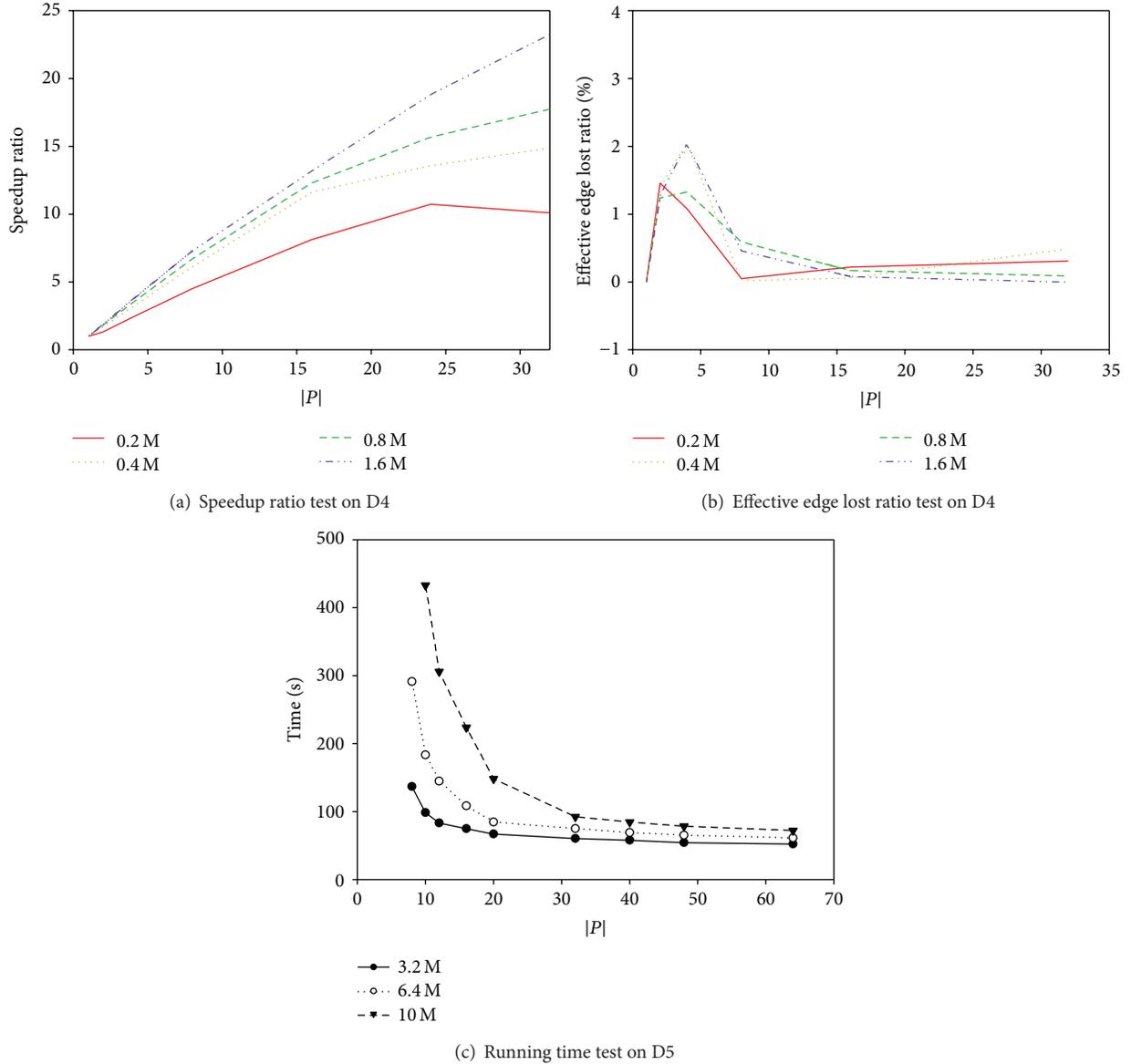


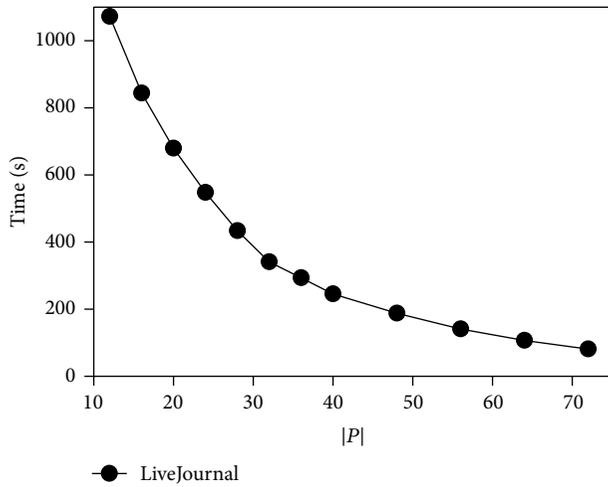
FIGURE 10: Scalability tests on data sets D4 and D5.

remain at about $(|P| - 1)/|P|$. Effective edge loss decreases linearly with the increase of mix for the same reason when total edge loss ratio increases linearly with the increase of mix of multilevel partitioning method in Figures 8(a) and 8(b).

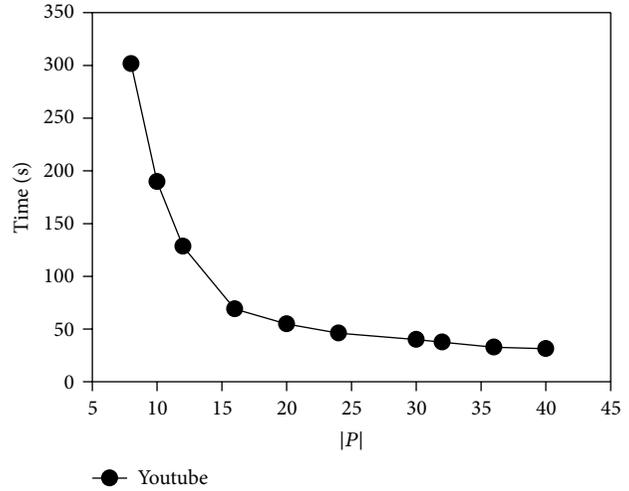
In addition, we conduct a degree distribution test on a real network-LiveJournal to verify performance of the improved multilevel partitioning method. The network is divided into 4 partitions by the improved multilevel k -way partitioning method, and the degree distributions corresponding to the original network and the 4 subnetworks are shown in Figure 9. Comparative observation indicates that the subnetworks got from the improved multilevel k -way partitioning method are able to maintain the degree distribution characteristics of the original network.

4.3. Scalability and Performance Experiment. Our study aims to uncover community structures in big social networks and improve resource utilization as much as possible. Here, we unify the two problems together by means of MapReduce. With a small portion of expense of performance, we will achieve the goal. In this section, we will test the scalability and performance of the parallel community detection method, and data sets used are D4, D5, LiveJournal (<http://snap.stanford.edu/data/com-LiveJournal.html>), Youtube (<http://snap.stanford.edu/data/as-skitter.html>), and Orkut (<http://snap.stanford.edu/data/com-Orkut.html>).

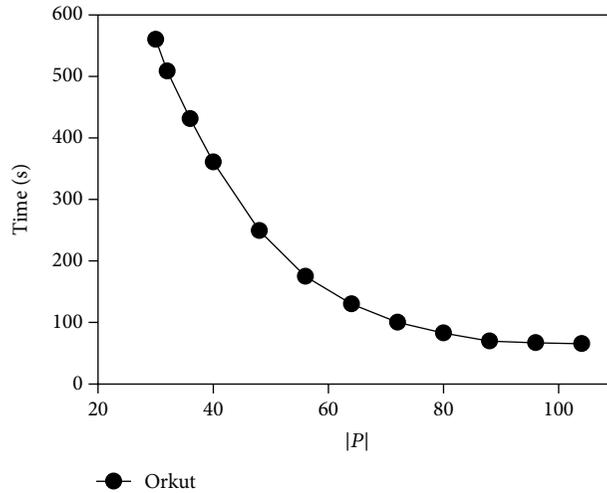
For a certain network in Figures 10(a) and 10(b), when $|P|$ increases, the speedup ratio will increase but the acceleration will become slow, since MapReduce needs some time to



(a) Scalability test on LiveJournal



(b) Scalability test on Youtube



(c) Scalability test on Orkut

FIGURE 11: Scalability tests on real networks.

initiate before map tasks start to run and transmit data from map phase to reduce phase. For a certain $|P|$, as network size grows, the speedup ratio will become higher. For Figure 10(c), we just present the running time of parallel community detection method on MapReduce because the capacity of the servers cannot deal with such large networks on one server or with $|P| = 1$.

Finally, we apply the same process onto the real networks. Experiments on real networks shown in Figure 11 also confirm that our parallel community detection method has excellent scalability. From the results, we can conclude that, when $|P|$ increases, the subnetwork size assigned to each map task will be smaller, and the total edge lost ratio will increase, which will further reduce the subnetwork size. From Figures 10 and 11, we can get the following: in the case of constant data size, the running time and $|P|$ are linear approximation when $|P|$ is small. When partition number is small, the running time is affected by the number of partitions

significantly. When the partition number $|P|$ reaches the “critical point” (Figure 11(c) Orkut, $|P| = 72$ and Figure 11(b) Youtube, $|P| = 20$), running time is less affected by the changes of partition number and shows “long tail effect” to some extent. The reason is that the cost of MapReduce is basically fixed. For a larger social network with the same number of map tasks, Mapreduces initial time accounts for a smaller proportion of the total running time. When partition number increases and the total running time decreases, the proportion of the initial time is not negligible. It makes our method exhibiting “long tail effect” in different data sets.

5. Conclusion

Community detection has become an important research topic in social networks. Traditional algorithms on community mining cannot effectively adapt to the current big social network scenarios [29, 30]. Infomap is excellent standalone

community detection method and, by means of multilevel k -way partitioning method enhanced by k -shell deposition, we are able to develop a parallel community discovery method on MapReduce framework. Related experiments verified the validity of the proposed work in this paper, and it may possess some reference meaning for social network analysis and social community mining with the big data techniques. Next, we will try to use some overlapping partitioning methods to further improve the community detection accuracy.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to express their sincere gratitude to Zhang Yuchao from Beijing Institute of System Engineering for providing great assistance through the entire research process, Lancichinetti A. from Amaral Lab of Northwestern University for supporting their work unselfishly with implementation of some algorithms, and Chen Siming from University of Illinois at Chicago for his careful review, comments, and feedback on this paper. In addition, this research is supported by the National High-Tech R&D Program of China (nos. 2012AA012600, 2012AA01A401, and 2012AA01A402), the National Natural Science Foundation of China (no. 61202362), the State Key Development Program of Basic Research of China (no. 2013CB329601), and Project funded by the China Postdoctoral Science Foundation (no. 2013M542560).

References

- [1] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [2] C. Lee and P. Cunningham, "Community detection: effective evaluation on large social networks," *Journal of Complex Networks*, vol. 2, no. 1, pp. 19–37, 2013.
- [3] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4–5, pp. 175–308, 2006.
- [5] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [6] M. E. J. Newman and E. A. Leicht, "Mixture models and exploratory analysis in networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 23, pp. 9564–9569, 2007.
- [7] G. R. Chen, X. F. Wang, and X. Li, *Introduction to Complex Networks: Models, Structures and Dynamics*, Higher Education Press, Beijing, China, 2012.
- [8] S. Cooper, "The Largest Social Networks in the World Include Some Big Surprises [Business Insider]," January 2014, <http://www.businessinsider.com/the-largest-social-networks-in-the-world-2013-12>.
- [9] T. N. Bui and C. Jones, "A heuristic for reducing fill-in in sparse matrix factorization," in *Proceedings of the 6th SIAM Conference on Parallel Processing for Scientific Computing (PPSC '93)*, pp. 445–452, 1993.
- [10] K. Andreev and H. Räcke, "Balanced graph partitioning," *Theory of Computing Systems*, vol. 39, no. 6, pp. 929–939, 2006.
- [11] A. Lancichinetti and S. Fortunato, "Community detection algorithms: a comparative analysis," *Physical Review E*, vol. 80, no. 5, Article ID 056117, 2009.
- [12] M. Rosvall and C. T. Bergstrom, "Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems," *PLoS ONE*, vol. 6, no. 4, Article ID e18209, 2011.
- [13] X. Ding, Advancement of operating system to manage critical resources in increasingly complex computer architecture. (Electronic Thesis or Dissertation), 2010, <https://etd.ohiolink.edu/>.
- [14] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [15] S. Jin, A. Li, S. Yang, W. Lin, B. Deng, and S. Li, "A MapReduce and information compression based social community structure mining method," in *Proceedings of the IEEE 16th International Conference on Computational Science and Engineering (CSE '13)*, pp. 971–980, IEEE, Sydney, Australia, December 2013.
- [16] M. Kitsak, L. K. Gallos, S. Havlin et al., "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.
- [17] D. Zhao, L. Li, H. Peng, Q. Luo, and Y. Yang, "Multiple routes transmitted epidemics on multiplex networks," *Physics Letters A*, vol. 378, no. 10, pp. 770–776, 2014.
- [18] D. Zhao, H. Peng, L. Li, Y. Yang, and S. Li, "An efficient patch dissemination strategy for mobile networks," *Mathematical Problems in Engineering*, vol. 2013, Article ID 896187, 13 pages, 2013.
- [19] G. Karypis and V. Kumar, "Analysis of multilevel graph partitioning," in *Proceedings of the ACM/IEEE Conference on Supercomputing*, p. 29, ACM, 1995.
- [20] L. Li, D. Alderson, J. C. Doyle, and W. Willinger, "Towards a theory of scale-free graphs: definition, properties, and implications," *Internet Mathematics*, vol. 2, no. 4, pp. 431–523, 2005.
- [21] A. Abou-Rjeili and G. Karypis, "Multilevel algorithms for partitioning power-law graphs," in *Proceedings of the Parallel and Distributed Processing Symposium*, p. 10, 2006.
- [22] B. W. Kernighan and S. Lin, "An efficient heuristic procedure for partitioning graphs," *The Bell System Technical Journal*, vol. 49, no. 2, pp. 291–307, 1970.
- [23] P. D. Grnwald, I. J. Myung, and M. A. Pitt, Eds., *Advances in Minimum Description Length: Theory and Applications*, MIT press, New York, NY, USA, 2005.
- [24] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.
- [25] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.
- [26] G. K. Orman, V. Labatut, and H. Cherifi, "Towards realistic artificial benchmark for community detection algorithms evaluation," *International Journal of Web Based Communities*, vol. 9, no. 3, pp. 349–370, 2013.
- [27] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, "Finding statistically significant communities in networks," *PLoS ONE*, vol. 6, no. 4, Article ID e18961, 2011.

- [28] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, Article ID P10008, 2008.
- [29] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [30] M. Li, W. Chen, and L. Han, "Correlation matching method for the weak stationarity test of LRD traffic," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 181–195, 2010.

Research Article

On the Propagation of Longitudinal Stress Waves in Solids and Fluids by Unifying the Navier-Lame and Navier-Stokes Equations

Ahmad Barzkar and Hojatollah Adibi

Department of Applied Mathematics, Faculty of Mathematics and Computer Sciences, Amirkabir University of Technology, 424 Hafez Avenue, P.O. Box 15875-4413, Tehran, Iran

Correspondence should be addressed to Hojatollah Adibi; adibih@aut.ac.ir

Received 2 September 2014; Accepted 14 December 2014

Academic Editor: Florin Pop

Copyright © 2015 A. Barzkar and H. Adibi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Propagation of mechanical waves' phenomenon is the result of infinitely small displacements of integrated individual particles in the materials. These displacements are governed by Navier-Lame and Navier-Stokes equations in solids and fluids, respectively. In the present work, a generalized Kelvin-Voigt model of viscoelasticity has been proposed with the aim of bridging the gap between solids and fluids leading to a new concept of viscoelasticity which unifies the Navier-Lame and the Navier-Stokes equations. On solving this equation in one dimension, propagation of stress disturbance in the so-called "Kelvin-Voigt materials" will be studied. The model of these materials involves all the elastic and viscoelastic solids, as well as fluids and soft materials.

1. Introduction

The subject matter of mechanics is the study of motion, in how a physical object changes position with time and why [1]. In continuum mechanics we are concerned with the mechanical behavior and shape of materials under load. The physical reasons for this behavior can be quite different for different materials. Solid media will deform when forces are applied on them. These materials are called elastic, if the object will return to its initial shape and size when these forces are removed. Hence elasticity is the tendency of solid materials to return to their original shape after being deformed [2]. In continuum mechanics we consider the basic equations describing the physical effects created by external forces acting upon solids and fluids. In addition to the basic equations that are applicable to all continua, there are other equations, called constitutive equations which are constructed to take into account material characteristics. In the study of solids the constitutive equations for a linear elastic material are a set of relations between stress and strain. Hook's law represents the material behavior and relates the Cauchy stress tensor $T = (\sigma_{ij})$ and infinitesimal strain

tensor $e = (e_{ij})$. The general form of Hook's law in components is

$$\sigma_{ij} = C_{ijkl}e_{lk}, \quad (1)$$

where C_{ijkl} is the fourth-order stiffness tensor [3]. If the body is isotropic and homogenous, then this law is simply written as [3]

$$T = \lambda\theta I + 2\mu e \quad (2)$$

in which θ is the trace of e , I is the unit tensor, and λ and μ are the Lamé constants. Perfect elasticity is an approximation of the real world and few materials remain purely elastic even after very small deformations. When an elastic material is not stressed in tension or compression beyond its elastic limit, its individual particles perform elastic movement. The displacement of the particles mass center, denoted by vector field $u = (u_1, u_2, u_3)$, is related to the strain tensor by the relation [4]

$$e = \frac{1}{2} [\text{grad}u + (\text{grad}u)^T]. \quad (3)$$

Newton's second law of motion is based on the conservation of momentum and expressed as [5]

$$\operatorname{div} T = \rho \frac{\partial^2 u}{\partial t^2} \quad (4)$$

in which ρ is the mass density of the body. Now, in an isotropic, homogenous elastic solid, by combining (2), (3), and (4) we have

$$u_{tt} = \frac{\lambda + \mu}{\rho} \operatorname{grad} \operatorname{div} u + \frac{\mu}{\rho} \nabla^2 u \quad (5)$$

which is the vector form of the Navier-Lame equations and governs the infinitesimal movements of the body's integrated individual particles [5].

Viscous materials resist shear flow and strain linearly with time, when stress is applied [6]. If a material exhibits a linear response, it is categorized as a Newtonian material [6]. In this case the stress is linearly proportional to the strain rate. If the material exhibits a nonlinear response to the strain rate, it is categorized as non-Newtonian fluid. To get the equation that governs the small movement of integrated particles of fluids, let us first consider the continuity equation, governing the continuity of the integrated individual particles which are derived from the principle of conservation of mass and is given as

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho V) = 0, \quad (6)$$

where the vector field V represents the velocity of the particles and ρ is the mass density of the fluid [7]. In the absence of body forces, conservation of momentum, on which Newton's second law is based, is expressed as [7]

$$\operatorname{div} T = \rho \frac{\partial V}{\partial t}. \quad (7)$$

In the study of fluids, the constitutive equations consist of a set of relations between stress and rate of strain; these equations in tensor form are given as [8]

$$T = -pI + \lambda \operatorname{tr}(D)I + 2\mu D, \quad (8)$$

where the pressure p is induced by tension, that is, the difference between the dynamical and thermodynamical pressures, λ and μ are independent parameters characterizing viscosity, I is the unit tensor, and D is the rate of strain which is a symmetric tensor of order two.

Finally the kinematical relation of fluids describing the relation between tensor D and velocity vector field V is given as

$$D = \frac{1}{2} [\operatorname{grad} V + (\operatorname{grad} V)^T]. \quad (9)$$

Moreover, the pressure can be written as [8]

$$p = -\left(\lambda + \frac{2}{3}\mu\right) \operatorname{tr}(D). \quad (10)$$

It can be deduced from (7), (8), (9), and (10) that

$$\frac{\partial V}{\partial t} = \frac{\lambda + \mu}{\rho} \operatorname{grad}(\operatorname{div} V) + \frac{\mu}{\rho} \nabla^2 V. \quad (11)$$

However, in terms of displacement vector field u ,

$$u_{tt} = \frac{\lambda + \mu}{\rho} \operatorname{grad} \operatorname{div} \dot{u} + \frac{\mu}{\rho} \nabla^2 \dot{u} \quad (12)$$

which is the set of Navier-Stokes equations.

In fluid continuum, the motion of substances is described by the Navier-Stokes equations. These equations are strictly the statement of conservation of momentum and are based on the assumption that the fluid, at the scale of interest, is a continuum. In other words, it is not made up of discrete particles but rather a continuous substance. Another necessary assumption is that all the fields of interest like pressure, velocity, density, and temperature are differentiable weakly at least [9].

There are materials for which a suddenly applied and maintained state of uniform stress induces an instantaneous deformation followed by a flow process which may or may not be limited in magnitude as time grows [10]. These materials exhibit both solid and fluid characteristics. Behavior of these materials clearly cannot be described by either elasticity or viscosity theories alone, as it combines features of each and is called viscoelastic. Viscoelasticity is a generalization of elasticity and viscosity [11].

All materials exhibit some viscoelastic response. In common metals such as steel or aluminum, as well as in quartz, at room temperature, and at small strain, the behavior does not deviate much from linear elasticity, which is the simplest response of a viscoelastic material. Synthetic polymers, wood, and human tissue as well as metals at high temperature display significant viscoelastic effects. Some phenomena in viscoelastic materials are as follows.

- (i) If the stress is held constant, the strain increases with time (creep).
- (ii) If the strain is held constant, the stress decreases with time (relaxation).
- (iii) Acoustic waves experience attenuation.

The material creeps, that gives the prefix visco-, and the material fully recovers, which gives the suffix-elasticity [12]. The viscoelastic materials play an important role in many engineering structures. Those materials such as polymers are being used, for example, to dissipate and to insulate vibration caused by rotating or reciprocal movements [13]. They also have potential application in a new Hopkinson pressure bar testing apparatus [14–18]. Therefore, having the knowledge of the behavior of these materials, particularly related to their mechanical parameters, is essential. The theory which illustrates this behavior is the viscoelasticity theory. This theory is used in many fields, such as solid mechanics, seismology, exploration geophysics, acoustics, and engineering [19].

Modeling and model parameter estimation are of great importance for a correct prediction of the foundation behavior [19]. In many cases, elastic constitutive models work well

when time dependent effects can be neglected. However in those cases when time dependent effects cannot be neglected, we will need to utilize different constitutive models. Basically, time dependent effects indicate that the stress-strain behavior of material will change with time. The elastic material model for time dependent effects is viscoelasticity. Many researchers as Alfrey [20], Barberán and Herrera [21], Achenbach and Reddy [22], Bhattacharya and Sengupta [23], and Acharya et al. [24] formulated and developed this theory [19]. Further, Bert and Egle [25], Abd-Alla and Ahmed [26], and Batra [27] successfully applied this theory to wave propagation in homogenous, elastic media [19]. Murayama and Shibata [28] and Schiffman et al. [29] have proposed higher order viscoelastic models of five and seven parameters to represent the soil behavior. Jankowski et al. [30] discussed the linear viscoelastic model and the nonlinear viscoelastic model [19].

One of the most important classic models which is focused on, in this paper is Kelvin-Voigt model. The constitutive relation of this model is expressed as a linear first-order differential equation, which can be derived as below.

The most simple one-dimensional stress and strain tensors are of the form

$$T = \begin{bmatrix} \sigma & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad e = \begin{bmatrix} e_1 & 0 & 0 \\ 0 & e_2 & 0 \\ 0 & 0 & e_3 \end{bmatrix} \quad (13)$$

which, according to the generalized Hook's law, are related by

$$\sigma = Ee_1, \quad (14)$$

where E is Young's modulus of elasticity [3]. In the Kelvin-Voigt model for viscoelastic materials, (14) is augmented to include viscosity, leading to the generalized equation

$$\sigma = Ee_1 + \eta \frac{\partial e_1}{\partial t}, \quad (15)$$

where η is the constant of viscosity [31, 32].

This model represents a solid undergoing reversible, viscoelastic strain and is extremely good with modeling creep in materials, but with regard to relaxation the model is much less accurate [33].

All material substances are comprised of particles. When a material is not stressed in tension or compression beyond its elastic limit, its individual particles may be forced into vibrational motion about their equilibrium positions. Thus, these particles perform elastic oscillations. Elastic waves are focused on particles that move in unison to produce a mechanical wave. In solids, elastic waves can propagate in four principle modes that are based on the way the particles oscillate. These waves can propagate as longitudinal, shear, and surface waves and in the thin materials as plate waves.

In longitudinal waves, the oscillations occur in the longitudinal direction or the direction of wave propagation. Since compressional and dilatational forces are active in these waves, they are also called pressure or compressional waves. They are also sometimes called density waves because their particle density fluctuates as they move. Compression waves can be generated in liquids, as well as solids, because

the energy travels through the atomic structure by a series of compressions and expansion movements [34]. In the transverse or shear wave, the particles oscillate at a right angle or perpendicular to the direction of propagation. Shear waves require a solid material for effective propagation and therefore are not effectively propagated in materials such as liquids or gasses [34]. Waves in an isotropic elastic solid are governed by the vector Navier-Lame equations.

2. New Mathematical Model

2.1. Generalization of Kelvin-Voigt Model. In Kelvin-Voigt model for viscoelastic materials, the viscosity term is augmented to (14), leading to the generalized equation (15). But this model is a generalization of one-dimensional constitutive equation (14). In order to obtain a model for viscoelastic materials in most general form, the constitutive equation (2) should be extended. In this paper this equation is rewritten as

$$T = \lambda \theta I + 2\mu e + \lambda_f \dot{\theta} I \quad (16)$$

in which λ_f is called the first constant of viscosity. Also by symmetric of (16) this equation can be written as

$$T = \lambda \theta I + 2\mu e + \lambda_f \dot{\theta} I + 2\mu_f \dot{e} \quad (17)$$

in which μ_f is the second constant of viscosity. In the simple case (13), (16) is reduced as

$$\sigma = Ee_1 + \frac{\mu\lambda_f}{\lambda + \mu} \dot{e}_1 - \frac{\mu\alpha}{\lambda + \mu} \exp\left(-\frac{\lambda + \mu}{\lambda_f} t\right) \quad (18)$$

in which α is a positive constant. By comparing (18) and (15), it can be easily inferred that

$$\eta = \frac{\mu\lambda_f}{\lambda + \mu}. \quad (19)$$

In the next sections, it has been shown that (16) and (17) can justify the physics.

2.2. Generalization of Navier-Lame Equation. In the first section, the Navier-Lame equation was deduced by combining (2), (3), and (4). Now (2) is replaced by (16); therefore, by combining (16), (3), and (4), a generalized Navier-Lame equation is obtained as follows:

$$u_{tt} = \frac{\lambda + \mu}{\rho} \text{grad div } u + \frac{\mu}{\rho} \nabla^2 u + \frac{\lambda_f}{\rho} \text{grad div } \dot{u}, \quad (20)$$

where $\dot{u} = \partial u / \partial t$. Also if (2) is replaced by (17), then, by combining (17), (3), and (4), another generalization of Navier-Lame equation can be deduced as follows:

$$u_{tt} = \left(\frac{\lambda + \mu}{\rho} \text{grad div } u + \frac{\mu}{\rho} \nabla^2 u \right) + \left(\frac{\lambda_f + \mu_f}{\rho} \text{grad div } \dot{u} + \frac{\mu_f}{\rho} \nabla^2 \dot{u} \right), \quad (21)$$

where λ_f and μ_f resemble the form of equations for λ and μ , respectively.

2.3. *Unifying Navier-Lame and Navier-Stokes Equations.* At this point, by comparing (21) and (12), it is apparent that (21) can be written in two parts. The first and second parts are concerned with the solid and fluid properties of the body, respectively. In fact, (21) governs the displacements of the particles in both solids and fluids and in bodies with both solid and fluid properties, but not in materials involving two or more different phases. Equation (21) is a unification of Navier-Lame and Navier-Stokes equations; therefore, (16) and (17) can be used as constitutive equation to all viscoelastic materials.

3. Longitudinal Waves in Materials

Equation (21) governs displacement of integrated individual particles of any continuum body. To study the propagation of longitudinal stress waves in a body, we note that (21) can be written in one dimension as follows:

$$u_{tt} = c^2 u_{xx} + d^2 u_{txx}, \quad \infty < x < +\infty, \quad t \geq 0 \quad (22)$$

in which

$$c^2 = \frac{\lambda_s + 2\mu_s}{\rho}, \quad d^2 = \frac{\lambda_f + 2\mu_f}{\rho}, \quad (23)$$

where the only initial condition is $u(0, 0) = 0$. This equation can be solved by separation of variables method, so $u(x, t) = X(x)T(t)$. Using this in (22) leads to

$$\frac{X}{X''} = c^2 \frac{T}{T''} + d^2 \frac{T'}{T''} = \gamma \quad (24)$$

in which γ is constant. Now, the following five cases are considered.

Case 1 (γ is positive). In this case one can assume $\gamma = 1/k^2$, $k \neq 0$; hence, (24) becomes

$$\begin{cases} X'' - k^2 X = 0 \\ T'' - k^2 d^2 T' - k^2 c^2 T = 0 \end{cases} \quad (25)$$

$$\xrightarrow{\text{yields}} \begin{cases} X(x) = c_1 e^{kx} + c_2 e^{-kx} \\ T(t) = d_1 e^{r_1 t} + d_2 e^{r_2 t} \end{cases}$$

in which

$$\begin{aligned} r_1 &= \frac{1}{2} \left(k^2 d^2 + k \sqrt{k^2 d^4 + 4c^2} \right), \\ r_2 &= \frac{1}{2} \left(k^2 d^2 - k \sqrt{k^2 d^4 + 4c^2} \right). \end{aligned} \quad (26)$$

The solution, therefore, can be written as

$$u(x, t) = (c_1 e^{kx} + c_2 e^{-kx}) (d_1 e^{r_1 t} + d_2 e^{r_2 t}). \quad (27)$$

By applying the initial condition, one arrives at

$$\begin{aligned} u(0, 0) \\ = (c_1 + c_2)(d_1 + d_2) = 0 \longrightarrow c_2 = -c_1 \quad \text{or} \quad d_2 = -d_1. \end{aligned} \quad (28)$$

If $c_2 = -c_1$, then the velocity of the particle, placed at the origin, is zero, since

$$u_t(0, t) = (c_1 + c_2)(d_1 r_1 e^{r_1 t} + d_2 r_2 e^{r_2 t}). \quad (29)$$

Therefore, in this case, there is no motion in the medium.

On the other hand, since the medium is assumed to be elastic, the movement of the particle, located at the origin, stops after a while. Hence, if $d_2 = -d_1$, at time T , then (29) can be written as

$$u_t(0, T) = d_1 (c_1 + c_2)(r_1 e^{r_1 T} - r_2 e^{r_2 T}) = 0. \quad (30)$$

Since r_1 is positive and r_2 is negative, the second part of (30) is not zero and so $c_2 = -c_1$. This means that there is no motion in the medium. Therefore, in this case, the equation does not have a nontrivial solution.

Case 2. If $\gamma = 0$, then the solution is trivial.

Case 3. If γ is negative and $k^2 d^4 - 4c^2$ is positive, then one can assume $\gamma = -1/k^2$, $k \neq 0$, so

$$\begin{cases} X'' + k^2 X = 0 \\ T'' + k^2 d^2 T' + k^2 c^2 T = 0 \end{cases} \quad (31)$$

$$\xrightarrow{\text{yields}} \begin{cases} X(x) = c_1 \cos kx + c_2 \sin kx \\ T(t) = d_1 e^{r_1 t} + d_2 e^{r_2 t} \end{cases}$$

in which

$$\begin{aligned} r_1 &= \frac{1}{2} \left(-k^2 d^2 + k \sqrt{k^2 d^4 - 4c^2} \right), \\ r_2 &= \frac{1}{2} \left(-k^2 d^2 - k \sqrt{k^2 d^4 - 4c^2} \right). \end{aligned} \quad (32)$$

Thus the solution is as follows:

$$u(x, t) = (c_1 \cos kx + c_2 \sin kx)(d_1 e^{r_1 t} + d_2 e^{r_2 t}). \quad (33)$$

Also, the velocity is

$$u_t(x, t) = (c_1 \cos kx + c_2 \sin kx)(r_1 d_1 e^{r_1 t} + r_2 d_2 e^{r_2 t}). \quad (34)$$

Now, from the initial condition, one obtains

$$u(0, 0) = c_1 (d_1 + d_2) = 0 \xrightarrow{\text{yields}} c_1 = 0, \quad \text{or} \quad d_2 = -d_1. \quad (35)$$

If $c_1 = 0$, then $u_t(0, t) = 0$, which means that there is no motion and the solution is trivial. Otherwise, $d_2 = -d_1$, which yields

$$u_t(0, t) = c_1 d_1 (r_1 e^{r_1 t} - r_2 e^{r_2 t}). \quad (36)$$

Since after a time T the particle that is at the origin stops, then

$$r_1 e^{r_1 T} - r_2 e^{r_2 T} = 0 \xrightarrow{\text{yields}} T = \frac{1}{r_1 - r_2} \ln \left(\frac{r_2}{r_1} \right). \quad (37)$$

Consequently, the solution can be written as

$$u(x, t) = d_1(c_1 \cos kx + c_2 \sin kx)(e^{r_1 t} - e^{r_2 t}). \quad (38)$$

On the other hand, it can be written as follows:

$$u(x, t) = (\alpha_1 \cos kx + \alpha_2 \sin kx)(e^{r_1 t} - e^{r_2 t}). \quad (39)$$

Case 4 (γ is negative and $k^2 d^4 - 4c^2 = 0$). In this case the solution is trivial.

Case 5 (γ is negative and $k^2 d^4 - 4c^2 < 0$). In this case one can assume $\gamma = -1/k^2$, $k \neq 0$, so that

$$\begin{cases} X'' + k^2 X = 0 \\ T'' + k^2 d^2 T' + k^2 c^2 T = 0 \end{cases} \quad (40)$$

$$\xrightarrow{\text{yields}} \begin{cases} X(x) = c_1 \cos kx + c_2 \sin kx \\ T(t) = e^{-(k^2 d^2 t)/2} (d_1 \cos rt + d_2 \sin rt) \end{cases}$$

in which

$$r = \frac{1}{2} k \sqrt{4c^2 - k^2 d^4}. \quad (41)$$

Hence, the solution is

$$u(x, t) = e^{-(k^2 d^2 t)/2} (c_1 \cos kx + c_2 \sin kx) \times (d_1 \cos rt + d_2 \sin rt). \quad (42)$$

Also, the velocity is given as

$$u_t(x, t) = e^{-(k^2 d^2 t)/2} (c_1 \cos kx + c_2 \sin kx) \times \left[\left(r d_2 - \frac{1}{2} k^2 d^2 d_1 \right) \cos rt - \left(r d_1 + \frac{1}{2} k^2 d^2 d_2 \right) \sin rt \right]. \quad (43)$$

From the initial condition one obtains

$$u(0, 0) = c_1 d_1 = 0 \xrightarrow{\text{yields}} c_1 = 0, \quad \text{or} \quad d_1 = 0. \quad (44)$$

If $c_1 = 0$, then $u_t(0, t) = 0$, and hence there is no motion in the medium and the solution is trivial. Therefore, $c_1 \neq 0$, and d_1 must be zero, so that

$$u_t(0, t) = e^{-(k^2 d^2 t)/2} c_1 d_2 \left[r \cos rt - \frac{1}{2} k^2 d^2 \sin rt \right]. \quad (45)$$

Since after a time T the particle which is at the origin stops, then

$$\begin{aligned} r \cos rT - \frac{1}{2} k^2 d^2 \sin rT &= 0 \\ \longrightarrow T_n &= \frac{1}{r} \left(\tan^{-1} \left(\frac{2r}{k^2 d^2} \right) + n\pi \right). \end{aligned} \quad (46)$$

Consequently the solution can be written as

$$u(x, t) = 2c_1 d_2 \sqrt{c_1^2 + c_2^2} \times e^{-(k^2 d^2 t)/2} \cos(kx - \varphi) \sin(rt) \quad (47)$$

in which

$$\varphi = \cos^{-1} \left(\frac{c_1}{\sqrt{c_1^2 + c_2^2}} \right). \quad (48)$$

Therefore, the solution can be written as

$$u(x, t) = \beta e^{-(k^2 d^2 t)/2} \cos(kx - \varphi) \sin(rt). \quad (49)$$

4. Results

4.1. Longitudinal Waves in a Perfectly Elastic Solid. In a metal like aluminum, whose characteristics are listed in Table 1, the wave that propagates in the medium is of the form (49).

If one considers $k = 2$, $\beta = 0.003$, and $\varphi = 1$, the depiction of the longitudinal wave that propagates from the origin is shown in Figure 1. It is clear that it is a wave with no attenuation.

4.2. Longitudinal Waves in Viscoelastic Solids. If the viscoelastic properties are characterized as in Table 2, the depiction of the longitudinal stress wave is shown in Figure 2. It is also clear that this wave is attenuated.

4.3. Longitudinal Waves in a Viscoelastic Newtonian Fluid. In a liquid with parameters shown in Table 3, the wave propagating in the medium is of the form (39). So then the wave characteristics are discussed in the following for material with properties as mentioned in Table 3.

If one assumes $k = 2$, $\alpha = 0.01$, and $\beta = 0.003$, then the displacement graph of particles in this medium is as shown in Figure 3. In Figure 3(a) the displacement of particles at the origin without returning to their initial position and in Figure 3(b) the displacement of a particle in each position x without returning to its initial position are seen.

4.4. Longitudinal Waves in an Ideal Newtonian Fluid. If a liquid is viscoelastic with viscosity coefficients shown in Table 4, the graph of longitudinal stress wave in this medium is as shown in Figure 4. In Figure 4(a), the distance between the particle originally located in the origin and the same particle moved away from the origin in time T is seen. This distance starts to decrease with the particles returning to the origin. In Figure 4(b), the distance between the particle at the time T and the same particle approaching its equilibrium position at the origin is shown; in Figure 4(c) the distance between the particles located in their respective equilibrium position and the same particles moved away from the equilibrium position in time T is shown. This distance then starts to decrease with the particles returning to their respective equilibrium position.

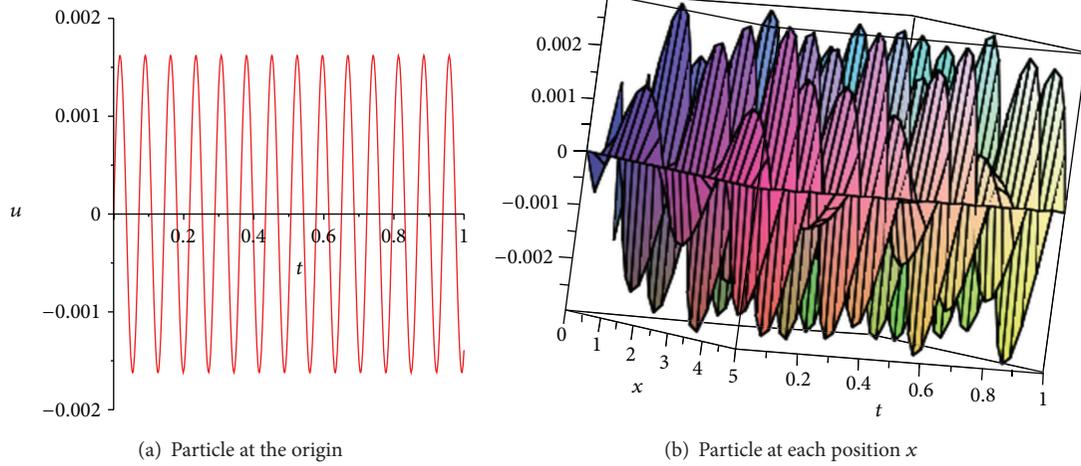


FIGURE 1: The vibration of particles.

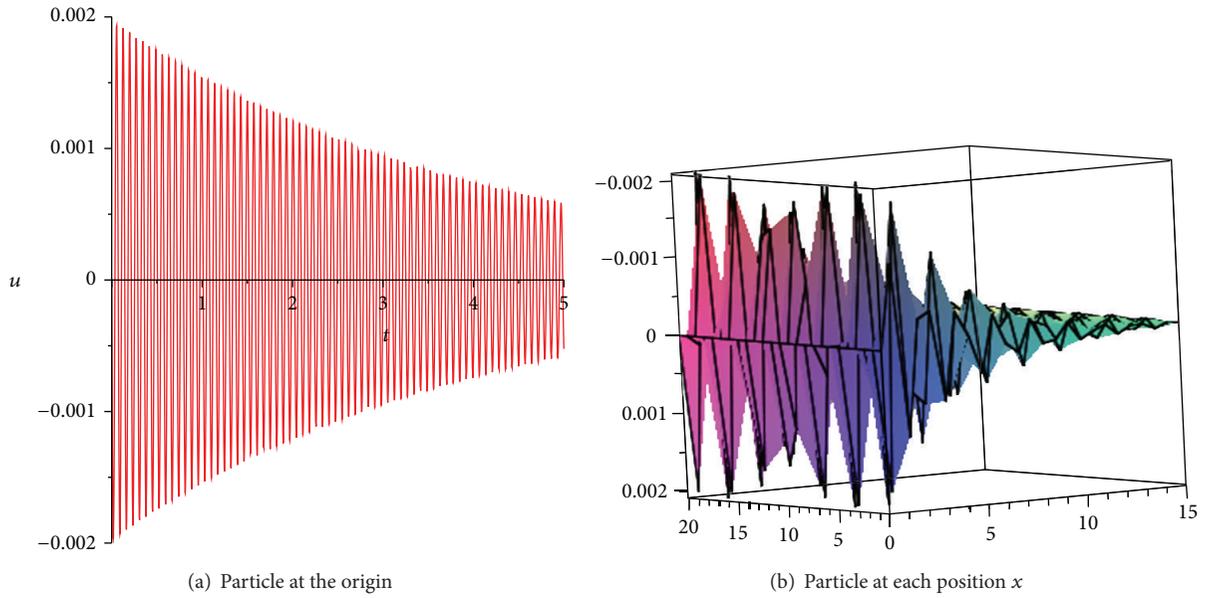


FIGURE 2: Vibration of a particle with attenuation.

TABLE 1: Parameters of aluminum that is considered to be perfectly elastic.

λ_s	λ_f	μ_s	μ_f	ρ
$50.6 \times 10^6 \text{ KN/m}^2$	0	$27.6 \times 10^6 \text{ KN/m}^2$	0	56000 Kg/m^3

TABLE 2: Parameters of aluminum with fluid properties pertaining to viscoelasticity.

λ_s	λ_f	μ_s	μ_f	ρ
$50.6 \times 10^6 \text{ KN/m}^2$	3016	$27.6 \times 10^6 \text{ KN/m}^2$	2130	56000 Kg/m^3

TABLE 3: Parameters of a Newtonian liquid like water which is perfectly fluid with no solid parameters.

λ_s	λ_f	μ_s	μ_f	ρ
0	30.12×10^2	0	21.3×10^2	996.3 Kg/m^3

TABLE 4: Parameters of a liquid like water which is not perfectly fluid and has solid parameters.

λ_s	λ_f	μ_s	μ_f	ρ
50.6	30.12×10^2	27.6	21.3×10^2	996.3 Kg/m^3

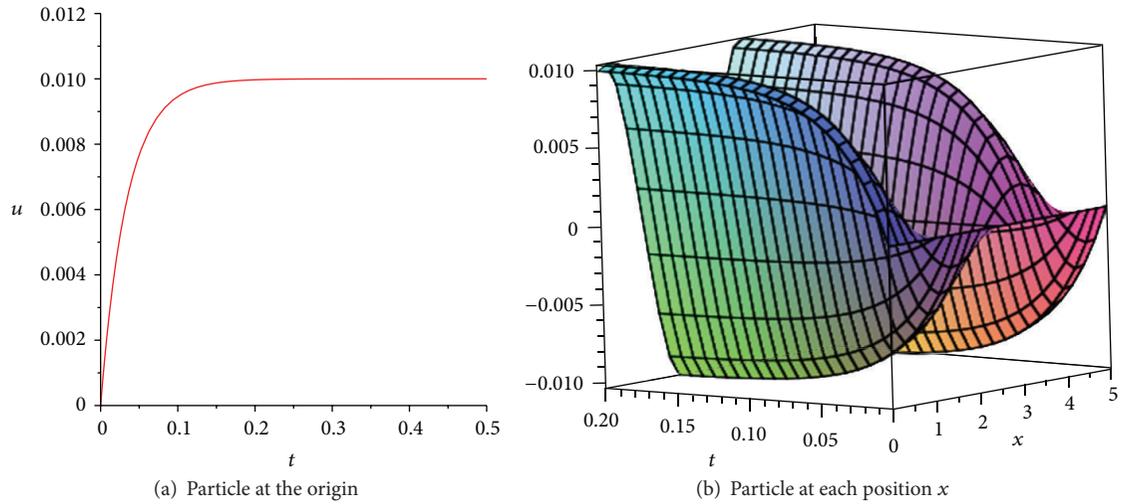
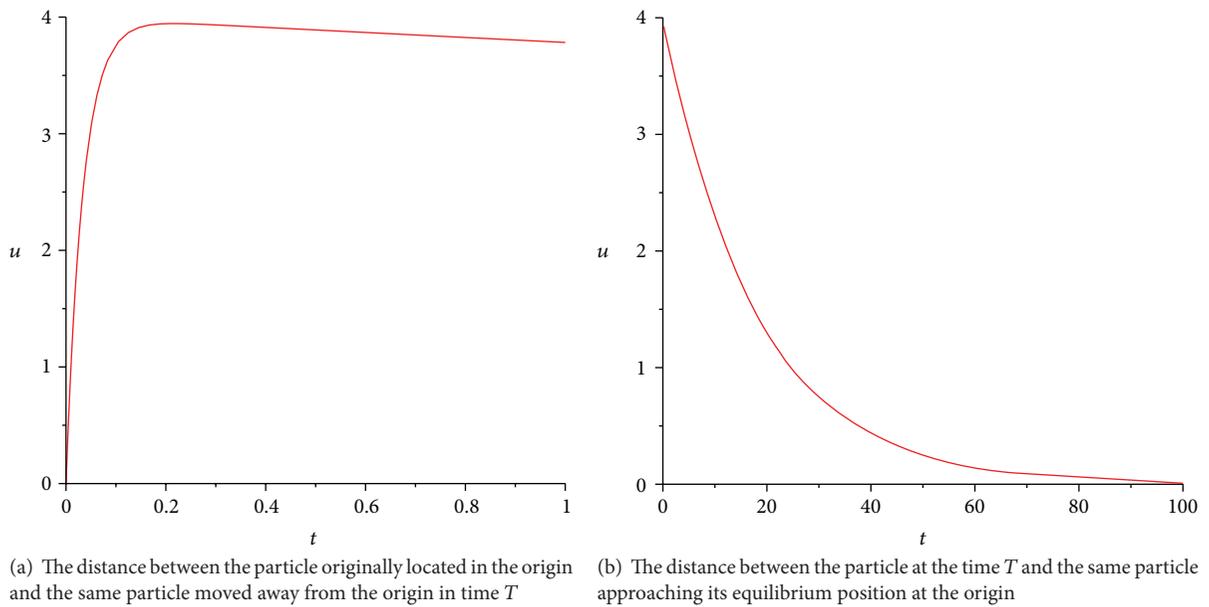
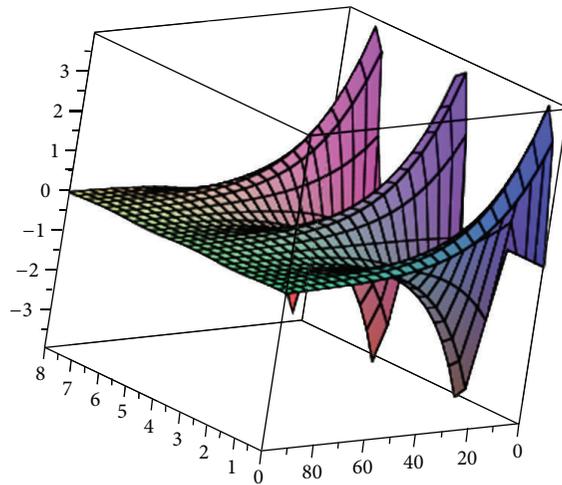


FIGURE 3: The displacement of particles.



(a) The distance between the particle originally located in the origin and the same particle moved away from the origin in time T (b) The distance between the particle at the time T and the same particle approaching its equilibrium position at the origin



(c) The distance between the particles located in their respective equilibrium position and the same particles moved away from the equilibrium position in time T

FIGURE 4: The distance between the particles at the time T .

5. Conclusions

When a material is subjected to an impulsive force, there occur infinitesimal displacements of the integrated particles of that body. In solids and in fluids the Navier-Lame and the Navier-Stokes equations govern these displacements, respectively. In the present work it was shown that through the use of a generalized form of Kelvin-Voigt model of viscoelasticity the unification of these two equations is possible and one obtains an equation that can represent solid materials, fluids, and soft materials. A new concept of viscoelasticity was defined in which every material has some degree of both solid and fluid properties depending on the particular parameters of viscosity. This is of particular importance in relation to the soft materials. Using this generalized equation, propagation of stress disturbance, pulse, and waves in solids as well as fluids can be studied in one single frame.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] K. Hackl and M. Goodarzi, *An Introduction to Linear Continuum Mechanics*, Ruhr-University Bochum, 2010.
- [2] L. Treloar, *The Physics of Rubber Elasticity*, Clarendon Press, Oxford, UK, 1975.
- [3] N. Muskhilishvili, *Some Basic Problem of the Mathematical Theory of Elasticity*, Noordhoff, Leiden, The Netherlands, 1975.
- [4] C. Truesdell, *Mechanics of Solids*, vol. 2, Springer, New York, NY, USA, 1972.
- [5] J. Allard, *Propagation of Sound in Porous Media*, Elsevier Science, Essex, UK, 1993.
- [6] M. A. Meyers and K. K. Chawla, *Mechanical Behavior of Materials*, Cambridge University Press, London, UK, 2008.
- [7] C.-e. Foias, *Navier-Stokes Equation and Turbulence*, Cambridge University, Cambridge, UK, 2001.
- [8] "Overheads," 2014, http://www.ucd.ie/mecheng/ccm/Overheads_6,7.pdf.
- [9] "Derivation of Navier Stokes equations," 2014, http://en.wikipedia.org/wiki/Derivation_of_the_Navier%E2%80%93Stokes_equations.
- [10] R. M. Christensen, *Theory of Viscoelasticity*, Academic Press, New York, NY, USA, 1971.
- [11] A. Musa, "Predicting wave velocity and stress of impact of elastic slug and viscoelastic rod using viscoelastic discontinuity analysis (standard linear solid model)," in *Proceedings of the World Congress on Engineering*, London, UK, 2012.
- [12] N. McCrum, C. Buckley, and C. Bucknall, *Principles of Polymer Engineering*, Oxford University Press, Oxford, UK, 1997.
- [13] A. D. Drozdov and A. Dorfmann, "The nonlinear viscoelastic response of carbon black-filled natural rubbers," *International Journal of Solids and Structures*, vol. 39, no. 23, pp. 5699–5717, 2002.
- [14] H. Zhao and G. Gary, "A three dimensional analytical solution of the longitudinal wave propagation in an infinite linear viscoelastic cylindrical bar. Application to experimental techniques," *Journal of the Mechanics and Physics of Solids*, vol. 43, no. 8, pp. 1335–1348, 1995.
- [15] H. Zhao, G. Gary, and J. R. Klepaczko, "On the use of a viscoelastic split Hopkinson pressure bar," *International Journal of Impact Engineering*, vol. 19, no. 4, pp. 319–330, 1997.
- [16] C. Salisburg, *Wave propagation through a polymeric Hopkinson bar [M.S. thesis]*, Department of Mechanical Engineering, University of Waterloo, Waterloo, Canada, 2001.
- [17] A. Benatar, D. Rittel, and A. L. Yarin, "Theoretical and experimental analysis of longitudinal wave propagation in cylindrical viscoelastic rods," *Journal of the Mechanics and Physics of Solids*, vol. 51, no. 8, pp. 1413–1431, 2003.
- [18] D. T. Casem, W. Fournery, and P. Chang, "Wave separation in viscoelastic pressure bars using single-point measurements of strain and velocity," *Polymer Testing*, vol. 22, no. 2, pp. 155–164, 2003.
- [19] K. Rajneesh and K. Kanwaljeet, "Mathematical analysis of five parameter model on the propagation of cylindrical shear waves in non-homogeneous viscoelastic media," *International Journal of Physical and Mathematical Science*, vol. 4, no. 1, pp. 45–52, 2013.
- [20] T. Alfrey, "Non-homogeneous stresses in visco-elastic media," *Quarterly of Applied Mathematics*, vol. 2, pp. 113–119, 1944.
- [21] J. Barberán and I. Herrera, "Uniqueness theorems and the speed of propagation of signals in viscoelastic materials," *Archive for Rational Mechanics and Analysis*, vol. 23, pp. 173–190, 1966.
- [22] J. D. Achenbach and D. P. Reddy, "Note on wave propagation in linearly viscoelastic media," *Zeitschrift für angewandte Mathematik und Physik ZAMP*, vol. 18, no. 1, pp. 141–144, 1967.
- [23] S. Bhattacharya and P. Sengupta, "Disturbances in a general viscoelastic medium due to impulsive forces on a spherical cavity," *Gerlands Beiträge zur Geophysik*, vol. 87, no. 8, pp. 57–62, 1978.
- [24] D. P. Acharya, I. Roy, and P. K. Biswas, "Vibration of an infinite inhomogeneous transversely isotropic viscoelastic medium with cylindrical hole," *Applied Mathematics and Mechanics (English Edition)*, vol. 29, no. 3, pp. 367–378, 2008.
- [25] C. Bert and D. Egle, "Wave propagation in a finite length bar with variable area of cross-section," *Journal of Applied Mechanics*, vol. 36, pp. 908–909, 1969.
- [26] A. M. Abd-Alla and S. M. Ahmed, "Rayleigh waves in an orthotropic thermoelastic medium under gravity field and initial stress," *Earth, Moon, and Planets*, vol. 75, no. 3, pp. 185–197, 1996.
- [27] R. C. Batra, "Linear constitutive relations in isotropic finite elasticity," *Journal of Elasticity*, vol. 51, pp. 269–273, 1998.
- [28] S. Murayama and T. Shibata, "Rheological properties of clays," in *Proceedings of the 5th International Conference of Soil Mechanics and Foundation Engineering*, Paris, France, 1961.
- [29] R. Schiffman, C. Ladd, and A. Chen, "The secondary consolidation of clay, rheology and soil mechanics," in *Proceedings of the International Union of Theoretical and Applied Mechanics Symposium*, Berlin, Germany, 1964.
- [30] R. Jankowski, K. Wilde, and Y. Fujino, "Pounding of superstructure segments in isolated elevated bridge during earthquakes," *Earthquake Engineering and Structural Dynamics*, vol. 27, no. 5, pp. 487–502, 1998.

- [31] V. N. Nikolaevskij, *Mechanics of Porous and Fractured Media*, World Scientific, 1990.
- [32] N. Thin, *Understanding Viscoelasticity*, Springer, Heidelberg, Germany, 2002.
- [33] “viscoelasticity,” 2014, <http://en.wikipedia.org/wiki/Viscoelasticity>.
- [34] “Wavepropagation,” 2014, <https://www.nde-ed.org/Education-Resources/CommunityCollege/Ultrasonics/Physics/wavepropagation.htm>.

Research Article

A Secure and Effective Anonymous Integrity Checking Protocol for Data Storage in Multicloud

Lingwei Song,¹ Dawei Zhao,² Xuebing Chen,³ Chenlei Cao,¹ and Xinxin Niu¹

¹Information Security Center, Beijing University of Posts and Telecommunications, Beijing 100876, China

²Shandong Provincial Key Laboratory of Computer Network, Shandong Computer Science Center (National Supercomputer Center in Jinan), Jinan 250014, China

³University of International Business and Economics, Beijing 100029, China

Correspondence should be addressed to Lingwei Song; songlw@bupt.edu.cn

Received 17 September 2014; Accepted 24 December 2014

Academic Editor: Florin Pop

Copyright © 2015 Lingwei Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to verify the integrity of outsourced data is an important problem in cloud storage. Most of previous work focuses on three aspects, which are providing data dynamics, public verifiability, and privacy against verifiers with the help of a third party auditor. In this paper, we propose an identity-based data storage and integrity verification protocol on untrusted cloud. And the proposed protocol can guarantee fair results without any third verifying auditor. The theoretical analysis and simulation results show that our protocols are secure and efficient.

1. Introduction

With the growing popularity of clouds, the tools and technologies for hybrid clouds have been emerging recently; cloud storage has become a hot research topic that aims to provide a comparably low cost, scalable, position-independent platform for data owners data [1]. However, this new paradigm of data hosting service also introduces new security challenges [2]. A list of security threats to cloud computing is presented in [3]. These issues range from the required trust in the cloud server for storage and attacks on cloud interfaces to misusing the cloud services for attacks in the complex systems. When considering using the complex cloud service, the data owner must be aware of the fact that all data given to the cloud server leave the owner control protection sphere [4]. Huge measurement data, huge environment monitoring data, hydrological data, marine biological data, and GIS information are provided by the complex multicloud. In this situation, it is a strong demand that the data owners can check the data integrity confidentially, dynamically, and publicly; besides, the anonymous is also demanded for smart phone users.

In the past few years, some work has been done on insuring remote data integrity checking, which allows data

integrity to be checked without completely downloading the data. Prior studies were based on two-party storage checking protocols that the data owner can check the data integrity [4–12]. Deswarte et al. [5] and Filho and Barreto [9] introduced RSA-based methods for solving remote data integrity checking. After that Shah et al. [12] proposed a remote storage auditing method based on precomputed challenge-response pairs. In practical application, to guarantee fair results, neither the cloud service provider nor the data owner should be the auditor in a cloud storage system. In this case, the protocols [13–15] employed the third party audit (TPA) performing the verification. However, none of them provided privacy against third party verifiers under the condition of introducing TPA. Wang et al. [14, 16] recognized the need of privacy against third party verifiers and proposed a random masking technique to cope with this problem. Scheme [17–21] required an additional trusted organizer to send a commitment to the auditor to ensure data privacy during auditing. The auditing protocol may make a performance bottleneck for the auditor. On some cases, without requiring any trusted organizer during the batch auditing for multiple clouds the client may delegate the remote data integrity checking task to the third party. It results in the untrusted

third party auditing in cloud computing [22, 23]. Yang and Jia [22] introduced an index table (ITable) to record the abstract information of the data; they proposed that the cloud server could be dishonest and may launch attacks just like replay attack, forge attack, and replace attack but only used ITable with time stamps to solve the problems. Wang [23] introduced identity-based distributed provable data possession in multicloud storage to check the certificate when it checks the remote data integrity. Chen et al. [24] also propose a new secure outsourcing algorithm for (variable-exponent, variable-base) exponentiation modulo a prime in the two untrusted program models.

However, one of benefits of cloud storage is to enable universal data access with independent geographical locations. This implies that the end devices may be mobile and limited in computation and storage. Efficient integrity checking protocols are more suitable for cloud clients equipped with mobile end devices. Meanwhile, when a mobile user remotes into a foreign network, mutual authentication must first be solved to prevent illegal use from accessing services and to ensure that mobile users are connected to trusted networks [25]. Both Zhao and Liu used smart-card to resolve the authentication. To compensate for these shortcomings, our construction can be observed as an adaptation of the protocol of [20, 22, 23, 25, 26].

This paper aims to fill the gap on a secure and effective anonymous authentication protocol for remote verification protocol in multicloud storage based on complex system. To the best of our knowledge, our scheme is the first to provide the authentication and establishment of remote verification scheme when mobile user is located in his/her home network; therefore it is more practical and universal for complex multicloud storage system. The scheme does not use timestamp; thus it avoids the clock synchronization problem. Additionally, the performance and cost analysis also show that our scheme is more suitable for low-power and resource-limited mobile devices and thus availability for real implementation.

The rest of the paper is organized as follows. The layered security architecture and definitions are present in Section 2. In Section 3, a novel anonymous authentication protocol for remote verification user authentication scheme is proposed in multicloud storage. In Section 4, we analyze the security of our proposed scheme. Next, we analyze the functionality and performance of our proposed scheme and make comparisons with other related schemes in Section 5. Finally, Section 6 gives the concluding remark of the whole paper.

2. Definitions and Preliminaries

In this section, we present our system model and briefly introduce the elliptic curve cryptosystem and some related mathematical assumptions.

2.1. Definitions of System Model. A representative network architecture for a secure and effective anonymous dynamics integrity checking protocol for data storage in multicloud (SA-DVCP) in global mobility networks is illustrated in

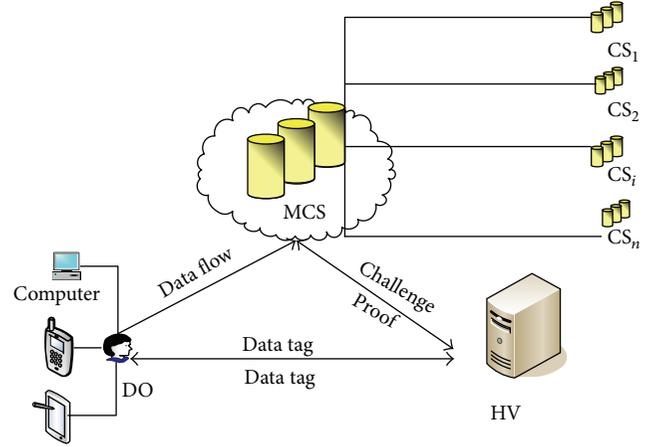


FIGURE 1: The system model of SA-DVCP.

Figure 1. Three different network entities can be identified as follows.

- (1) The data owner, that has massive data to be stored on the multicloud for maintenance and computation, can be either individual consumer or corporation who has large amount of data files to be stored in the cloud. DO has the ability to check the storage integrity of their outsourced data, while hoping to keep their data private from any entity which is untrusted. The checking devices may be mobile and limited in computation and storage, which need a secure and effective anonymous integrity checking protocol.
- (2) The data user/client/requested (DU), who accesses the CS or downloads the data from CS, has capabilities to check the integrity of data.
- (3) Data stakeholder (DS): we define both DO and DU as data stakeholder.
- (4) The multicloud server (MCS), which has significant storage space and computation resources to store the owners data and provides the data access to data users (data client/requesters), stores its whole data on the different cloud servers according to their importance and sensitivity.
- (5) The HV (Home Verifier) is a home third party that has expertise and capabilities to provide data storage auditing service for both the DS and DU. The HV can provide unbiased result for both the DO and the CS.

2.2. Notation and Preliminaries. Let f be a pseudorandom function and let π be a pseudorandom permutation. They can be described in detail as follows:

$$f: Z_p^* \times \{1, 2, \dots, n\} \rightarrow Z_p^*,$$

$$\pi: Z_p^* \times \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\},$$

in which k and d are two security parameters. Furthermore, denote the length of n in bits by $|P|$. We now introduce some necessary cryptographic background for our proposed scheme.

Bilinear Map. Let G_1 be a cyclic additive group generated by P and let G_2 be a cyclic multiplicative group generated by P with a bilinear map $\tilde{e} : G_1 \times G_2 \rightarrow G_T$.

- (a) $\forall P \in G_1, \forall Q \in G_2, \tilde{e}(aP, bQ) = \tilde{e}(P, Q)^{ab}$.
- (b) Nondegenerate: there exists $P, Q \in G_1$, such that $\tilde{e}(P, Q) \neq I_G$, where denotes the identity element of the group G_2 .
- (c) Computational discrete logarithm (CDL) problem: given $R = aP$, where $P, R \in G_1$. It is easy to calculate R given a and P , but it is hard to determine a given P and R .
- (d) Computational co-Diffie-Hellman: given $P, aP \in G_1$, and $aQ \in G_2$, compute $aQ \in G_2$.

For providing high insecurity level of the proposed scheme, some important mathematical assumptions are introduced for bilinear pairings defined on elliptic curves.

- (e) Define $A = aP$, $B = bP$, and $C = cP$; the computational bilinear Diffie-Hellman (CBDH) problem is computing the value $\text{bdh}(A, B, C)$ given randomly. The CBDH assumption asserts that the CBDH problem is hard that is for all PPT algorithms A .
- (f) Decision co-Diffie-Hellman: given $P, aP \in G_1$, and $Q, bQ \in G_2$, output is yes if $a = b$ and no otherwise. When the answer is yes we say that it is a co-Diffie-Hellman tuple.

3. The Proposed Schemes

In this section, we propose a novel anonymous dynamics integrity checking protocol for data storage in multicloud (SA-DVCP), using elliptic curve cryptosystem to not only protect the scheme from security breaches but also emphasize the efficient features. Before describing the auditing protocol definition, some notations are defined as in Notations and Descriptions section.

Suppose a file F has m data components as $M = (m_1, \dots, m_n)$. Each data component has its physical meanings and can be updated dynamically by the data owners. For public data components, the data owner does not need to encrypt it, but for private data component, the DO needs to encrypt it with its corresponding key.

For simplicity, we only consider one data component in our construction and constant number of sectors for each data block. Suppose there is a data component M , which is divided into n data blocks, and each data block is further split into s sectors. For data blocks with different sector number.

Then for $1 \leq i \leq n$, each block M_i is split into s sectors; that is, $M_i = \{M_{i1}, \dots, M_{is}\}$. Our storage auditing protocol consists of the following algorithms.

Setup (1^k) \rightarrow (pk, sk). Input the security parameter k and the bilinear map $e : G_P \times G_Q \rightarrow G_T$. Let G_T be multiplicative cyclic groups of prime order p , $G_P = \langle P \rangle$, $G_Q = \langle Q \rangle$, and $g = e(P, Q)$. Let $e(P, Q) \neq I_{G_T}$, $pk = (P, Q, g)$, and $sk = (PW_{DS}, ID_{DS})$; pk is the public key and sk is the private key. Let $h : \{0, 1\}^* \rightarrow G_P$ be a keyed secure hash function that maps the M_{ij} to a point in G_P .

TagGen (pk, sk, MCS, M) $\rightarrow D_i$. The tag generation algorithm takes as inputs each data component M and a set of CSP = $\{CS_j\}$, the private key sk . For each data block F_i , it computes a data tag D_i as $D_i = h(w_i, CS_i) \cdot \prod_{j=1}^s M_{ij} \cdot P$.

Where $w_i = \text{name} \parallel i$ and name is chosen by the DO uniformly at random from Z_p as the identifier of file M and i represents the block number of m_{ij} . It outputs the set of data tags $D_m = \{D_1, D_2, D_3, \dots, D_n\}$. Without loss of generality, we suppose that every block has its uniqueness. After finishing computing all the block tags, the DO sends the file M to MCS and releases D_m to be publicly known to everyone.

Proof (P, C (MCS), V ($Home Verifier$)). SA-DVCP is a protocol among P, C , and V . At the end of the interactive protocol, HV outputs the auditing result as 0 or 1. If DS delegates the verification task to HV, it needs to register himself/herself to his/her HV.

(1) **Registration.** The details of DS registration phase are shown in Figure 2.

The interaction protocol can be given in detail as follows.

Step R1. DS freely chooses his/her identity ID_{DS} and password PW_{DS} and generates a random number N_{DS} . Then DS submits $\{ID_{DS}, h(PW_{DS} \parallel N_{DS})\}$ to HV for registration via a secure channel.

Step R2. When receiving the message $\{ID_{DS}, h(PW_{DS} \parallel N_{DS})\}$ HV computes $Q = h(ID_{DS} \parallel y) \oplus h(PW_{DS} \parallel N_{DS})$ and $H = h(ID_{DS} \parallel h(PW_{DS} \parallel N_{DS}))$, $C = cP$, where y is a secret number of HV, and picks the challenge $\text{chal} = (c, k_1, k_2)$, $1 \leq c \leq n$, $k_1, k_2 \in Z_p^*$. Then HV submits $\{Q, H, ID_{HV}, C, \text{chal}\}$ to DS through a secure channel.

(2) **The Authentication and Proof.** The details of the authentication and proof DS registration phase are shown in Figure 3. When roaming into a foreign network MCS, DS needs to verify the validity of MCS and proves to DS that he is a legitimate user. The authentication and proof phase used to solve the above issue in our proposed scheme is described as follows.

Step P1. DS generates a random number a and computes $A = aP$, $R_{AC} = acP$, $N = Q \oplus h(PW_{DS} \parallel N_{DS})$, $DID_{DS} = ID \oplus h(R_{AC})$, and $V_1 = h(N \parallel ID_{DS} \parallel A \parallel C \parallel AC)$ and

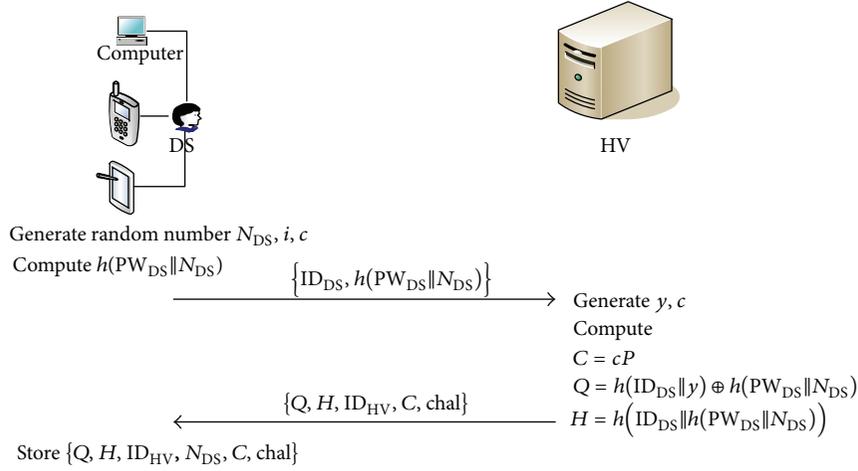


FIGURE 2: Registration phase of SA-DVCP.

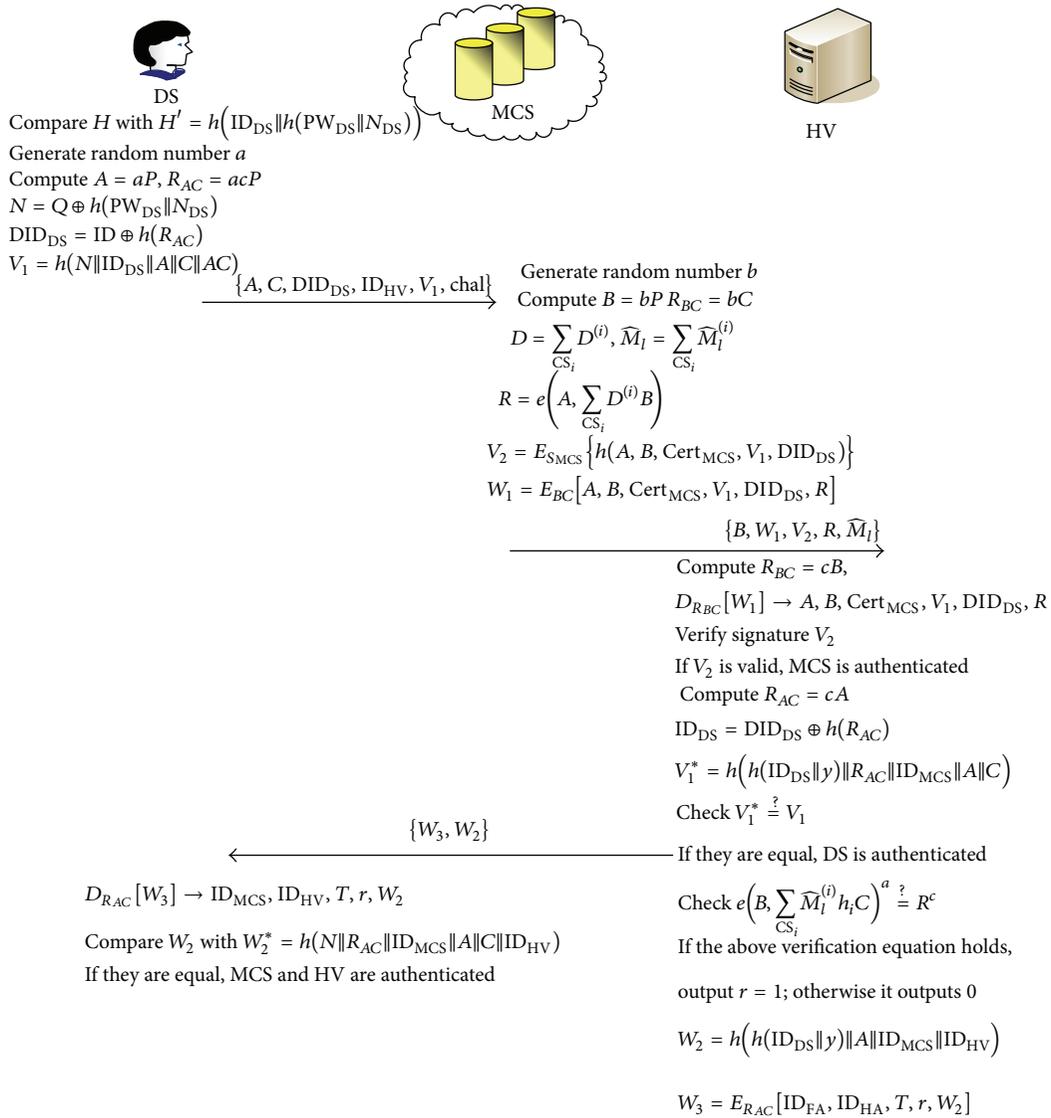


FIGURE 3: Proof phase of SA-DVCP.

DS sends the request message $\{A, C, DID_{DS}, ID_{HA}, V_1, chal\}$ to MCS over a public channel.

Step P2. After receiving the message $\{A, C, DID_{DS}, ID_{HA}, V_1, chal\}$, MCS generates a random number b and computes $B = bP$ and $R_{BC} = bC$ and $V_2 = E_{S_{MCS}}\{h(A, B, Cert_{MCS}, V_1, DID_{DS})\}$ and $W_1 = E_{BC}[A, B, Cert_{MCS}, V_1, DID_{DS}, R]$. Here, S_{MCS} is the private key of MCS, and $Cert_{MCS}$ is MCS's certificate. Next, the MCS calculates $v_i = \pi_{k_1}(i)$, $1 \leq i \leq c$ and looks up the table T_o to get the records that correspond to $\{v_1, v_2, \dots, v_c\} = F_1 \cup F_2 \cup \dots \cup F_n$ denoting the index set where the corresponding block-tag pair is stored in CS_i . Then, C sends (F_i, k_2) to $CS_i \in P$.

Response 1 ($P \leftarrow C$). For $CS_i \in P$, it performs the following procedures.

- For $v_l \in M_i$, CS_i splits M_{v_l} into s sectors $M_{v_l} = \{\widetilde{M}_{v_l1}, \widetilde{M}_{v_l2}, \dots, \widetilde{M}_{v_ls}\}$ and $1 \leq j \leq s$.
- CS_i calculates $a_l = f_{k_2}(l)$, $v_l \in M_i$, and $D^{(i)} = \sum_{v_l \in M_i} D_{v_l} a_l$.
- For $1 \leq j \leq s$, CS_i calculates $\widetilde{M}_j^{(i)} = \sum_{v_l \in F_i} a_l M_{v_lj}$ denoting $M_i = (\widetilde{M}_1^{(i)}, \dots, \widetilde{M}_s^{(i)})$.
- CS_i sends $\rho_i = (\widetilde{M}^{(i)}, D^{(i)})$ to C.

Response 2 ($C \rightarrow V$). After receiving all the responses from $CS_i \in P$, the combiner aggregates $\{\rho_i\}_{CS_i \in P}$ into the final response as $D = \sum_{CS_i} D^{(i)}$, $\widetilde{M}_l = \sum_{CS_i} \widetilde{M}_l^{(i)}$ denote $\widetilde{M} = (\widetilde{M}_1, \widetilde{M}_2, \dots, \widetilde{M}_s)$. The MCS generates the data proof R as $R = e(A, Db)$. Then MCS sends $\{B, W_2, V_2, R, \widetilde{M}_l\}$ to HV.

Step P3. When receiving $\{B, W_2, V_2, R\}$, HV first computes $R_{BC} = cB$ and decrypts $D_{R_{BC}}[W_1]$ to reveal $A, B, Cert_{MCS}, V_1$ and DID_{DS} . Then, HV verifies the MCS's signature V_2 by using the MCS's certificate $Cert_{MCS}$. If they are valid, MCS is authenticated. After that, HV computes the following:

$$R_{AC} = cA, \quad ID_{DS} = DID_{DS} \oplus h(R_{AC}),$$

$$V_1^* = h(h(ID_{DS} \parallel y) \parallel R_{AC} \parallel ID_{MCS} \parallel A \parallel C).$$

Then HV checks whether $V_1^* \stackrel{?}{=} V_1$. If they are equal, DS is authenticated by HV. Next, HV calculates

$$v_i = \pi_{k_1}(i),$$

$$h_i = h(N_{v_i}, CS_{i_{v_i}}, v_i),$$

$$a_i = f_{k_2}(i).$$

Then, it verifies whether the following formula holds: $e(B, \sum_{CS_i} \widetilde{M}_l^{(i)} h_i C)^a = R^c$.

If the formula holds, then the verifier outputs $r = 1$. Otherwise, the verifier outputs $r = 0$. Next compute

$$W_2 = h(h(ID_{DS} \parallel y) \parallel A \parallel ID_{MCS} \parallel ID_{HV})$$

$$W_3 = E_{R_{AC}}[ID_{MCS}, ID_{HV}, T, r, W_2]; \text{ at last, HV sends } \{W_3, W_2\} \text{ to DS.}$$

Step P4. DS decrypts $D_{R_{AC}}[W_3]$ to reveal $D_{MCS}, ID_{HA}, T, r, W_2$. Then, the MU compare W_2 with $W_2^* = h(N \parallel R_{AC} \parallel ID_{MCS} \parallel A \parallel C \parallel ID_{HV})$. If it is valid, HV and MCS are all authenticated by DS.

4. Security Analysis of the Proposed Scheme

In this section, we show that the proposed scheme can withstand all possible security attacks.

4.1. Storage Correctness Guarantee

Theorem 1. A SA-DVCP protocol must be workable and correct. That is, if the DS, MCS, and HV are honest and follow the specified procedures, the response $\{R, \widetilde{M}_l\}$ can pass HV's checking. The correctness follows from

$$e\left(B, \sum_{CS_i} \widetilde{M}_l^{(i)} h_i C\right)^a = e\left(bP, \sum_{CS_i} \widetilde{M}_l^{(i)} h_i cP\right)^a$$

$$= e\left(P, \sum_{CS_i, v_l \in M_i} h_i \sum_{j=1}^s a_l M_{v_lj} cP\right)^{ab} \quad (1)$$

$$= e\left(aP, \sum_{CS_i} D^{(i)} P\right)^{bc}$$

$$= e(A, Db)^c = R^c.$$

This completes the proof.

4.2. Privacy-Preserving Guarantee

Theorem 2. The proposed protocol can provide users privacy-preserving.

Proof. In our proposed scheme, the DS sends the login request message $\{A, C, DID_{DS}, ID_{HA}, V_1, chal\}$ to MCS, where $DID_{DS} = ID \oplus h(R_{AC})$ is used to protect the real identity ID_{DS} of DS. Based on the CDL problem, any attacker cannot obtain the random number a from A and thus cannot retrieve ID_{DS} from DID_{DS} . At the same time, the attacker cannot trace the moving history and current location of DS according to the login request message since A, DID_{DS} , and V_1 are dynamically changed in different login request messages of DS. Therefore, the proposed scheme can provide privacy-preserving of DS. \square

4.3. Resist Impersonation Attack. Our proposed protocol can efficiently prevent impersonation attacks by considering the following scenarios.

Proof. Our proposed scheme can efficiently prevent impersonation attacks by considering the following scenarios.

- (1) Any attacker cannot impersonate DS to cheat MCS and HV. In the proposed scheme, whether DS is located in a foreign network or in his/her home

TABLE 1: Comparison of cost.

	TagGen	Verify	$P + C$
Zhu et al. [17]	$(s + n(s + 2))C_{\text{exp}} + nsC_{\text{mul}}$	$3C_e + (c + s)C_{\text{exp}} + (c + s - 2)C_{\text{mul}}$	$nsC_e + (3n + c + 2)C_{\text{exp}} + (2\hat{n} + c - 1)C_{\text{mul}}$
Zhu et al. [20]	$(s + 2n)C_{\text{exp}} + nC_{\text{mul}}$	$3C_e + (c + s)C_{\text{exp}} + (c + s - 2)C_{\text{mul}}$	$cC_{\text{exp}} + (c - 1)C_{\text{mul}}$
Wang [23]	$n(s + 1)C_{\text{exp}} + nsC_{\text{mul}} + nsC_{h_1}$	$2C_e + (c + s + 1)C_{\text{exp}} + (c + s)C_{\text{mul}}$	$cC_{\text{exp}} + (c - 1)C_{\text{mul}} + csC_{h_1}$
Our protocol	$nsC_{\text{mul}} + nsC_{h_1}$	$C_e + (c + s)C_{\text{mul}}$	$C_e + (c + 1)C_{\text{mul}} + 2C_{\text{exp}} + csC_{h_1}$

network, the HV authenticates DS by verifying the computed $V_1^* = h(h(\text{ID}_{\text{DS}} \parallel y) \parallel R_{\text{AC}} \parallel \text{ID}_{\text{MCS}} \parallel A \parallel C)$ with the received $V_1 = h(N \parallel \text{ID}_{\text{DS}} \parallel A \parallel C \parallel \text{AC})$. Since the attacker does not possess DSs password PW_{DS} , he/she cannot compute the correct $N = Q \oplus h(\text{PW}_{\text{DS}} \parallel N_{\text{DS}})$ and thus cannot cheat HV by forging a login request message. At the same time, since a is a one-time random number and only possessed by DS, V_1 is dynamically changed in each login request message. Therefore, the attacker cannot cheat the HV by replaying a previous login request message. Besides, when DS is located in a foreign network, the authentication of MCS to DS is completely dependent on the authentication of HV to DS. If an attacker cannot successfully cheat HV by masquerading as DS, he/she cannot cheat MCS successfully.

- (2) Any attacker cannot impersonate MCS to cheat HV and DS. In the proposed scheme, the HV authenticates MCS by checking whether $D_{\text{P}_{\text{MCS}}}\{V_2\}$ equals $h(A, B, \text{Cert}_{\text{MCS}}, V_1, \text{DID}_{\text{DS}})$, where V_2 is MCSs digital signature. Obviously, the attacker cannot compute the correct MCSs digital signature without knowing MCSs private key S_{MCS} . Therefore, the attacker cannot cheat HV successfully by masquerading as MCS. At the same time, the authentication of DS to MCS is completely dependent on the authentication of HV to MCS. If an attacker cannot successfully cheat HV by masquerading as MCS, he/she cannot cheat DS successfully.
- (3) Any attacker cannot impersonate HV to cheat DS. In the proposed scheme, the DS authenticates HV by verifying $W_2^* = h(N \parallel R_{\text{AC}} \parallel \text{ID}_{\text{MCS}} \parallel A \parallel C \parallel \text{ID}_{\text{HV}})$ with the received $W_2 = h(h(\text{ID}_{\text{DS}} \parallel y)R \parallel A \parallel \text{ID}_{\text{MCS}} \parallel \text{ID}_{\text{HV}})$. Obviously, any attacker cannot compute the correct W_2 without knowing ID_{DS} and y , and the attacker cannot cheat DS successfully. \square

4.4. Forward Secrecy

Theorem 3. *The proposed protocol meets the security requirement for perfect forward secrecy.*

Proof. Perfect forward secrecy means that even if an attacker compromises all the passwords of the entities of the system, he/she still cannot compromise the session key. In the proposed scheme, these three one-time random numbers $a, b,$

and c are only held by the DS, MCS, and HV, respectively, and cannot be retrieved from $A = aP, B = bP, R_{\text{AC}} = aC = cA,$ and $R_{\text{BC}} = bC = cB$ based on the security of CDL and CDH problem. Thus, even if an adversary obtains all the passwords of the entities, previous session keys, and all the transmitted messages, he/she still cannot compromise other session keys. Hence, the proposed scheme achieves perfect forward secrecy. \square

5. Performance Comparison and Functionality Analysis

It is well known that most of the mobile devices have limited energy resources and computing capability. Hence, one of the most important issues in wireless networks is power consumption caused by communication and computation. In fact, the communication cost in the GLOMONET is higher than computation cost in terms of power consumption. In Table 1, we list the numbers of the TagGen, Verify and the $P + C$ phases of our scheme and some related previous schemes.

Computation. Suppose there are n message blocks which will be stored in n cloud servers. The blocks sector number is s and the challenged block number is c . We will consider the computation overhead in the different phases. On group G_p , bilinear pairings, exponentiation, multiplication, and the hash function h_1 contribute most computation cost. Compared with them, the hash function h and the operations on Z_p and G_Q are faster; the hash function H can be done once for all. On the DS, the computation cost mainly comes from the procedures of TagGen and verification (i.e., phase 5 in the protocol proof (P, C, V)). In the phase TagGen, the client performs ns multiplication on $G_p, n,$ and hash function h_1 . At the same time, for every file, the corresponding record ρ_i is stored by DS and CS. This stored metadata is small. In the phase proof, in order to respond the challenge $\text{chal} = (c, k_1, k_2)$ and generate the response ρ and the MCS perform $c + 1$ multiplication on the group G_p, cs hash function h_1 . In the verification of the response, HV performs 2 exponentiations, 2 pairings, and $c + s$ multiplication on the group G_p and c hash function h . On the other hand, in 2012, Zhu et al. proposed the cooperative provable data possession for integrity in multicloud storage [17]. Almost at the same time, Zhu et al. proposed the dynamic audit services for outsourced storage in clouds [20]. Compared with them, our proposed scheme is more efficient in the computation cost. The computation comparison can be summarized in Table 1.

In Table 1, C_{exp} denotes the time cost of exponentiation on the group $G_p; C_{\text{mul}}$ denotes the time cost of multiplication

TABLE 2: Comparison of communication cost.

Protocols	Chal	Response	Communication (rounds)
Zhu et al. [17]	$o(kc(\log_2 n + \log_2 q))$	$1G_P + 1G_Q + \text{slog}_2 q$	5
Zhu et al. [20]	$c(\log_2 n + \log_2 q)$	$1G_P + \text{slog}_2 q$	5
Wang [23]	$\log_2 n + 2 \log_2 q$	$G_P + \text{slog}_2 q$	5
Our protocol	$\log_2 n + 2 \log_2 q$	$\text{slog}_2 q$	5

on the group G_P ; C_e denotes the time cost of bilinear pairing; c_{h_1} denotes the time cost of the hash function h_1 . In other schemes, the sector must be in Z_P . Our scheme only requires the hash function h_1 's value which lies in Z_P . Thus, the hash function h_1 can be used to generate less block-tag pairs for the same file. Less block-tag pairs only incur less computation cost. This shows that our protocol can be implemented in mobile devices which have limited computation power.

Communication. In the phase proof, the communication overhead mainly comes from the challenge chal and response. The block-tag pairs are uploaded once and for all. After that, the phase proof will be performed periodically. Thus, the communication overheads mainly come from the Chal and responses. Suppose there are n message blocks stored in the CS. G_P and G_Q have the same order q . In chal, the verifier sends the challenge $\text{chal} = (c, k_1, k_2)$ to MCS. That is, the communication overhead is $\log_2 n + 2 \log_2 q$. On the other hand, Zhu et al. [17], Zhu et al. [20], and Wang [23] proposed three different provable data possession schemes. We do the comparison under the same probability of detection. Our scheme and Wang's ID-PDP have the same total communication cost during the challenge phase. During the proof phase, the communication cost of the proof incurs less communication cost than Wang's ID-PDP. Compared with these three schemes, our scheme is more efficient in the communication cost. The communication comparison can be summarized in Table 2. In Table 2, $1G_1$ denotes one element of G_P and $1G_1$ denotes one element of G_Q .

6. Conclusion

In this paper, we propose a novel anonymous authentication scheme for roaming service in global mobility networks. Security and performance analysis show that the proposed scheme is more suitable for the low-power and resource-limited mobile devices and is secure against various attacks and has many excellent features.

Notations and Descriptions

- G_P : Cyclic multiplicative group with generator q
- G_Q : Cyclic multiplicative group with generator Q
- Z_P^* : $\{1, 2, \dots, P-1\}$
- h, h_1 : Three cryptographic hash functions
- f : Pseudorandom function

- π : Pseudorandom permutation
- n : The block number
- s : The sector number
- $M = (M_1, \dots, M_n)$: The stored file M is split into n blocks
- $M_i = (\tilde{M}_{i1}, \dots, \tilde{M}_{in})$: The block M_i is split into s blocks
- \hat{n} : The cloud server number
- l_i : The index of the CS which stores the i th block-tag pair
- CS_{l_i} : The CS which stores the i th block
- $D_i = h(w_i, CS_i) \cdot \prod_{j=1}^s M_{ij} \cdot P$: The record where i denotes the i th block
- C_{exp} : The computation of exponentiation
- C_{h_1} : The computation of hash function
- C_{mul} : The computation of multiplications in group G
- C_e : The computation of bilinear pairings
- DO: Data owner
- DU: The data user/client/requested
- DS: Data stakeholder used to define both DO and DU
- MCS: The multicloud server
- HV: Home Verifier
- v_i : The permuted index of $v_i = \pi_{k_1}(i)$.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by Digital Right Management Technology Research and Development Project (1681300000119), Beijing Higher Education Young Elite Teacher Project (YETP0448), Specialized Research Fund for the Doctoral Program of Higher Education (2013114), Fundamental Research Funds for the Central Universities (2013RC0310), National Key Technology Research and Development Program (2012BAH08B02), National Natural Science Foundation of China (U1433105), and National 863 Program (2012AA012606).

References

- [1] Y. Zhu, H. Hu, G. J. Ahn et al., "Collaborative integrity verification in hybrid clouds," in *Proceedings of the 7th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom '11)*, pp. 191–200, IEEE, 2011.
- [2] T. Velte, A. Velte, and R. Elsenpeter, *Cloud Computing: A Practical Approach*, chapter 7, McGraw-Hill, New York, NY, USA, 1st edition, 2010.
- [3] D. Hubbard and M. Sutton, "Top Threats to Cloud Computing VI.0, Cloud Security Alliance," 2010, <https://cloudsecurityalliance.org/research/top-threats/>.
- [4] M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, and M. Isard, "A cooperative internet backup scheme," in *Proceedings of the Annual Conference on USENIX Annual Technical Conference*, pp. 29–41, 2003.
- [5] Y. Deswarte, J. Quisquater, and A. Saidane, "Remote integrity checking," in *Integrity and Internal Control in Information Systems VI: IFIP TC11/WG11.5 Sixth Working Conference on Integrity and Internal Control in Information Systems (IICIS) 13-14 November 2003, Lausanne, Switzerland*, vol. 140 of *IFIP International Federation for Information Processing*, pp. 1–11, 2004.
- [6] M. Naor and G. N. Rothblum, "The complexity of online memory checking," *Journal of the ACM*, vol. 56, no. 1, article 2, 2009.
- [7] A. Juels and B. S. Kaliski Jr., "Pors: proofs of retrievability for large files," in *Proceedings of the ACM Conference on Computer and Communications Security*, P. Ning, S. D. C. di Vimercati, and P. F. Syverson, Eds., pp. 584–597, 2007.
- [8] S. J. Thomas Schwarz and E. L. Miller, "Store, forget, and check: Using algebraic signatures to check remotely administered storage," in *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS '06)*, p. 12, July 2006.
- [9] D. L. G. Filho and P. S. L. M. Barreto, "Demonstrating data possession and uncheatable data transfer," *IACR Cryptology ePrint archive*, 2006.
- [10] F. Seb e, J. Domingo-Ferrer, A. Mart inez-Ballest e, Y. Deswarte, and J.-J. Quisquater, "Efficient remote data possession checking in critical information infrastructures," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1034–1038, 2008.
- [11] G. Yamamoto, S. Oda, and K. Aoki, "Fast integrity for large data," in *Proceedings of the ECRYPT Workshop Software Performance Enhancement for Encryption and Decryption*, pp. 21–32, June 2007.
- [12] M. A. Shah, M. Baker, J. C. Mogul, and R. Swaminathan, "Auditing to keep online storage services honest," in *Proceedings of the 11th USENIX Workshop Hot Topics in Operating Systems (HOTOS '07)*, G. C. Hunt, Ed., 2007.
- [13] G. Ateniese, R. Burns, R. Curtmola et al., "Provable data possession at untrusted stores," in *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS '07)*, pp. 598–610, ACM, New York, NY, USA, November 2007.
- [14] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 5, pp. 847–859, 2011.
- [15] H. Shacham and B. Waters, "Compact proofs of retrievability," in *Advances in Cryptology: Proceedings of the 14th International Conference on the Theory and Application of Cryptology and Information Security, Melbourne, Australia, December 2008*, vol. 5350 of *Lecture Notes in Computer Science*, pp. 90–107, Springer, Berlin, Germany, 2008.
- [16] C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for secure cloud storage," *IEEE Transactions on Computers*, vol. 62, no. 2, pp. 362–375, 2013.
- [17] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, "Cooperative provable data possession for integrity verification in multicloud storage," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 12, pp. 2231–2244, 2012.
- [18] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in *Proceedings of the 4th International Conference on Security and Privacy in Communication Networks (SecureComm '08)*, pp. 1–10, ACM, Istanbul, Turkey, September 2008.
- [19] C. Erway, A. K upc u, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in *Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS '09)*, pp. 213–222, New York, NY, USA, November 2009.
- [20] Y. Zhu, G.-J. Ahn, H. Hu, S. S. Yau, H. G. An, and C.-J. Hu, "Dynamic audit services for outsourced storages in clouds," *IEEE Transactions on Services Computing*, vol. 6, no. 2, pp. 227–238, 2013.
- [21] H. Wang, "Proxy provable data possession in public clouds," *IEEE Transactions on Services Computing*, vol. 6, no. 4, pp. 551–559, 2013.
- [22] K. Yang and X. Jia, "An efficient and secure dynamic auditing protocol for data storage in cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 9, pp. 1717–1726, 2013.
- [23] H. Wang, "Identity-based distributed provable data possession in multi-cloud storage," *IEEE Transactions on Services Computing*, 2014.
- [24] X. Chen, X. Huang, J. Li, J. Ma, W. Lou, and D. S. Wong, "New algorithms for secure outsourcing of large-scale systems of linear equations," *IEEE Transactions on Information and Forensics Security*, vol. 10, no. 1, pp. 69–78, 2015.
- [25] D. Zhao, H. Peng, L. Li, and Y. Yang, "A secure and effective anonymous authentication scheme for roaming service in global mobility networks," *Wireless Personal Communications*, vol. 78, no. 1, pp. 247–269, 2014.
- [26] J. Liu, Z. Zhang, X. Chen, and K. S. Kwak, "Certificateless remote anonymous authentication schemes for wireless body area networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 2, pp. 332–342, 2014.

Research Article

Stochastic Fractional Heat Equations Driven by Fractional Noises

Xichao Sun¹ and Ming Li²

¹Department of Mathematics and Physics, Bengbu College, 1866 Caoshan Road, Bengbu, Anhui 233030, China

²School of Information Science and Technology, East China Normal University, Shanghai 200241, China

Correspondence should be addressed to Xichao Sun; sunxichao626@126.com

Received 31 July 2014; Accepted 30 August 2014

Academic Editor: Ezzat G. Bakhoun

Copyright © 2015 X. Sun and M. Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper is concerned with the following stochastically fractional heat equation on $(t, x) \in [0, T] \times \mathbb{R}^d$ driven by fractional noise: $\partial u(t, x)/\partial t = \mathcal{D}_0^\alpha u(t, x) + W^H(t, x) \diamond u(t, x)$, where the Hurst parameter $H = (h_0, h_1, \dots, h_d)$ and \diamond denotes the Skorokhod integral. A unique solution of that equation in an appropriate Hilbert space is constructed. Moreover, the Lyapunov exponent of the solution is estimated, and the Hölder continuity of the solution on both space and time parameters is discussed. On the other hand, the absolute continuity of the solution is also obtained.

1. Introduction

Recently, there has been intense interest (see, e.g., Podlubny [1], Samko et al. [2], Heydari et al. [3, 4], Cattani et al. [5, 6], Liao [7], and Hu [8]) in fractional calculus and its applications. Many mathematical problems in physics and engineering with respect to systems and processes are represented by a kind of equations, more precisely fractional order differential equations driven by fractional order noise. One of the emerging branches of this study is the theory and applications of fractional (partial) differential equations. The increasing interest in this class of equations is motivated both by their applications to fluid dynamic traffic model, viscoelasticity, heat conduction in materials with memory, electrodynamics with memory and also because they can be employed to approach nonlinear conservation laws (see [9] and references therein). For instance, the research discussed by Gurtin and Pipkin [10] and Nunziato [11] provides a description of heat conduction in materials with fading memory. Besides, noise or stochastic perturbation is unavoidable and omnipresent in nature as well as in man-made systems. Therefore, it is of great significance to import the stochastic effects into the investigation of fractional (partial) differential systems.

As known, publications treating stochastic partial differential equation (SPDE) involving fractional derivatives gain

interests. In fact, these kinds of equations may be widely used in physics, fractal medium, quantum fields, risk management, statistical mechanics, and other areas (see Droniou and Imbert [12], Uchaikin and Zolotarev [13], Toma [14], Bakhoun and Toma [15], and Li et al. [16, 17]). Most of them investigate evolution type equations driven by a fractional power of the Laplacian. Mueller [18] and Wu [19] proved the existence of a solution of stochastically fractional heat and Burgers equation perturbed by a stable noise, respectively. Boulamba et al. [20] studied the existence, uniqueness, Hölder regularity, and absolute continuity of the solution for a class of fractional stochastic partial differential equations driven by spatially correlated noise. Other related references include Chang and Lee [21], Debbi and Dozzi [22], Sun et al. [23], Truman and Wu [24], Liu et al. [25], and Wu [26].

On the other hand, there has been an increasing interest in studying SPDEs driven by fractional noise. Recall that a fractional Brownian motion (fBm) is a centered Gaussian process $(B_t)_{t \geq 0}$ with the covariance given by

$$E(B_t B_s) = R_H(t, s) := \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H}), \quad (1)$$

with $H \in (0, 1)$, referring Biagini et al. [27], Mishura [28], and Nualart [29] for a comprehensive account on the fBm.

The fBm has some interesting properties, such as self-similarity, Hölder continuity, and long-range dependence. It has been applied in various scientific areas including telecommunications, turbulence, image processing, and finance engineering. Moreover, there have been many papers about SPDEs driven by fractional noise. Hu [30] showed the existence and uniqueness of the solutions for a class of second-order stochastic heat equations, via chaos expansion. Nualart and Ouknine [31] explored the existence and uniqueness of mild solution to a class of second-order heat equations with additive fractional noise (fractional in time and white in space) when the Hurst parameter $H > 1/2$. In a successive paper, Tindel et al. [32] studied a linear stochastic evolution equation driven by an infinite-dimensional fBm in the cases of the Hurst parameter above and below $1/2$, respectively. Hu and Nualart [33] studied the d -dimensional stochastic heat equation with a multiplicative Gaussian noise which is white in space and has the covariance of an fBm with $H \in (0, 1)$ in time. More works for the fields can be found in Balan [34], Balan and Tudor [35, 36], Bo et al. [37, 38], Jiang et al. [39–41], and the references therein.

Let $B^H(t, x), (t, x) \in [0, T] \times \mathbb{R}^d$, be the canonical fractional Brownian fields (FBF) with the Hurst parameter H on the canonical probability space (Ω, \mathcal{F}, P) , where $H = (h_0, h_1, \dots, h_d)$ with $1/2 < h_i < 1$, $0 \leq i \leq d$. The fractional white noise is denoted by $W^H(t, x)$; that is, formally $W^H(t, x) = (\partial^{d+1}/\partial t \partial x_1 \partial x_2 \dots \partial x_d) B^H(t, x)$. In this paper, we focus on the following stochastically fractional partial differential equation $(t, x) \in [0, T] \times \mathbb{R}^d$ driven by fractional noise:

$$\frac{\partial u(t, x)}{\partial t} = \mathcal{D}_\delta^\alpha u(t, x) + W^H(t, x) \diamond u(t, x), \quad (2)$$

where “ \diamond ” denotes the Skorokhod integral, $(t, x) \in [0, T] \times \mathbb{R}^d$, $d \geq 1$, $\alpha = (\alpha_1, \dots, \alpha_d)$, $\delta = (\delta_1, \dots, \delta_d)$, and $\mathcal{D}_\delta^\alpha$ denotes a nonlocal fractional differential operator defined by

$$\mathcal{D}_\delta^\alpha = \sum_{i=1}^d D_{\delta_i}^{\alpha_i}, \quad (3)$$

where $D_{\delta_i}^{\alpha_i}$ denotes the fractional differential derivative with respect to the i th coordinate defined via its Fourier transform \mathcal{F} by

$$\mathcal{F}(D_\delta^\alpha \varphi)(\xi) = -|\xi|^\alpha \exp\left(-i\delta \frac{\pi}{2} \operatorname{sgn}(\xi)\right) \mathcal{F}(\varphi)(\xi), \quad (4)$$

where $i^2 + 1 = 0$. The precise meaning of the solution of (2) will be stated in Section 2.

The structure of this paper is as follows. In Section 2, we briefly present some basic notations and preliminaries. Section 3 consists of the existence and the Lyapunov exponent estimate of the solution to (2). In Section 4, we check the Hölder continuity of the solution $\{u(t, x), (t, x) \in [0, T] \times \mathbb{R}^d\}$ with both space and time parameters. In Section 5, we prove that the law of the solution of (2) is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^d . Finally, Section 6 concludes the paper.

2. Preliminaries

In this section, we will first introduce multiple stochastic integrals with respect to fractional Brownian fields and define a solution of (2) in S_ρ sense after proposing the fractional differential operator $\mathcal{D}_\delta^\alpha$ and introducing some properties. Then, we will recall the Malliavin calculus with respect to fractional noises.

2.1. Skorokhod Integral with respect to Fractional Brownian Fields

Definition 1. A multiparameter fractional Brownian field $B^H = \{B^H(t, x), (t, x) \in [0, T] \times \mathbb{R}^d\}$ with multiparameter $H = (h_0, h_1, \dots, h_d)$ for $h_i \in (0, 1)$ and $i \in \{0, 1, 2, \dots, d\}$ is a centered Gaussian field defined on some probability space (Ω, \mathcal{F}, P) with the covariance as follows:

$$\begin{aligned} E[B^H(t, x)B^H(s, y)] &= \frac{1}{2^{d+1}} (t^{2h_0} + s^{2h_0} - |t-s|^{2h_0}) \\ &\quad \times \prod_{i=1}^d (x_i^{2h_i} + y_i^{2h_i} - |x_i - y_i|^{2h_i}) \end{aligned} \quad (5)$$

for all $t, s \in [0, T]$, $x = (x_1, \dots, x_d)$, and $y = (y_1, \dots, y_d) \in \mathbb{R}^d$.

Throughout the paper, we restrict our consideration on the multiparameter fractional Brownian field with the parameter $H = (h_0, h_1, \dots, h_d)$ for $h_i \in (1/2, 1)$.

Firstly, we briefly introduce the stochastic integral with respect to the fractional Brownian field $B^H(t, x)$.

For any $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$, we write $\varphi_H(x, y) := \prod_{i=1}^d \varphi_{h_i}(x_i, y_i)$, where $\varphi_h := h(2h-1)|x-y|^{2h-2}$. Introduce the following Hilbert space:

$$\mathcal{H}_1 := \left\{ f : [0, T] \times \mathbb{R}^d \longrightarrow \mathbb{R}; \right.$$

$$\begin{aligned} \|f\|_H^2 &= \int_{[0, T]^2} \int_{\mathbb{R}^{2d}} \varphi_{h_0}(s-r) \varphi_H(x, y) f(s, x) \\ &\quad \times f(r, y) dx dy ds dr < \infty \left. \right\}, \end{aligned} \quad (6)$$

where $dx = dx_1 dx_2 \dots dx_d$ and $dy = dy_1 dy_2 \dots dy_d$. Let $f \in \mathcal{H}_1$; one can define the following stochastic integral:

$$\left\{ \int_0^t \int_{\mathbb{R}^d} f(s, x) dB^H(t, x); t \in [0, T] \right\} \quad (7)$$

(see, e.g., Hu [30]).

Proposition 2. Let $f, g \in \mathcal{H}_1$. Then,

- (1) $E[\int_0^T \int_{\mathbb{R}^d} f(s, x) dB^H(s, x)] = 0$.
- (2) $E[\int_0^T \int_{\mathbb{R}^d} f(s, x) dB^H(s, x) \int_0^T \int_{\mathbb{R}^d} g(s, x) dB^H(s, x)] = \langle f, g \rangle_{\mathcal{H}_1}$.

Denote by \otimes the symmetric tensor product. Let $\{e_1, e_2, \dots\}$ be an orthonormal basis of \mathcal{H}_1 . Then, $\{e_{i_1} \otimes e_{i_2} \cdots \otimes e_{i_n}, 1 \leq i_1, i_2, \dots, i_n < \infty\}$ is an orthonormal basis of $\mathcal{H}_n := \mathcal{H}_1^{\otimes n}$. It is easy to see that

$$\begin{aligned} \mathcal{H}_n := \{f : [0, T]^n \times \mathbb{R}^{nd} \longrightarrow \mathbb{R}; \\ \int_{[0, T]^{2n}} \int_{\mathbb{R}^{2nd}} \prod_{i=1}^n \varphi_{h_0}(s_i - r_i) \varphi_H(u_i, v_i) f(s, u) \\ \times f(r, v) ds dr du_1 dv_1 \cdots du_n dv_n \\ < \infty\}, \end{aligned} \quad (8)$$

where we denote

$$u = (u_1, u_2, \dots, u_n), \quad v = (v_1, v_2, \dots, v_n). \quad (9)$$

Define

$$\begin{aligned} \mathcal{H}_n := \{f \in \mathcal{H}_n; f((s_1, u_1), \dots, (s_n, u_n)) \\ = f((s_{\sigma(1)}, u_{\sigma(1)}), \dots, (s_{\sigma(n)}, u_{\sigma(n)})) \\ \text{for all permutations} \\ \sigma = \{\sigma(1), \dots, \sigma(n)\} \text{ of } \{1, 2, \dots, n\}\}. \end{aligned} \quad (10)$$

We call $f \in \mathcal{H}_n$ symmetric to n ($(d+1)$ -dimensional) variables $(s_1, u_1), \dots, (s_n, u_n)$. Now, define

$$\mathcal{D}_n := \{f \in \mathcal{H}_n; f = \sum a_{i_1, \dots, i_n} e_{i_1} \otimes e_{i_2} \cdots \otimes e_{i_n}, a_{i_1, \dots, i_n} \in \mathbb{R}\}. \quad (11)$$

Denote by \otimes the symmetric tensor product. Let $H_n(x)$ be the Hermite polynomial of degree $n \geq 0$. It is defined by

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}, \quad x \in \mathbb{R}. \quad (12)$$

For $e \in \mathcal{H}_1$ and $\|e\|_{\mathcal{H}_1} = 1$, define the multiple integral of Itô type of the function $e^{\otimes n}$ by

$$\begin{aligned} \int_{[0, T]^n} \int_{\mathbb{R}^{nd}} e^{\otimes n}(s_1, \dots, s_n, u_1, \dots, u_n) \\ \cdot dB^H(s_1, u_1) \cdots dB^H(s_n, u_n) \\ = H_n \left(\int_{[0, T]} \int_{\mathbb{R}^d} e(s, u) dB^H(s, u) \right). \end{aligned} \quad (13)$$

Then,

$$\begin{aligned} I_n(e_{i_1} \otimes e_{i_2} \cdots \otimes e_{i_n}) = \int_{[0, T]^n} \int_{\mathbb{R}^{nd}} e^{\otimes n}(s_1, \dots, s_n, u_1, \dots, u_n) \\ \cdot dB^H(s_1, u_1) \cdots dB^H(s_n, u_n), \end{aligned} \quad (14)$$

by the polarization argument.

For each $f \in \mathcal{D}_n$, we write

$$I_n(f) = \sum a_{i_1, \dots, i_n} I_n(e_{i_1} \otimes e_{i_2} \cdots \otimes e_{i_n}). \quad (15)$$

Then, the following isometry holds:

$$E|I_n(f)|^2 = n! |f|_{\mathcal{H}_n}^2. \quad (16)$$

Note that, for $f \in \mathcal{H}_n$, there exists a sequence $\{f_k\}_{k \in \mathbb{N}} \subset \mathcal{D}_n$ such that $f_n \rightarrow f$ in \mathcal{D}_n . It follows from (15) that $\{I_n(f_k)\}_{k \in \mathbb{N}}$ is Cauchy in $L^2(\Omega)$, and the limit point of $I_n(f_k)$ (as $k \rightarrow \infty$) is independent of the choice of $f \in \mathcal{H}_n$. We call the limit point the multiple integral of Itô type and denote it by

$$\begin{aligned} I_n(f) := \int_{[0, T]^n} \int_{\mathbb{R}^{nd}} f(s, u) dB^H(s_1, u_1) \cdots dB^H(s_n, u_n) \\ = \lim_{k \rightarrow \infty} I_n(f_k), \quad \text{in } L^2(\Omega) \text{ sense.} \end{aligned} \quad (17)$$

It is easy to see that, for $f, g \in \mathcal{H}_n$,

$$E[I_n(f) I_n(g)] = n! \langle f, g \rangle_{\mathcal{H}_n}. \quad (18)$$

Let $F = \bigoplus_{n=0}^{\infty} F_n$, where F_n is the n th chaos of F (see, e.g., Hu [42]). For each $\rho \in \mathbb{R}$, we introduce a Hilbert space denoted by

$$S_\rho = \left\{ F = \bigoplus_{n=0}^{\infty} F_n; \sum_{n=0}^{\infty} [n!]^\rho E|F_n|^2 < \infty \right\} \quad (19)$$

and define

$$|F|_\rho := \sqrt{\sum_{n=0}^{\infty} [n!]^\rho E|F_n|^2}. \quad (20)$$

In particular, if $\rho = 0$, one has $S_\rho = L^2(\Omega)$.

2.2. Definition of the Solution. In order to define the solution of (2), we will introduce the nonlocal fractional differential operator defined by

$$\mathcal{D}_\delta^\alpha = \sum_{i=1}^d D_{\delta_i}^{\alpha_i}, \quad (21)$$

where $\alpha = (\alpha_1, \dots, \alpha_d)$, $\delta = (\delta_1, \dots, \delta_d)$, and $D_{\delta_i}^{\alpha_i}$ denotes the fractional differential derivative with respect to the i th coordinate defined via its Fourier transform \mathcal{F} by

$$\mathcal{F}(D_\delta^\alpha \varphi)(\xi) = -|\xi|^\alpha \exp\left(-i\delta \frac{\pi}{2} \operatorname{sgn}(\xi)\right) \mathcal{F}(\varphi)(\xi). \quad (22)$$

In this paper, we will assume that $|\delta_i| \leq \min\{\alpha_i - [\alpha_i]_2, 2 + [\alpha_i]_2 - \alpha_i\}$, $i = 1, \dots, d$, $[\alpha_i]_2$, is the largest even integer less or equal to α_i (even part of α_i), and $\delta_i \in 2\mathbb{N} + 1$.

In one space dimension, the operator D_δ^α is a closed, densely defined operator on $L^2(\mathbb{R})$. It is the infinitesimal generator of a semigroup which is in general not symmetric

and not a contraction. This operator is a generalization of various well-known operators, such as the Laplacian operator when $\alpha = 2$, the inverse of the generalized Riesz-Feller potential if $\alpha > 2$, and the Riemann-Liouville differential operator when $|\delta| = 2 + [\alpha]_2$ or $|\delta| = \alpha - [\alpha]$. It is self-adjoint only when $\delta = 0$ and, in this case, it coincides with the fractional power of the Laplacian, citing Debbi [43], Debbi and Dozzi [22], and Komatsu [44] for more details about this operator.

According to Komatsu [44], D_δ^α can be represented for $1 < \alpha < 2$ by

$$D_\delta^\alpha \varphi(x) = \int_{\mathbb{R}} \frac{\varphi(x+y) - \varphi(x) - y\varphi'(x)}{|y|^{1+\alpha}} \times (\kappa_-^\delta 1_{(-\infty,0)}(y) + \kappa_+^\delta 1_{(0,+\infty)}(y)) dy \quad (23)$$

and for $0 < \alpha < 1$ by

$$D_\delta^\alpha \varphi(x) = \int_{\mathbb{R}} \frac{\varphi(x+y) - \varphi(x)}{|y|^{1+\alpha}} \times (\kappa_-^\delta 1_{(-\infty,0)}(y) + \kappa_+^\delta 1_{(0,+\infty)}(y)) dy, \quad (24)$$

where κ_-^δ and κ_+^δ are two nonnegative constants satisfying $\kappa_-^\delta + \kappa_+^\delta > 0$ and φ is a smooth function for which the integral exists and φ' is its derivative. This representation identifies it as the infinitesimal generator for a nonsymmetric α -stable Lévy process.

Let $G_{\alpha,\delta}(t, x)$ be the fundamental solution of the following Cauchy problem:

$$\begin{aligned} \frac{\partial u}{\partial t}(t, x) &= D_\delta^\alpha(t, x), \\ u(0, x) &= \delta_0(x), \quad t > 0, x \in \mathbb{R}, \end{aligned} \quad (25)$$

where $\delta_0(\cdot)$ is the Dirac distribution. By the Fourier transform, we see that $D_{\alpha,\delta}(t, x)$ is given by \mathbb{F} as follows:

$$\begin{aligned} G_{\alpha,\delta}(t, x) &= \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-izx - t|z|^\alpha \exp(-i\delta \frac{\pi}{2} \operatorname{sgn}(z))) dz. \end{aligned} \quad (26)$$

The relevant parameters α called the index of stability and δ (related to the asymmetry) improperly referred to as the skewness are real numbers satisfying $|\delta| \leq \min\{\alpha - [\alpha]_2, 2 + [\alpha]_2 - \alpha\}$, and $\delta = 0$ when $\delta \in 2\mathbb{N} + 1$.

Let us list some known results on $G_{\alpha,\delta}(t, x)$ that will be used later on (see, e.g., Debbi [43] and Debbi and Dozzi [22]).

Lemma 3. *Let $\alpha \in (1, \infty) \setminus \{\mathbb{N}\}$; we have the following.*

- (1) *The function $G_{\alpha,\delta}(t, \cdot)$ is not in general symmetric relatively to x and it is not everywhere positive.*
- (2) *For any $s, t \in (0, \infty)$ and $x \in \mathbb{R}$,*

$$\frac{\partial^n}{\partial x^n} G_{\alpha,\delta}(t, x) = (s)^{-(n+1)/\alpha} G_{\alpha,\delta}(s^{-1}t, s^{-(1/\alpha)}x) \quad (27)$$

or equivalently

$$\frac{\partial^n}{\partial x^n} G_{\alpha,\delta}(t, x) = (t)^{-((n+1)/\alpha)} G_{\alpha,\delta}(1, (t)^{-(1/\alpha)}x). \quad (28)$$

$$(3) G_\alpha(s, \cdot) * G_{\alpha,\delta}(t, \cdot) = G_{\alpha,\delta}(s+t, \cdot) \text{ for any } s, t \in (0, \infty).$$

$$(4) \text{ For } n \geq 1, \text{ there exist some constants } C \text{ and } C_n > 0 \text{ such that, for all } x \in \mathbb{R},$$

$$|G_{\alpha,\delta}(1, x)| \leq C \frac{1}{1 + |x|^{1+\alpha}}, \quad (29)$$

$$\left| \frac{\partial^n}{\partial x^n} G_{\alpha,\delta}(1, x) \right| \leq C_n \frac{|x|^{\alpha+n-1}}{(1 + |x|^{\alpha+n})^2}.$$

$$(5) \int_0^T \int_{\mathbb{R}} |G_{\alpha,\delta}(t, x)|^\lambda dt dx < \infty \text{ if and only if } 1/\alpha < \lambda < \alpha.$$

For $d \geq 1$ and any multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$ and $\delta = (\delta_1, \dots, \delta_d)$, denote by $\mathcal{G}_{\alpha,\delta}(t, x)$ the Green function of the deterministic equation as follows:

$$\frac{\partial u}{\partial t}(t, x) = \mathcal{D}_\delta^\alpha(t, x), \quad (30)$$

$$u(0, x) = \delta_0(x), \quad t > 0, x \in \mathbb{R}^d.$$

Clearly,

$$\begin{aligned} \mathcal{G}_{\alpha,\delta}(t, x) &= \prod_{i=1}^d G_{\alpha_i, \delta_i}(t, x_i) \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp\left(-i \langle \xi, x \rangle - t \sum_{i=1}^d |\xi_i|^{\alpha_i} \times \exp\left(-i\delta \frac{\pi}{2} \operatorname{sgn}(\xi_i)\right)\right) d\xi. \end{aligned} \quad (31)$$

Now, we describe a solution of (2) in S_ρ sense.

Definition 4. We say that a stochastic field $\{u(t, x), (t, x) \in [0, T] \times \mathbb{R}^d\}$ is a solution of (2) in S_ρ sense, if

- (1) $u : \mathbb{R}_+ \times \mathbb{R} \times \Omega$ is jointly measurable;
- (2) $\int_0^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t-s, x-z) u(s, z) dB^H(s, z)$ is well defined for all $t \in \mathbb{R}_+$ and $x \in \mathbb{R}^d$ as an element of \mathcal{S}_ρ for certain $\rho \in \mathbb{R}$;
- (3) the following equation holds in \mathcal{S}_ρ :

$$\begin{aligned} u(t, x) &= \mathcal{G}_{\alpha,\delta}(t) * u(0, x) \\ &+ \int_0^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t-s, x-z) u(s, z) dB^H(s, z). \end{aligned} \quad (32)$$

The following embedding proposition given by Mémin et al. [45] is useful for our derivations below.

Lemma 5. *If $h \in (1/2, 1)$ and $f, g \in L^{1/h}([a, b])$, one has*

$$\int_a^b \int_a^b f(u)g(v)|u-v|^{2h-2} du dv \leq C_h \|f\|_{L^{1/h}([a,b])} \|g\|_{L^{1/h}([a,b])}, \quad (33)$$

where $C_h > 0$ is a constant depending only on h .

2.3. Malliavin Calculus. In this subsection, let us brief the basic elements of the Malliavin calculus (see, e.g., Nualart [29]).

Since $B^H(t, x), (t, x) \in [0, T] \times \mathbb{R}^d$, is Gaussian, we might develop the Malliavin calculus with respect to fractional noises in order to prove the existence of the laws of the solutions of SPDE driven by fractional noises.

Let $B^H(\varphi) = \int_0^T \int_{\mathbb{R}^d} \varphi(t, x) B^H(dt, dx)$ for $\varphi \in \mathcal{H}_1$ and

$$\begin{aligned} \mathcal{S} = \{ & F = f(B^H(\varphi_1), \dots, B^H(\varphi_n)); \\ & f \in C_b^\infty(\mathbb{R}^n), \varphi_i \in \mathcal{H}_1, n \geq 1 \} \end{aligned} \quad (34)$$

be the space of all ‘‘smooth cylindrical’’ random variables, where $C_b^\infty(\mathbb{R}^n)$ denotes the class of all bounded infinitely differentiable functions on \mathbb{R}^n , whose partial derivatives are also bounded. The Malliavin derivative of an element $F = f(B^H(\varphi_1), \dots, B^H(\varphi_n))$, with respect to B^H , is defined by

$$D_{t,x}F := \sum_{i=1}^n \frac{\partial f}{\partial x_i} (B^H(\varphi_1), \dots, B^H(\varphi_n)) \varphi_i(t, x). \quad (35)$$

Let $\mathcal{D}^{1,2}$ be the completion of \mathcal{S} under the norm

$$\|F\|_{1,2}^2 := E|F|^2 + E\|DF\|_{\mathcal{H}_1}^2. \quad (36)$$

Then, $\mathcal{D}^{1,2}$ is the domain of the closed operator D . For each $h \in \mathcal{H}_1$ and $F \in \mathcal{S}$, define

$$\begin{aligned} D_h F := \lim_{\varepsilon \rightarrow 0} \frac{d}{dh} f(B^H(h_1) + \varepsilon \langle h_1, h \rangle_{\mathcal{H}_1}, \dots, B^H(h_n) \\ + \varepsilon \langle h_n, h \rangle_{\mathcal{H}_1}), \end{aligned} \quad (37)$$

which might be extended as a closed operator on $L^2(\Omega)$ with the domain \mathcal{D}_h being the closure of \mathcal{S} under the norm

$$\|F\|_h^2 = E[|F|^2 + |D_h F|^2]. \quad (38)$$

Let $\{h_n, n \geq 1\}$ be an orthonormal basis of \mathcal{H}_1 . Then, $F \in \mathcal{D}^{1,2}$ if and only if $F \in \mathbb{D}_{h_n}$ for each $n \in \mathbb{N}$ and

$$\sum_{n=1}^{\infty} E|D_{h_n} F|^2 < \infty. \quad (39)$$

In this case,

$$D_h F = \langle DF, h \rangle_{\mathcal{H}_1}. \quad (40)$$

On the other hand, the divergence operator $\delta : \text{Dom } \delta \subset L^2(\Omega; \mathcal{H}_1) \rightarrow L^2(\Omega)$ is the adjoint of the operator D and is

uniquely defined by the following duality relationship: $u \in \text{Dom } \delta$ if and only if

$$E(F\delta(u)) = E\langle DF, u \rangle_{\mathcal{H}_1}, \quad \forall F \in \mathcal{S}. \quad (41)$$

Note that $u \in \text{Dom } \delta$ if and only if u is integrable with respect to B^H . In the literature, δ is called the Skorokhod integral with respect to B^H .

The following propositions (see Wei [46] for the case of fractional noises), which are the Malliavin calculus with respect to fractional-colored noises, can deduce the laws for solutions to the corresponding stochastic partial differential equations.

Proposition 6. *Let $A \in \mathcal{F}$. If F is a square integrable random variable that is measurable with respect to the σ -field \mathcal{F}_{A^c} , then*

$$DF1_A = 0, \quad \text{a.s.} \quad (42)$$

Remark 7. Let $\{u(t, x), (t, x) \in [0, T] \times \mathbb{R}^d\}$ be an $\{\mathcal{F}_t, t \in [0, T]\}$ -adapted random field. By Proposition 6, we have $D_{s,y}u(t, x) = 0$, a.s., for any $0 \leq t < s \leq T$ and $x, y \in \mathbb{R}^d$.

Proposition 8. *Suppose $F \in \mathcal{D}^{1,2}$; if $\|DF\|_{\mathcal{H}_1}^2 > 0$, a.s., then the distribution of the random variable F is absolutely continuous with respect to Lebesgue measure.*

Remark 9. Propositions 6 and 8 can be proved similarly to those in Nualart [29].

3. Lyapunov Exponent Estimate of the Solution

In this section, we will establish the existence and uniqueness of the solution of (2) and give the Lyapunov exponent estimate of the solution. We denote $T_n = \{0 < s_1 < s_2 < \dots < s_n < t\}$ and $ds = ds_1 ds_2 \dots ds_n$. In fact, iterating (32) yields formally that

$$u(t, x) = \mathcal{G}_{\alpha,\delta}(t) * u(0, x) + \sum_{n=1}^{\infty} I_n(f_n(t, x)), \quad (43)$$

where

$$\begin{aligned} & f_n(t, x; s_1, x_1, s_2, x_2, \dots, s_n, x_n) \\ & \text{Sym} \left\{ \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t - s_n, x - x_n) \dots \mathcal{G}_{\alpha,\delta}(s_2 - s_1, x_2 - x_1) \right. \\ & \left. \times \mathcal{G}_{\alpha,\delta}(s_1, x_1 - y) u(0, y) dy \right\}, \end{aligned} \quad (44)$$

where Sym is the symmetrization with respect to $n((d+1)$ -dimensional) variables $(s_1, x_1), \dots, (s_n, x_n)$ and

$$\begin{aligned} I_n(f_n(t, x)) = & \int_{([0,T] \times \mathbb{R}^d)^n} f_n(t, x; s_1, x_1, s_2, x_2, \dots, s_n, x_n) \\ & \cdot dB^H(s_1, x_1) dB^H(s_2, x_2) \dots dB^H(s_n, x_n). \end{aligned} \quad (45)$$

We compute the \mathcal{L}_2 norm of each chaos; that is,

$$\Theta_n(t, x) = E(I_n(f_n(t, x)))^2. \tag{46}$$

Assume that $|u(0, x)| \leq C < \infty$, where, in what follows, inessential constants will be denoted generically by C , even if they vary from line to line. By the isometric equality (18), we obtain

$$\begin{aligned} \Theta_n(t, x) &= n! \int_{T_n^2} \int_{\mathbb{R}^{2n(d+1)}} \prod_{i=1}^n \varphi_{h_0}(s_i - r_i) \varphi_H(\xi_i - \eta_i) \\ &\quad \times \mathcal{G}_{\alpha, \delta}(t - s_n, x - \xi_n) \\ &\quad \cdots \mathcal{G}_{\alpha, \delta}(s_2 - s_1, \xi_2 - \xi_1) \\ &\quad \times \mathcal{G}_{\alpha, \delta}(s_1, \xi_1 - y) u_0(y) \\ &\quad \times \mathcal{G}_{\alpha, \delta}(t - r_n, x - \eta_n) \\ &\quad \cdots \mathcal{G}_{\alpha, \delta}(r_2 - r_1, \eta_2 - \eta_1) \\ &\quad \times \mathcal{G}_{\alpha, \delta}(r_1, \eta_1 - y) \\ &\quad \times u_0(\bar{y}) dy d\bar{y} d\xi d\eta ds dr \tag{47} \\ &\leq Cn! \int_{T_n^2} \int_{\mathbb{R}^{2nd}} \prod_{i=1}^n \varphi_{h_0}(s_i - r_i) \varphi_H(\xi_i - \eta_i) \\ &\quad \times \mathcal{G}_{\alpha, \delta}(t - s_n, x - \xi_n) \\ &\quad \cdots \mathcal{G}_{\alpha, \delta}(s_2 - s_1, \xi_2 - \xi_1) \\ &\quad \times \mathcal{G}_{\alpha, \delta}(t - r_n, x - \eta_n) \cdots \mathcal{G}_{\alpha, \delta} \\ &\quad \times (r_2 - r_1, \eta_2 - \eta_1) d\xi d\eta ds dr. \end{aligned}$$

Let

$$\begin{aligned} G_{\alpha_i, \delta_i}(t, x) &= \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-izx - t|z|^{\alpha_i} \\ &\quad \times \exp(-i\delta_i \frac{\pi}{2} \operatorname{sgn}(z))) dz, \tag{48} \\ &\quad 1 \leq i \leq d. \end{aligned}$$

Thus,

$$\Theta_n(t, x) = Cn! \int_{T_n^2} \prod_{i=1}^n \varphi_{h_0}(s_i - r_i) \prod_{i=1}^d \Theta_{i,n}(t, x_i, s, r) ds dr, \tag{49}$$

where

$$\begin{aligned} \Theta_{i,n}(t, x_i, s, r) &= \int_{\mathbb{R}^{2n}} \prod_{k=1}^n \varphi_{h_i}(\rho_k - \tau_k) G_{\alpha_i, \delta_i}(t - s_n, x_i - \rho_n) \\ &\quad \cdots G_{\alpha_i, \delta_i}(s_2 - s_1, \rho_2 - \rho_1) \cdot G_{\alpha_i, \delta_i}(t - r_n, x_i - \tau_n) \\ &\quad \cdots G_{\alpha_i, \delta_i}(r_2 - r_1, \tau_2 - \tau_1) \\ &\quad \times d\rho_1 d\rho_2 \cdots d\rho_n d\tau_1 d\tau_2 \cdots d\tau_n. \tag{50} \end{aligned}$$

Now, we estimate $\Theta_{i,n}(t, x_i, s, r)$.

Lemma 10. *There is a constant C_{h_i} such that*

$$\begin{aligned} \Theta_{i,n}(t, x_i, s, r) &\leq C_{h_i}^n \prod_{k=1}^n (s_{k+1} - s_k)^{(h_i-1)/\alpha_i} (r_{k+1} - r_k)^{(h_i-1)/\alpha_i}, \tag{51} \end{aligned}$$

where $s_{n+1} = r_{n+1} = t$.

Proof. Applying Lemma 5, we have

$$\begin{aligned} \widetilde{\Theta}_{i,n}(t, x_i, s_1, r_1) &= \int_{\mathbb{R}^2} \varphi_{h_i}(\rho_1 - \tau_1) G_{\alpha_i, \delta_i}(s_2 - s_1, \rho_2 - \rho_1) \\ &\quad \times G_{\alpha_i, \delta_i}(r_2 - r_1, \tau_2 - \tau_1) d\rho_1 d\tau_1 \\ &= h_i(2h_i - 1) \int_{\mathbb{R}^2} |\rho_1 - \tau_1|^{2h_i-2} \\ &\quad \times G_{\alpha_i, \delta_i}(s_2 - s_1, \rho_2 - \rho_1) \\ &\quad \times G_{\alpha_i, \delta_i}(r_2 - r_1, \tau_2 - \tau_1) d\rho_1 d\tau_1 \\ &\leq C |G_{\alpha_i, \delta_i}(s_2 - s_1, \rho_2 - \cdot)|_{L^{1/h_i}(\mathbb{R})} \\ &\quad \times |G_{\alpha_i, \delta_i}(r_2 - r_1, \tau_2 - \cdot)|_{L^{1/h_i}(\mathbb{R})}. \tag{52} \end{aligned}$$

Note that

$$\begin{aligned} &|G_{\alpha_i, \delta_i}(s_2 - s_1, \rho_2 - \cdot)|_{L^{1/h_i}(\mathbb{R})} \\ &= \left(\int_{\mathbb{R}} |G_{\alpha_i, \delta_i}(s_2 - s_1, \rho_2 - \rho_1)|^{1/h_i} d\rho_1 \right)^{h_i} \\ &= \left(\int_{\mathbb{R}} (s_2 - s_1)^{-(1/h_i)\alpha_i} |G_{\alpha_i, \delta_i}(1, (s_2 - s_1)^{-(1/\alpha_i)}(\rho_2 - \rho_1))|^{1/h_i} \right. \\ &\quad \left. \times d\rho_1 \right)^{h_i} \\ &= \left(\int_{\mathbb{R}} (s_2 - s_1)^{(1/\alpha_i) - (1/h_i)\alpha_i} |G_{\alpha_i, \delta_i}(1, \rho_1)|^{1/h_i} d\rho_1 \right)^{h_i} \\ &\leq (s_2 - s_1)^{(1/\alpha_i)(h_i-1)} \int_{\mathbb{R}} \left(\frac{c}{1 + |\rho_1|^2} \right)^{1/h_i} d\rho_1 \\ &\leq C_{h_i} (s_2 - s_1)^{(1/\alpha_i)(h_i-1)}. \tag{53} \end{aligned}$$

Therefore,

$$\widetilde{\Theta}_{i,n}(t, x_i, s_1, r_1) \leq C_{h_i, \alpha_i, \delta_i} [(s_2 - s_1)(r_2 - r_1)]^{(1/\alpha_i)(h_i-1)}. \tag{54}$$

Thus, we show that (51) is true when $n = 1$. The lemma follows from iteration. \square

Lemma 11. Let $2h_0 - \sum_{i=1}^d ((1 - h_i)/\alpha_i) > 1$ and $p \in (1/(2h_0 - 1), 1/\sum_{i=1}^d ((1 - h_i)/\alpha_i))$. Then, $u(t, x) \in \mathcal{S}_\rho$ if $\rho < 2/p - 2\sum_{i=1}^d ((1 - h_i)/\alpha_i - 1)$. Moreover,

$$\limsup_{t \rightarrow \infty} \frac{\log \sup_{x \in \mathbb{R}^d} \|u(t, x)\|_\rho^2}{t^k} < \infty, \quad (55)$$

for $k = 2(h_0 - \sum_{i=1}^d ((1 - h_i)/\alpha_i))/(2/p - 2\sum_{i=1}^d ((1 - h_i)/\alpha_i - 1 - \rho))$.

Proof. Let $\gamma = 2(\sum_{i=1}^d ((1 - h_i)/\alpha_i))$, $p, q > 1$, and $1/p + 1/q = 1$. Then, by the Hölder inequality, one gets

$$\begin{aligned} & \Theta_n(t, x) \\ & \leq n! C_{H, \alpha, \delta}^n \int_{T_n^2} \prod_{k=1}^n |s_k - r_k|^{2h_0 - 2} \\ & \quad \times [(s_{k+1} - s_k)(r_{k+1} - r_k)]^{-(\gamma/2)} ds dr \\ & \leq n! C_{H, \alpha, \delta}^n \left\{ \int_{T_n^2} \prod_{k=1}^n [(s_{k+1} - s_k)(r_{k+1} - r_k)]^{-(p\gamma/2)} ds dr \right\}^{1/p} \\ & \quad \cdot \left\{ \int_{T_n^2} \prod_{k=1}^n |s_k - r_k|^{(2h_0 - 2)q} ds dr \right\}^{1/q} \\ & = n! C_{H, \alpha, \delta}^n \left\{ \int_{T_n} \prod_{k=1}^n (s_{k+1} - s_k)^{-(p\gamma/2)} ds \right\}^{2/p} \\ & \quad \cdot \left\{ \int_{T_n^2} \prod_{k=1}^n |s_k - r_k|^{(2h_0 - 2)q} ds dr \right\}^{1/q}. \end{aligned} \quad (56)$$

Following Hu [30], one can get a constant C_{p, h_0} such that, for any value q such that $1 < q < 1/(2 - 2h_0)$ (i.e., $p > 1/(2h_0 - 1)$),

$$\left\{ \int_{T_n} \prod_{k=1}^n |s_k - r_k|^{(2h_0 - 2)q} ds dr \right\}^{1/q} \leq C_{p, h_0}^n t^{((2 + (2h_0 - 2)q)n)/q}. \quad (57)$$

Set $\theta = p\gamma/2 \in (0, 1)$ (i.e., $p < 1/\sum_{i=1}^d ((1 - h_i)/\alpha_i)$); similar to the proof of (3.6) in Bo et al. [37], we get

$$\int_{T_n} \prod_{k=1}^n (s_{k+1} - s_k)^{-\theta} ds \leq C_H^n \frac{t^{(1-\theta)n}}{\Gamma(n(1-\theta) + 1)}, \quad (58)$$

where $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1}$ is the Gamma function. Thus, we obtain

$$\begin{aligned} & \Theta_n(t, x) \\ & \leq C_{H, \alpha, \delta}^n n! \frac{t^{(2n/p)(1-p\sum_{i=1}^d ((1-h_i)/\alpha_i))}}{[\Gamma(n(1-p\sum_{i=1}^d ((1-h_i)/\alpha_i)) + 1)]^{2/p}} \\ & \quad \cdot t^{(2+(2h_0-2)q)n/q} \\ & = C_{H, \alpha, \delta}^n n! \frac{t^{2n(h_0 - \sum_{i=1}^d ((1-h_i)/\alpha_i))}}{[\Gamma(n(1-p\sum_{i=1}^d ((1-h_i)/\alpha_i)) + 1)]^{2/p}}. \end{aligned} \quad (59)$$

We continue to use the notation introduced previously. The Stirling formula yields that, for $x > -1$,

$$\Gamma(x + 1) = K(x) x^x e^{-x}, \quad (60)$$

where the function $K(x)$ satisfies $\lambda^{-x} \leq K(x) \leq \lambda^x$. Hence, for $\beta = 1 - p\sum_{i=1}^d ((1 - h_i)/\alpha_i)$, by the definition of \mathcal{S}_ρ ,

$$\begin{aligned} & \|u(t, x)\|_\rho^2 \\ & \leq \sum_{n=0}^\infty (n!)^\rho \Theta_n(t, x) \\ & \leq \sum_{n=0}^\infty (n!)^\rho n! C_{H, \alpha, \delta}^n \frac{t^{2n(h_0 - \sum_{i=1}^d ((1-h_i)/\alpha_i))}}{[\Gamma(n\beta + 1)]^{2/p}} \\ & \leq \sum_{n=0}^\infty C_{H, \alpha, \delta}^n (\lambda^{4\beta/p})^n \left[\frac{(2\beta/p - (\rho + 1))^{2\beta/p - (\rho + 1)}}{\beta^{2\beta/p}} \right]^n \\ & \quad \times \frac{t^{2n(h_0 - \sum_{i=1}^d ((1-h_i)/\alpha_i))}}{\Gamma(k(2\beta/p - (\rho + 1)) + 1)} \\ & \leq E_{2\beta/p - (\rho + 1)}(\tilde{C} t^{2(h_0 - \sum_{i=1}^d ((1-h_i)/\alpha_i))}), \end{aligned} \quad (61)$$

where $\tilde{C} = C_{H, \alpha, \delta} \lambda^{4\beta/p} (2\beta/p - (\rho + 1))^{2\beta/p - (\rho + 1)} / \beta^{2\beta/p} > 0$ and $E_r(z)$ is the Mittag-Leffler function with the parameter $r > 0$. Note that $2\beta/p - \rho - 1 > 0$; then, by the asymptotic property of the Mittag-Leffler function (see, e.g., Podlubny [1]), one has

$$\begin{aligned} & \|u(t, x)\|_\rho^2 \leq \frac{C}{2\beta/p - (\rho + 1)} \\ & \quad \times \exp(\tilde{C} t^{2(h_0 - \sum_{i=1}^d ((1-h_i)/\alpha_i))/(2\beta/p - (\rho + 1))}) \\ & < \infty, \end{aligned} \quad (62)$$

if $\rho < 2/p - 2\sum_{i=1}^d ((1 - h_i)/\alpha_i - 1)$. This completes the proof of the lemma. \square

Lemma 12. $u(t, x)$ defined by (43) is the solution of (2) in the sense of Definition 4.

Proof. Let $u(t, x)$ be given by (43). It suffices to verify (32). By the Fubini lemma and the definition of the integral and the definition of $f_n(t, x)$, we have

$$\begin{aligned} & \int_0^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha, \delta}(t - s, x - z) I_n(f_n(s, z)) dB^H(s, z) \\ & = I_{n+1}(f_{n+1}(t, x)). \end{aligned} \quad (63)$$

Thus, the Lemma follows. \square

From the above lemmas, we get the following.

Theorem 13 (existence, uniqueness, and Lyapunov exponent estimate). Let $2h_0 - \sum_{i=1}^d ((1 - h_i)/\alpha_i) > 1$ and $p \in (1/(2h_0 - 1), 1/\sum_{i=1}^d ((1 - h_i)/\alpha_i))$. If $u_0 \in L^\infty(\mathbb{R}^d)$, then (2) has a unique

solution $u(t, x) \in \mathcal{S}_\rho$ when $\rho < 2/p - 2 \sum_{i=1}^d ((1 - h_i)/\alpha_i - 1)$. Moreover,

$$\limsup_{t \rightarrow \infty} \frac{\log \sup_{x \in \mathbb{R}^d} \|u(t, x)\|_\rho^2}{t^k} < \infty, \quad (64)$$

for $k = 2n(h_0 - \sum_{i=1}^d ((1 - h_i)/\alpha_i))/(2/p - 2 \sum_{i=1}^d ((1 - h_i)/\alpha_i - 1 - \rho))$.

4. Hölder Regularity

Theorem 14. Assume that $2h_0 - \sum_{i=1}^d ((1 - h_i)/\alpha_i) > 1$. If there exists some $\gamma \in (0, 1)$ such that, for all $z \in K$ (compact subset of \mathbb{R}^d),

$$\sup_{x \in \mathbb{R}^d} \mathbb{E} \left(|u_0(x+z) - u_0(x)|^p \right) < c |z|^{\gamma p_0} \quad (65)$$

for some p_0 large enough in $[2, \alpha_0/\gamma]$, where $\alpha_0 = \min_{1 \leq i \leq d} \{\alpha_i\}$, then the solution $\{u(t, x), (t, x) \in [0, T] \times \mathbb{R}^d\}$ is θ Hölder continuous in t and μ Hölder continuous in x , where $\theta \in (0, \min_{1 \leq i \leq d} \{1, \sum_{j=1}^d (\gamma/\alpha_j), h_0 \alpha_i (1 + \sum_{j=1}^d ((h_j - 1)/h_0 \alpha_j))\})$ and $\mu \in (0, \min\{\gamma, (h_0 - \sum_{i=1}^d ((1 - h_i)/\alpha_i))(1 + \sum_{i=1}^d (1/\alpha_i))^{-1}\})$.

Proof. It follows from the proof of Lemma 11. If $p \in (1/(2h_0 - 1), 2(1 + 2 \sum_{i=1}^d ((1 - h_i)/\alpha_i)^{-1}))$, then $u(t, x) \in \mathcal{S}_\rho$ when $\rho < 2/p - 2 \sum_{i=1}^d ((1 - h_i)/\alpha_i - 1)$. In particular, $u(t, x) \in \mathcal{S}_0 = L^2(\Omega)$ since $2/p - 2 \sum_{i=1}^d ((1 - h_i)/\alpha_i - 1) > 0$. On the other hand, from (43), it follows that

$$\begin{aligned} u(t, x) &= \mathcal{G}_{\alpha, \delta}(t) * u(0, x) + \sum_{n=1}^{\infty} I_n(f_n(t, x)) \\ &= \mathcal{G}_{\alpha, \delta}(t) * u(0, x) \\ &\quad + \int_0^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha, \delta}(t-r, x-z) u(s, z) dB^H(r, z). \end{aligned} \quad (66)$$

Thus, for $s, t \in [0, T]$ and $x, y \in \mathbb{R}^d$, we have

$$\begin{aligned} &E |u(t, x) - u(s, y)|^2 \\ &\leq E |\mathcal{G}_{\alpha, \delta}(t) * u(0, x) - \mathcal{G}_{\alpha, \delta}(s) * u(0, y)|^2 \\ &\quad + E \left| \int_s^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha, \delta}(t-r, x-z) u(r, z) dB^H(r, z) \right|^2 \\ &\quad + E \left| \int_0^s \int_{\mathbb{R}^d} (\mathcal{G}_{\alpha, \delta}(t-r, x-z) - \mathcal{G}_{\alpha, \delta}(s-r, y-z)) \right. \\ &\quad \quad \left. \times u(r, z) dB^H(r, z) \right|^2 \end{aligned} \quad (67)$$

:= I + II + III.

Now, we will estimate A, B, and C, respectively. Following Boulanba et al. [20], one gets

$$I \leq C_T (|x - y|^{2\gamma_1} + |t - s|^{2\gamma_2}), \quad (68)$$

where $0 < \gamma_1 < \rho$ and $0 < \gamma_2 < \min\{\sum_{i=1}^d (\gamma/\alpha_i), 1\}$.

By Lemma 5 and (62) with $\rho = 0$, one can get

$$\begin{aligned} \text{II} &= \int_{[s, t]^2} \int_{\mathbb{R}^{2d}} \varphi_{h_0}(r - \bar{r}) \varphi_H(z - \bar{z}) |\mathcal{G}_{\alpha, \delta}(t - r, x - z)| \\ &\quad \times |\mathcal{G}_{\alpha, \delta}(t - \bar{r}, x - \bar{z})| \\ &\quad \times E(|u(r, z) u(\bar{r}, \bar{z})|) dz d\bar{z} dr d\bar{r} \\ &\leq \int_{[s, t]^2} \int_{\mathbb{R}^{2d}} \varphi_{h_0}(r - \bar{r}) \varphi_H(z - \bar{z}) |\mathcal{G}_{\alpha, \delta}(t - r, x - z)| \\ &\quad \times |\mathcal{G}_{\alpha, \delta}(t - \bar{r}, x - \bar{z})| \|u(r, z)\|_0 \\ &\quad \times \|u(\bar{r}, \bar{z})\|_0 dz d\bar{z} dr d\bar{r} \\ &\leq C_T \int_{[s, t]^2} \int_{\mathbb{R}^{2d}} \varphi_{h_0}(r - \bar{r}) \varphi_H(z - \bar{z}) |\mathcal{G}_{\alpha, \delta}(t - r, x - z)| \\ &\quad \times |\mathcal{G}_{\alpha, \delta}(t - \bar{r}, x - \bar{z})| dz d\bar{z} dr d\bar{r} \\ &= C_T \int_{[s, t]^2} \varphi_{h_0}(r - \bar{r}) dr d\bar{r} \\ &\quad \times \int_{\mathbb{R}^{2d}} \varphi_H(z - \bar{z}) |\mathcal{G}_{\alpha, \delta}(t - r, x - z)| \\ &\quad \times |\mathcal{G}_{\alpha, \delta}(t - \bar{r}, x - \bar{z})| dz d\bar{z} \\ &\leq \left(\int_s^t \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \cdots \left(\int_{\mathbb{R}} |\mathcal{G}_{\alpha, \delta}(t - r, x - z)|^{1/h_d} dz_d \right)^{h_d/h_{d-1}} \right. \right. \right. \\ &\quad \left. \left. \left. \cdots dz_2 \right)^{h_2/h_1} dz_1 \right)^{h_1/h_0} dr \right)^{2h_0} \\ &=: C_T \|\mathcal{G}_{\alpha, \delta}(t - \cdot, x - \cdot)\|_{L^{1/h_d - (1/h_1)(1/h_0)}}^2. \end{aligned} \quad (69)$$

By the definition of the Green function $\mathcal{G}_{\alpha, \delta}$ and Lemma 3, one gets

$$\begin{aligned} &\int_{\mathbb{R}} |\mathcal{G}_{\alpha, \delta}(t - r, x - z)|^{1/h_d} dz_d \\ &= \prod_{i=1}^{d-1} |G_{\alpha_i, \delta_i}(t - r, x_i - z_i)|^{1/h_d} \\ &\quad \times \int_{\mathbb{R}} (t - r)^{-(1/\alpha_d h_d)} \\ &\quad \times |G_{\alpha_i, \delta_i}(1, (t - r)^{-(1/\alpha_d h_d)}(x_i - z_i))|^{1/h_d} dz_d \\ &= \prod_{i=1}^{d-1} |G_{\alpha_i, \delta_i}(t - r, x_i - z_i)|^{1/h_d} \\ &\quad \times \int_{\mathbb{R}} (t - r)^{-(1/\alpha_d h_d) + 1/\alpha_d} |G_{\alpha_i, \delta_i}(1, z_d)|^{1/h_d} dz_d \\ &\leq \prod_{i=1}^{d-1} |G_{\alpha_i, \delta_i}(t - r, x_i - z_i)|^{1/h_d} \\ &\quad \times (t - r)^{-(1/\alpha_d h_d) + 1/\alpha_d} \\ &\quad \times \int_{\mathbb{R}} \left(\frac{C}{1 + |z_d|^2} \right)^{1/h_d} dz_d \\ &\leq C \prod_{i=1}^{d-1} |G_{\alpha_i, \delta_i}(t - r, x_i - z_i)|^{1/h_d} (t - r)^{-(1/\alpha_d h_d) + 1/\alpha_d}. \end{aligned} \quad (70)$$

Using induction, one can get

$$\begin{aligned} & \|\mathcal{G}_{\alpha,\delta}(t-\cdot, x-\cdot)\|_{L^{1/h_d \dots (1/h_1)(1/h_0)}}^2 \\ &= \left(\int_s^t \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \cdots \left(\int_{\mathbb{R}} |\mathcal{G}_{\alpha,\delta}(t-r, x-z)|^{1/h_d} dz_d \right)^{h_d/h_{d-1}} \right. \right. \right. \\ & \quad \left. \left. \left. \cdots dz_2 \right)^{h_2/h_1} dz_1 \right)^{h_1/h_0} dr \right)^{2h_0} \\ &\leq C_{T,H,\alpha,\delta} \left(\int_s^t (t-r)^{\sum_{i=1}^d ((h_i-1)/h_0\alpha_i)} dr \right)^{2h_0} \\ &\leq C_{T,H,\alpha,\delta} (t-s)^{2h_0 \sum_{i=1}^d ((h_i-1)/h_0\alpha_i+1)}. \end{aligned} \quad (71)$$

Therefore,

$$\text{II} \leq C_{T,H,\alpha,\delta} (t-s)^{2h_0(\sum_{i=1}^d ((h_i-1)/h_0\alpha_i+1))}. \quad (72)$$

As for III, we have

$$\begin{aligned} \text{III} &\leq 2 \left[E \left| \int_0^s \int_{\mathbb{R}^d} (\mathcal{G}_{\alpha,\delta}(t-r, x-z) \right. \right. \\ & \quad \left. \left. - \mathcal{G}_{\alpha,\delta}(t-r, y-z)) \right. \right. \\ & \quad \left. \left. \times u(r, z) dB^H(r, z) \right|^2 \right. \\ & \quad \left. + E \left| \int_0^s \int_{\mathbb{R}^d} (\mathcal{G}_{\alpha,\delta}(t-r, y-z) \right. \right. \\ & \quad \left. \left. - \mathcal{G}_{\alpha,\delta}(s-r, y-z)) \right. \right. \\ & \quad \left. \left. \times u(r, z) dB^H(r, z) \right|^2 \right] \\ &=: \text{III}_1 + \text{III}_2. \end{aligned} \quad (73)$$

Let $\theta \in (0, 1)$. Then,

$$\begin{aligned} \text{III}_1 &= \int_{[0,s]^2} \int_{\mathbb{R}^{2d}} \varphi_{h_0}(r-\bar{r}) \varphi_H(z-\bar{z}) \\ & \quad \times |\mathcal{G}_{\alpha,\delta}(t-r, x-z) - \mathcal{G}_{\alpha,\delta}(t-r, y-z)| \\ & \quad \times |\mathcal{G}_{\alpha,\delta}(t-\bar{r}, x-\bar{z}) - \mathcal{G}_{\alpha,\delta}(t-\bar{r}, y-\bar{z})| \\ & \quad \times E(|u(r, z) u(\bar{r}, \bar{z})|) dz d\bar{z} dr d\bar{r} \\ &\leq C_T \int_{[0,T]^2} \int_{\mathbb{R}^{2d}} \varphi_{h_0}(r-\bar{r}) \varphi_H(z-\bar{z}) \\ & \quad \times |\mathcal{G}_{\alpha,\delta}(t-r, x-z) \\ & \quad \quad - \mathcal{G}_{\alpha,\delta}(t-r, y-z)| \\ & \quad \times |\mathcal{G}_{\alpha,\delta}(t-\bar{r}, x-\bar{z}) \\ & \quad \quad - \mathcal{G}_{\alpha,\delta}(t-\bar{r}, y-\bar{z})| dz d\bar{z} dr d\bar{r} \end{aligned}$$

$$\begin{aligned} &= C_T \int_{[0,T]^2} \varphi_{h_0}(r-\bar{r}) dr d\bar{r} \\ & \quad \times \int_{\mathbb{R}^{2d}} \varphi_H(z-\bar{z}) \\ & \quad \times |\mathcal{G}_{\alpha,\delta}(t-r, x-z) - \mathcal{G}_{\alpha,\delta}(t-r, y-z)| \\ & \quad \times |\mathcal{G}_{\alpha,\delta}(t-\bar{r}, x-\bar{z}) \\ & \quad \quad - \mathcal{G}_{\alpha,\delta}(t-\bar{r}, y-\bar{z})| dz d\bar{z} \\ &= C_T \|\mathcal{G}_{\alpha,\delta}(t-\cdot, x-\cdot) - \mathcal{G}_{\alpha,\delta}(t-\cdot, y-\cdot)\|_{L^{1/h_d \dots (1/h_1)(1/h_0)}}^2 \\ &\leq C_T \left(\|\mathcal{G}_{\alpha,\delta}(t-\cdot, x-\cdot) - \mathcal{G}_{\alpha,\delta}(t-\cdot, y-\cdot)\|^\theta \right. \\ & \quad \times \|\mathcal{G}_{\alpha,\delta}(t-\cdot, x-\cdot)\|_{L^{1/h_d \dots (1/h_1)(1/h_0)}}^{1-\theta} \|^2 \\ & \quad \left. + \|\mathcal{G}_{\alpha,\delta}(t-\cdot, x-\cdot) - \mathcal{G}_{\alpha,\delta}(t-\cdot, y-\cdot)\|^\theta \right. \\ & \quad \left. \times \|\mathcal{G}_{\alpha,\delta}(t-\cdot, y-\cdot)\|_{L^{1/h_d \dots (1/h_1)(1/h_0)}}^{1-\theta} \|^2 \right) \\ &=: C_T (\text{III}_{1,1} + \text{III}_{1,2}). \end{aligned} \quad (74)$$

Using the mean-value theorem, it holds that

$$\begin{aligned} \text{III}_{1,1} &\leq \| |x-y|^\theta |\nabla \mathcal{G}_{\alpha,\delta}(t-\cdot, \xi-\cdot)|^\theta \\ & \quad \times \|\mathcal{G}_{\alpha,\delta}(t-\cdot, x-\cdot)\|_{L^{1/h_d \dots (1/h_1)(1/h_0)}}^{1-\theta} \|^2 \\ &= |x-y|^{2\theta} \left\| \left(\sum_{i=1}^d \left| \frac{\partial}{\partial x_i} \mathcal{G}_{\alpha,\delta}(t-\cdot, \xi-\cdot) \right| \right) \right\|^\theta \\ & \quad \times \|\mathcal{G}_{\alpha,\delta}(t-\cdot, x-\cdot)\|_{L^{1/h_d \dots (1/h_1)(1/h_0)}}^{1-\theta} \|^2 \\ &\leq |x-y|^{2\theta} \left\| \sum_{i=1}^d \left| \frac{\partial}{\partial x_i} \mathcal{G}_{\alpha,\delta}(t-\cdot, \xi-\cdot) \right| \right\|^\theta \\ & \quad \times \|\mathcal{G}_{\alpha,\delta}(t-\cdot, x-\cdot)\|_{L^{1/h_d \dots (1/h_1)(1/h_0)}}^{1-\theta} \|^2 \\ &\leq C_{T,H,\theta} |x-y|^{2\theta} \\ & \quad \times \sum_{i=1}^d \left\| \left| \frac{\partial}{\partial x_i} \mathcal{G}_{\alpha,\delta}(t-\cdot, \xi-\cdot) \right| \right\|^\theta \\ & \quad \times \|\mathcal{G}_{\alpha,\delta}(t-\cdot, x-\cdot)\|_{L^{1/h_d \dots (1/h_1)(1/h_0)}}^{1-\theta} \|^2. \end{aligned} \quad (75)$$

Using Lemma 3, one has

$$\begin{aligned} & \int_{\mathbb{R}} \left| \frac{\partial}{\partial x_d} \mathcal{G}_{\alpha,\delta}(t-r, \xi-z) \right|^{\theta/h_d} \\ & \quad \times |\mathcal{G}_{\alpha,\delta}(t-r, x-z)|^{(1-\theta)/h_d} dz_d \end{aligned}$$

$$\begin{aligned}
&= \prod_{i=1}^{d-1} |G_{\alpha_i, \delta_i}(t-r, \xi_i - z_i)|^{1/h_d} \\
&\quad \times \int_{\mathbb{R}} (t-r)^{-(2\theta/\alpha_d h_d) - (1-\theta)/\alpha_d h_d} \\
&\quad \times \left| \frac{\partial}{\partial x_d} G_{\alpha_d, \delta_d}(1, (t-r)^{-(1/\alpha_d)}(\xi_d - z_d)) \right|^{\theta/h_d} \\
&\quad \times |G_{\alpha_d, \delta_d}(1, (t-r)^{-(1/\alpha_d)}(\xi_d - z_d))|^{(1-\theta)/h_d} dz_d \\
&= \prod_{i=1}^{d-1} |G_{\alpha_i, \delta_i}(t-r, x_i - z_i)|^{1/h_d} \\
&\quad \times \int_{\mathbb{R}} (t-r)^{-((1-\theta)/\alpha_d h_d) + 1/\alpha_d} \\
&\quad \times \left| \frac{\partial}{\partial x_d} G_{\alpha_d, \delta_d}(1, \xi_d - z_d) \right|^{\theta/h_d} \\
&\quad \times |G_{\alpha_d, \delta_d}(1, z_d)|^{(1-\theta)/h_d} dz_d \\
&= \prod_{i=1}^{d-1} |G_{\alpha_i, \delta_i}(t-r, x_i - z_i)|^{1/h_d} \\
&\quad \times (t-r)^{-(2\theta/\alpha_d h_d) - (1-\theta)/\alpha_d h_d + 1/\alpha_d} \\
&\quad \times \int_{\mathbb{R}} \left| \frac{\partial}{\partial x_d} G_{\alpha_d, \delta_d}(1, \xi_d - z_d) \right|^{\theta/h_d} \\
&\quad \times |G_{\alpha_d, \delta_d}(1, z_d)|^{(1-\theta)/h_d} dz_d \\
&\leq C \prod_{i=1}^{d-1} |G_{\alpha_i, \delta_i}(t-r, x_i - z_i)|^{1/h_d} \\
&\quad \times (t-r)^{-(2\theta/\alpha_d h_d) - (1-\theta)/\alpha_d h_d + 1/\alpha_d}.
\end{aligned} \tag{76}$$

Using induction, one can get

$$\begin{aligned}
&\left\| \left| \frac{\partial}{\partial x_i} \mathcal{G}_{\alpha, \delta}(t-\cdot, y-\cdot) \right|^{\theta} \right. \\
&\quad \times \left. |\mathcal{G}_{\alpha, \delta}(t-\cdot, x-\cdot)|^{1-\theta} \right\|_{L^{1/h_d - (1/h_1)(1/h_0)}}^2 \\
&= \left(\int_0^T \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \right. \right. \right. \\
&\quad \left. \left. \left. \cdots \left(\int_{\mathbb{R}} \left| \frac{\partial}{\partial x_i} \mathcal{G}_{\alpha, \delta}(t-r, y-z) \right|^{\theta/h_d} \right. \right. \right. \right. \\
&\quad \left. \left. \left. \times |\mathcal{G}_{\alpha, \delta}(t-r, x-z)|^{(1-\theta)/h_d} dz_d \right)^{h_d/h_{d-1}} \right. \right. \\
&\quad \left. \left. \left. \cdots dz_2 \right)^{h_2/h_1} dz_1 \right)^{h_1/h_0} dr \right)^{2h_0} \\
&\leq C_{T, H, \alpha, \delta} \left(\int_0^T (t-r)^{\sum_{j=1}^d ((h_j-1)/h_0 \alpha_j - \theta/\alpha_i h_0)} dr \right)^{2h_0}.
\end{aligned} \tag{77}$$

Therefore, if $\sum_{j=1}^d ((h_j-1)/h_0 \alpha_j - \theta/\alpha_i h_0 + 1) > 0$, that is, $\theta < h_0 \alpha_i (1 + \sum_{j=1}^d ((h_j-1)/h_0 \alpha_j))$, $i = 1, 2, \dots, d$, we have

$$III_{1,1} \leq C_{T, H, \theta, \alpha, \delta} |x-y|^{2\theta}. \tag{78}$$

Similarly,

$$III_{1,2} \leq C_{T, H, \theta, \alpha, \delta} |x-y|^{2\theta}. \tag{79}$$

Together with (78)-(79), one gets

$$III_1 \leq C_{T, H, \theta, \alpha, \delta} |x-y|^{2\theta}. \tag{80}$$

As for III_2 , by letting $\mu \in (0, 1)$, we have

$$\begin{aligned}
III_2 &= E \left[\left| \int_0^s \int_{\mathbb{R}^d} (\mathcal{G}_{\alpha, \delta}(t-r, y-z) - \mathcal{G}_{\alpha, \delta}(s-r, y-z)) \right. \right. \\
&\quad \left. \left. \times u(r, z) dB^H(r, z) \right|^2 \right] \\
&= \int_{[0, s]^2} \int_{\mathbb{R}^{2d}} \varphi_{h_0}(r-\bar{r}) \varphi_H(z-\bar{z}) \\
&\quad \times |\mathcal{G}_{\alpha, \delta}(t-r, y-z) - \mathcal{G}_{\alpha, \delta}(s-r, y-z)| \\
&\quad \times |\mathcal{G}_{\alpha, \delta}(t-\bar{r}, y-\bar{z}) - \mathcal{G}_{\alpha, \delta}(s-\bar{r}, y-\bar{z})| \\
&\quad \times E(|u(r, z) u(\bar{r}, \bar{z})|) dz d\bar{z} dr d\bar{r} \\
&\leq C_T \int_{[0, T]^2} \varphi_{h_0}(r-\bar{r}) dr d\bar{r} \\
&\quad \times \int_{\mathbb{R}^{2d}} \varphi_H(z-\bar{z}) \\
&\quad \times |\mathcal{G}_{\alpha, \delta}(t-r, y-z) - \mathcal{G}_{\alpha, \delta}(s-r, y-z)| \\
&\quad \times |\mathcal{G}_{\alpha, \delta}(t-\bar{r}, y-\bar{z}) \\
&\quad \quad - \mathcal{G}_{\alpha, \delta}(s-\bar{r}, y-\bar{z})| dz d\bar{z} \\
&= C_T \|\mathcal{G}_{\alpha, \delta}(t-\cdot, y-\cdot) \\
&\quad \quad - \mathcal{G}_{\alpha, \delta}(s-\cdot, y-\cdot)\|_{L^{1/h_d - (1/h_1)(1/h_0)}}^2 \\
&\leq C_T \left(\|\mathcal{G}_{\alpha, \delta}(t-\cdot, y-\cdot) - \mathcal{G}_{\alpha, \delta}(s-\cdot, y-\cdot)\|^\mu \right. \\
&\quad \times \|\mathcal{G}_{\alpha, \delta}(t-\cdot, y-\cdot)\|^{1-\mu} \| \cdot \|_{L^{1/h_d - (1/h_1)(1/h_0)}}^2 \\
&\quad \left. + \|\mathcal{G}_{\alpha, \delta}(t-\cdot, y-\cdot) - \mathcal{G}_{\alpha, \delta}(s-\cdot, y-\cdot)\|^\mu \right. \\
&\quad \times \|\mathcal{G}_{\alpha, \delta}(s-\cdot, y-\cdot)\|^{1-\mu} \| \cdot \|_{L^{1/h_d - (1/h_1)(1/h_0)}}^2 \left. \right) \\
&:= C_T (III_{2,1} + III_{2,2}).
\end{aligned} \tag{81}$$

By the mean-value theorem, it holds that

$$\begin{aligned} \text{III}_{2,1} &= \left\| |t-s|^\mu \left| \frac{\partial}{\partial t} \mathcal{G}_{\alpha,\delta}(\rho-\cdot, y-\cdot) \right|^\mu \right. \\ &\quad \cdot \left. |\mathcal{G}_{\alpha,\delta}(t-\cdot, y-\cdot)|^{1-\mu} \right\|_{L^{1/h_d-(1/h_1)(1/h_0)}}^2. \end{aligned} \quad (82)$$

Note that

$$\mathcal{G}_{\alpha,\delta}(t, y) = t^{-\sum_{i=1}^d (1/\alpha_i)} \mathcal{G}_{\alpha,\delta}\left(1, t^{-\sum_{i=1}^d (1/\alpha_i)} y\right). \quad (83)$$

Hence, one can attain

$$\begin{aligned} \frac{\partial}{\partial t} \mathcal{G}_{\alpha,\delta}(t, y) &= \left(-\sum_{i=1}^d \frac{1}{\alpha_i} \right) t^{-1-\sum_{i=1}^d (1/\alpha_i)} \mathcal{G}_{\alpha,\delta}\left(1, t^{-\sum_{i=1}^d (1/\alpha_i)} y\right) \\ &\quad + \left(-\sum_{i=1}^d \frac{1}{\alpha_i} \right) t^{-1-2\sum_{i=1}^d (1/\alpha_i)} \\ &\quad \times \left[\sum_{i=1}^d y_i \frac{\partial}{\partial y_i} \mathcal{G}_{\alpha,\delta}\left(1, t^{-\sum_{i=1}^d (1/\alpha_i)} y\right) \right]. \end{aligned} \quad (84)$$

After induction, therefore, one has

$$\begin{aligned} &\left\| \frac{\partial}{\partial t} \mathcal{G}_{\alpha,\delta}(\rho-\cdot, y-\cdot) \right\|^\mu \cdot \left\| \mathcal{G}_{\alpha,\delta}(t-\cdot, y-\cdot) \right\|_{L^{1/h_d-(1/h_1)(1/h_0)}}^{1-\mu} \\ &\leq C_T \left(\left\| (t-r)^{-1-\sum_{i=1}^d (1/\alpha_i)} \right. \right. \\ &\quad \times \left. \left. \mathcal{G}_{\alpha,\delta}\left(1, (\rho-r)^{-\sum_{i=1}^d (1/\alpha_i)} (y-z)\right) \right\|^\mu \right. \\ &\quad \cdot \left. \left. |\mathcal{G}_{\alpha,\delta}(t-r, y-z)|^{1-\mu} \right\|_{L^{1/h_d-(1/h_1)(1/h_0)}}^2 \right. \\ &\quad + \left\| (t-r)^{-1-2\sum_{i=1}^d (1/\alpha_i)} \right. \\ &\quad \times \left[\sum_{i=1}^d (y_i - z_i) \frac{\partial}{\partial y_i} \right. \\ &\quad \times \left. \left. \mathcal{G}_{\alpha,\delta}\left(1, (\rho-r)^{-\sum_{i=1}^d (1/\alpha_i)} (y-z)\right) \right\|^\mu \right. \\ &\quad \cdot \left. \left. |\mathcal{G}_{\alpha,\delta}(t-r, y-z)|^{1-\mu} \right\|_{L^{1/h_d-(1/h_1)(1/h_0)}}^2 \right) \end{aligned}$$

$$\begin{aligned} &= C_T \left(\int_0^T \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \right. \right. \right. \\ &\quad \left. \left. \left. \dots \left(\int_{\mathbb{R}} |t-r|^{-1-\sum_{i=1}^d (1/\alpha_i)} \right. \right. \right. \right. \\ &\quad \left. \left. \left. \times \mathcal{G}_{\alpha,\delta}\left(1, (\rho-r)^{-\sum_{i=1}^d (1/\alpha_i)} (y-z)\right) \right\|^\mu \right. \right. \\ &\quad \left. \left. \left. \cdot |\mathcal{G}_{\alpha,\delta}(t-r, y-z)|^{(1-\mu)/h_d} dz_d \right)^{h_d/h_{d-1}} \right. \right. \\ &\quad \left. \left. \left. \dots dz_2 \right)^{h_2/h_1} dz_1 \right)^{h_1/h_0} dr \right)^{2h_0} \\ &+ C_T \left(\int_0^T \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \right. \right. \right. \\ &\quad \left. \left. \left. \dots \left(\int_{\mathbb{R}} |t-r|^{-1-2-\sum_{i=1}^d (1/\alpha_i)} \right. \right. \right. \right. \\ &\quad \left. \left. \left. \times \left[\sum_{i=1}^d (y_i - z_i) \frac{\partial}{\partial y_i} \mathcal{G}_{\alpha,\delta} \right. \right. \right. \right. \\ &\quad \left. \left. \left. \times \left(1, (\rho-r)^{-\sum_{i=1}^d (1/\alpha_i)} (y-z)\right) \right\|^\mu \right. \right. \\ &\quad \left. \left. \left. \cdot |\mathcal{G}_{\alpha,\delta}(t-r, y-z)|^{(1-\mu)/h_d} dz_d \right)^{h_d/h_{d-1}} \right. \right. \\ &\quad \left. \left. \left. \dots dz_2 \right)^{h_2/h_1} dz_1 \right)^{h_1/h_0} dr \right)^{2h_0} \\ &\leq C_{T,H,\alpha,\delta} \left(\int_0^T (t-r)^{(-\mu-\sum_{i=1}^d ((1-h_i)/\alpha_i))(1/h_0)} dr \right)^{2h_0} \\ &\quad + C_{T,H,\alpha,\delta} \left(\int_0^T (t-r)^{[-\mu(1+\sum_{i=1}^d (1/\alpha_i))-\sum_{i=1}^d ((1-h_i)/\alpha_i)](1/h_0)} dr \right)^{2h_0}. \end{aligned} \quad (85)$$

Thus, if

$$\left[-\mu \left(1 + \sum_{i=1}^d \frac{1}{\alpha_i} \right) - \sum_{i=1}^d \frac{1-h_i}{\alpha_i} \right] \frac{1}{h_0} + 1 > 0, \quad (86)$$

that is,

$$\mu < \frac{h_0 - \sum_{i=1}^d ((1-h_i)/\alpha_i)}{1 + \sum_{i=1}^d (1/\alpha_i)}, \quad (87)$$

then

$$\text{III}_{2,1} \leq C_{\alpha,\delta,H,T,\mu} |t-s|^{2\mu}. \quad (88)$$

Similarly, we can check

$$\text{III}_{2,2} \leq C_{\alpha,\delta,H,T,\mu} |t-s|^{2\mu}. \quad (89)$$

From the estimates (78)-(79), it follows that

$$\text{III}_2 \leq C_{\alpha,\delta,H,T,\mu} |t-s|^{2\mu}. \quad (90)$$

This completes the proof of the theorem. \square

5. Absolute Continuity of the Solution

In this section, we will prove that the law of the solution for (2) is absolutely continuous with respect to Lebesgue measure on \mathbb{R}^d .

Theorem 15. *Assume that $2h_0 + \sum_{i=1}^d ((h_i - 1)/\alpha_i) > 1$ and $\sum_{i=1}^d (1/\alpha_i) < 1$. Then, for $(t, x) \in [0, T] \times \mathbb{R}^d$, the law of the solution $u(t, x)$ of (2) is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^d .*

Let us start with two lemmas for the proof of our main result.

Lemma 16. *Assume $2h_0 + \sum_{i=1}^d ((h_i - 1)/\alpha_i) > 1$. For $(t, x) \in [0, T] \times \mathbb{R}^d$, the solution $u(t, x) \in \mathcal{D}^{1,2}$. Its Malliavin derivative satisfies, if $s \leq t$,*

$$\begin{aligned} D_{s,y}u(t, x) &= \mathcal{G}_{\alpha,\delta}(t-s, x-y)u(s, y) \\ &+ \int_s^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t-r, x-z) D_{s,y}u(r, z) dB^H \\ &\times (r, z). \end{aligned} \quad (91)$$

And if $s > t$, $D_{s,y}u(t, x) = 0$.

Proof. For each $n \geq 1$, let $\{u^n(t, x), n \geq 1\}$ be the Picard iteration defined as

$$\begin{aligned} u^0(t, x) &= \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t, x-y)u(0, y) dy, \quad (92) \\ u^{n+1}(t, x) &= u^0(t, x) \\ &+ \int_0^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t-s, x-y)u^n(s, y) dB^H \\ &\times (s, y). \end{aligned} \quad (93)$$

Now, we will prove $u^n(t, x) \in \mathcal{D}^{1,2}$ by induction. Suppose that, for $n \geq 0$ and $(t, x) \in [0, T] \times \mathbb{R}^d$, $u^n(t, x) \in \mathcal{D}^{1,2}$ and satisfies

$$\sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \mathbb{E} \left(\int_0^T \int_{\mathbb{R}^d} |D_{s,y}u^n(t, x)|^2 dy ds \right) < \infty. \quad (94)$$

We take the Malliavin derivative at both sides of (93). Then, for $s \leq t$,

$$\begin{aligned} D_{s,y}u^{n+1}(t, x) &= \mathcal{G}_{\alpha,\delta}(t-s, x-y)u^n(s, y) \\ &+ \int_s^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t-r, x-z) D_{s,y} \\ &\times u^n(r, z) dB^H(r, z) \end{aligned} \quad (95)$$

and, for $s > t$, $D_{s,y}u^{n+1}(t, x) = 0$. Note that

$$\begin{aligned} &\mathbb{E} \left(\int_0^t \int_{\mathbb{R}^d} |D_{s,y}u^{n+1}(t, x)|^2 dy ds \right) \\ &= \mathbb{E} \left(\int_0^t \int_{\mathbb{R}^d} \left| \mathcal{G}_{\alpha,\delta}(t-s, x-y)u^n(s, y) \right. \right. \\ &\quad \left. \left. + \int_s^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t-r, x-z) \right. \right. \\ &\quad \left. \left. \times D_{s,y}u^n(r, z) dB^H(r, z) \right|^2 dy ds \right) \\ &\leq 2 \left[\mathbb{E} \left(\int_0^t \int_{\mathbb{R}^d} |\mathcal{G}_{\alpha,\delta}(t-s, x-y)u^n(s, y)|^2 dy ds \right) \right. \\ &\quad \left. + \mathbb{E} \left(\int_0^t \int_{\mathbb{R}^d} \left| \int_s^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t-r, x-z) \right. \right. \right. \\ &\quad \left. \left. \times D_{s,y}u^n(r, z) dB^H(r, z) \right|^2 dy ds \right) \right] \\ &:= 2 \left(B_n^1(t, x) + B_n^2(t, x) \right). \end{aligned} \quad (96)$$

Recall that $u^n(t, x) \in \mathcal{S}_0$; that is, $u^n(t, x) \in L^2(\Omega)$. Thus, we have

$$\begin{aligned} B_n^1(t, x) &= \int_0^t \int_{\mathbb{R}^d} \mathbb{E} |\mathcal{G}_{\alpha,\delta}(t-s, x-y)u^n(s, y)|^2 dy ds \\ &\leq \int_0^T \int_{\mathbb{R}^d} |\mathcal{G}_{\alpha,\delta}(t-s, x-y)|^2 dy ds \\ &< \infty. \end{aligned} \quad (97)$$

By the Fubini theorem, one gets

$$\begin{aligned} B_n^2(t, x) &= \mathbb{E} \left(\int_0^t \int_{\mathbb{R}^d} \left| \int_s^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha,\delta}(t-r, x-z) D_{s,y} \right. \right. \\ &\quad \left. \left. \times u^n(r, z) dB^H(r, z) \right|^2 dy ds \right) \end{aligned}$$

$$\begin{aligned}
 &= \int_0^t \int_{\mathbb{R}^d} \left[\int_{[s,t]^2} \int_{\mathbb{R}^{2d}} \varphi_{h_0}(r-\bar{r}) \varphi_H(z-\bar{z}) \right. \\
 &\quad \times |\mathcal{G}_{\alpha,\delta}(t-r, x-z)| \\
 &\quad \times |\mathcal{G}_{\alpha,\delta}(t-\bar{r}, x-\bar{z})| \\
 &\quad \times E \left(\left| D_{s,y} u^n(r,z) D_{s,y} u^n(\bar{r}, \bar{z}) \right| \right) \\
 &\quad \left. \times dz d\bar{z} dr d\bar{r} \right] dy ds \\
 &\leq C \int_0^t \int_{\mathbb{R}^d} \left[\int_{[s,t]^2} \int_{\mathbb{R}^{2d}} \varphi_{h_0}(r-\bar{r}) \varphi_H(z-\bar{z}) \right. \\
 &\quad \times |\mathcal{G}_{\alpha,\delta}(t-r, x-z)| \\
 &\quad \times |\mathcal{G}_{\alpha,\delta}(t-\bar{r}, x-\bar{z})| \\
 &\quad \left. \times E \left(\left| D_{s,y} u^n(r,z) \right|^2 \right) dz d\bar{z} dr d\bar{r} \right] dy ds \\
 &\leq C \int_{[s,t]^2} \int_{\mathbb{R}^{2d}} \sup_{z \in \mathbb{R}^d} \left(\int_0^t \int_{\mathbb{R}^d} E \left(\left| D_{s,y} u^n(r,z) \right|^2 \right) dy ds \right) \\
 &\quad \times \varphi_{h_0}(r-\bar{r}) \varphi_H(z-\bar{z}) \\
 &\quad \times |\mathcal{G}_{\alpha,\delta}(t-r, x-z)| \\
 &\quad \times |\mathcal{G}_{\alpha,\delta}(t-\bar{r}, x-\bar{z})| dz d\bar{z} dr d\bar{r}.
 \end{aligned} \tag{98}$$

Note that

$$\begin{aligned}
 &\int_{\mathbb{R}^{2d}} \varphi_H(z-\bar{z}) |\mathcal{G}_{\alpha,\delta}(t-r, y-z)| \\
 &\quad \times |\mathcal{G}_{\alpha,\delta}(t-\bar{r}, y-\bar{z})| dz d\bar{z} \\
 &= \int_{\mathbb{R}^{2(d-1)}} \prod_{i=1}^{d-1} \varphi_{h_i}(z_i - \bar{z}_i) |G_{\alpha_i, \delta_i}(t-r, x_i - z_i)| \\
 &\quad \times |G_{\alpha_i, \delta_i}(t-\bar{r}, x_i - \bar{z}_i)| dz_1 d\bar{z}_1 \cdots dz_{d-1} d\bar{z}_{d-1} \\
 &\quad \times \int_{\mathbb{R}^2} \varphi_{h_d}(z_d - \bar{z}_d) |G_{\alpha_d, \delta_d}(t-r, x_d - z_d)| \\
 &\quad \times |G_{\alpha_d, \delta_d}(t-\bar{r}, x_d - \bar{z}_d)| dz_d d\bar{z}_d, \\
 &\int_{\mathbb{R}^2} \varphi_{h_d}(z_d - \bar{z}_d) |G_{\alpha_d, \delta_d}(t-r, x_d - z_d)| \\
 &\quad \times |G_{\alpha_d, \delta_d}(t-\bar{r}, x_d - \bar{z}_d)| dz_d d\bar{z}_d \\
 &\leq C \|G_{\alpha_d, \delta_d}(t-r, x_d - \cdot)\|_{L^{1/h_d}} \\
 &\quad \times \|G_{\alpha_d, \delta_d}(t-\bar{r}, x_d - \cdot)\|_{L^{1/h_d}} \\
 &= C(t-r)^{(h_d-1)/\alpha_d} (t-\bar{r})^{(h_d-1)/\alpha_d}.
 \end{aligned} \tag{99}$$

So by induction, if $2h_0 + \sum_{i=1}^d ((h_i - 1)/\alpha_i) > 1$, we have

$$\begin{aligned}
 &B_n^2(t, x) \\
 &\leq C \int_{[s,t]^2} \sup_{z \in \mathbb{R}^d} \left(\int_0^t \int_{\mathbb{R}^d} E \left(\left| D_{s,y} u^n(r,z) \right|^2 \right) dy ds \right) \\
 &\quad \times \varphi_{h_0}(r-\bar{r}) \times (t-r)^{\sum_{i=1}^d ((h_i-1)/\alpha_i)} \\
 &\quad \times (t-\bar{r})^{\sum_{i=1}^d ((h_i-1)/\alpha_i)} dr d\bar{r} \\
 &\leq C \int_0^t \sup_{z \in \mathbb{R}^d} \left(\int_0^t \int_{\mathbb{R}^d} E \left(\left| D_{s,y} u^n(r,z) \right|^2 \right) dy ds \right) \\
 &\quad \times (t-r)^{\sum_{i=1}^d ((h_i-1)/\alpha_i)} \\
 &\quad \times \left(\int_0^r (r-\bar{r})^{2h_0-2} (t-\bar{r})^{\sum_{i=1}^d ((h_i-1)/\alpha_i)} d\bar{r} \right) dr \\
 &\leq C \int_0^t \sup_{z \in \mathbb{R}^d} \left(\int_0^t \int_{\mathbb{R}^d} E \left(\left| D_{s,y} u^n(r,z) \right|^2 \right) dy ds \right) \\
 &\quad \times (t-r)^{\sum_{i=1}^d ((h_i-1)/\alpha_i)} dr \\
 &= C \int_0^t \sup_{z \in \mathbb{R}^d} \left(\int_0^t \int_{\mathbb{R}^d} E \left(\left| D_{s,y} u^n(r,z) \right|^2 \right) dy ds \right) \\
 &\quad \times g_1(t-r) dr,
 \end{aligned} \tag{100}$$

where $\int_0^T t^{\sum_{i=1}^d ((h_i-1)/\alpha_i)} dt < +\infty$, since $\sum_{i=1}^d ((h_i - 1)/\alpha_i + 1) > 1 + (1 - 2h_0) > 0$. Together with (96)–(100), one gets

$$\begin{aligned}
 &\sup_{z \in \mathbb{R}^d} \mathbb{E} \left(\int_0^t \int_{\mathbb{R}^d} \left| D_{s,y} u^{n+1}(t, x) \right|^2 dy ds \right) \\
 &\leq C + C \int_0^t \sup_{z \in \mathbb{R}^d} \left(\int_0^t \int_{\mathbb{R}^d} E \left(\left| D_{s,y} u^n(r,z) \right|^2 \right) dy ds \right) \\
 &\quad \times g_1(t-r) dr.
 \end{aligned} \tag{101}$$

By the version of Gronwall's lemma given by Dalang [47], one has

$$\sup_{n \geq 0} \sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \mathbb{E} \left(\int_0^T \int_{\mathbb{R}^d} \left| D_{s,y} u^n(t, x) \right|^2 dy ds \right) < \infty. \tag{102}$$

Note that $E \|Du^n(t, x)\|_{\mathcal{H}_1} > 0$, if and only if $E \|Du^n(t, x)\|_{L^2([0,T] \times \mathbb{R}^d)} > 0$. Equation (102) implies $E \|Du^n(t, x)\|_{\mathcal{H}_1} > 0$. On the other hand, we note that $u^n(t, x) \in \mathcal{S}_0$; that is, $u^n(t, x) \in L^2(\Omega)$. Thus, we have proved $u^n(t, x) \in \mathcal{D}^{1,2}$. By Lemma 1.2.3 in Nualart [29], we get $u(t, x) \in \mathcal{D}^{1,2}$ and the sequence $Du^n(t, x)$ converges to $Du(t, x)$ in the weak topology of $L^2([0, T] \times \mathbb{R}^d)$. This concludes the proof of Lemma 16. \square

Lemma 17. Let $\sum_{i=1}^d (1/\alpha_i) < 1$. Then, there exists some constant $C > 0$ such that, for $0 < \varepsilon < t$,

$$\sup_{(s,y) \in [t-\varepsilon, t] \times \mathbb{R}^d} \mathbb{E} \left(\int_{t-\varepsilon}^t \int_{\mathbb{R}^d} |D_{r,z} u(s, y)|^2 dz dr \right) < C \varepsilon^{1-\sum_{i=1}^d (1/\alpha_i)}. \quad (103)$$

Proof. For $s \in [t - \varepsilon, t]$, set $I^\varepsilon(s, y) = \mathbb{E} \left(\int_{t-\varepsilon}^s \int_{\mathbb{R}^d} |D_{r,z} u(s, y)|^2 dz dr \right)$. By (102), we have

$$\sup_{(s,y) \in [0, T] \times \mathbb{R}^d} I^\varepsilon(s, y) < \infty. \quad (104)$$

Recall (91). Then,

$$\begin{aligned} I^\varepsilon(s, y) &\leq 2 \left[\left(\mathbb{E} \int_{t-\varepsilon}^s \int_{\mathbb{R}^d} |\mathcal{G}_{\alpha, \delta}(s-r, y-z) u(s, y)|^2 dz dr \right) \right. \\ &\quad + \mathbb{E} \left(\int_{t-\varepsilon}^s \int_{\mathbb{R}^d} \left| \int_r^s \int_{\mathbb{R}^d} \mathcal{G}_{\alpha, \delta}(s-r_1, y-x_1) \right. \right. \\ &\quad \quad \times D_{r,z} u(r_1, x_1) dB^H \\ &\quad \quad \left. \left. \times (r_1, z_1) \right|^2 dz dr \right) \Big] \\ &:= 2 \left(I^{\varepsilon,1}(s, y) + I^{\varepsilon,2}(s, y) \right). \end{aligned} \quad (105)$$

Note that

$$\begin{aligned} I^{\varepsilon,1}(s, y) &= \int_{t-\varepsilon}^s \int_{\mathbb{R}^d} \mathbb{E} |\mathcal{G}_{\alpha, \delta}(s-r, y-z) u(s, y)|^2 dz dr \\ &\leq \int_{t-\varepsilon}^t \int_{\mathbb{R}^d} |\mathcal{G}_{\alpha, \delta}(s-r, y-z)|^2 dz dr \\ &\leq C \int_{t-\varepsilon}^t (s-r)^{-\sum_{i=1}^d (1/\alpha_i)} dr \\ &< C \varepsilon^{1-\sum_{i=1}^d (1/\alpha_i)}, \end{aligned} \quad (106)$$

$$I^{\varepsilon,2}(s, y) \leq C \varepsilon^{1-\sum_{i=1}^d ((h_i-1)/\alpha_i)} \int_{t-\varepsilon}^s \sup_{x_1 \in \mathbb{R}^d} I^\varepsilon(r_1, x_1) dr_1. \quad (107)$$

Together with (105)–(107), one gets

$$\begin{aligned} \sup_{y \in \mathbb{R}^d} I^\varepsilon(s, y) &\leq \varepsilon^{1-\sum_{i=1}^d (1/\alpha_i)} + C \varepsilon^{1-\sum_{i=1}^d ((h_i-1)/\alpha_i)} \\ &\quad \times \int_{t-\varepsilon}^s \sup_{x_1 \in \mathbb{R}^d} I^\varepsilon(r_1, x_1) dr_1. \end{aligned} \quad (108)$$

Then by the Gronwall's lemma, we get (103). \square

Proof of Theorem 15. In order to prove Theorem 15, by Theorem 2.1.3 in Nualart [29], we only need to check that

$$\|Du(t, x)\|_{\mathcal{H}_1} > 0, \quad \text{a.s.} \quad (109)$$

Note that $\|Du(t, x)\|_{\mathcal{H}_1} > 0$, a.s., if and only if $\|Du(t, x)\|_{L^2([0, T] \times \mathbb{R}^d)} > 0$. Hence, we only need to prove that

$$\int_0^t \int_{\mathbb{R}^d} |D_{r,z} u(t, x)|^2 dz dr > 0, \quad \text{a.s.} \quad (110)$$

In (91), we note that

$$\begin{aligned} \int_0^t \int_{\mathbb{R}^d} |D_{r,z} u(t, x)|^2 dz dr &\geq \int_{t-\varepsilon}^t \int_{\mathbb{R}^d} |D_{r,z} u(t, x)|^2 dz dr \\ &\geq C \left(J^{\varepsilon,1}(t, x) - J^{\varepsilon,2}(t, x) \right), \end{aligned} \quad (111)$$

where

$$\begin{aligned} J^{\varepsilon,1}(t, x) &= \int_{t-\varepsilon}^t \int_{\mathbb{R}^d} |\mathcal{G}_{\alpha, \delta}(t-r, x-z) u(r, z)|^2 dz dr, \\ J^{\varepsilon,2}(t, x) &:= \int_{t-\varepsilon}^t \int_{\mathbb{R}^d} \left| \int_r^t \int_{\mathbb{R}^d} \mathcal{G}_{\alpha, \delta}(t-r_1, x-z_1) \right. \\ &\quad \left. \times D_{r,z} u(r_1, z_1) dB^H(r_1, z_1) \right|^2 dz dr. \end{aligned} \quad (112)$$

Similar to the proof of (106), there exists some constant $C > 0$ such that

$$\begin{aligned} \mathbb{E} \left(J^{\varepsilon,1}(t, x) \right) &\leq C \int_{t-\varepsilon}^t \int_{\mathbb{R}^d} |\mathcal{G}_{\alpha, \delta}(t-r, x-z)|^2 dz dr \\ &< C \varepsilon^{1-\sum_{i=1}^d (1/\alpha_i)}. \end{aligned} \quad (113)$$

By Lemma 17,

$$\begin{aligned} \mathbb{E} \left(J^{\varepsilon,2}(t, x) \right) &\leq C \int_{t-\varepsilon}^t \varepsilon^{\sum_{i=1}^d ((h_i-1)/\alpha_i)} \\ &\quad \times \mathbb{E} \left(\int_{t-\varepsilon}^t \int_{\mathbb{R}^d} |D_{r,z} u(r_1, z_1)|^2 dz_1 dr_1 \right) dr \\ &\leq C \varepsilon^{2-\sum_{i=1}^d (1/\alpha_i) + \sum_{i=1}^d ((h_i-1)/\alpha_i)}. \end{aligned} \quad (114)$$

Therefore, for each fixed $\varepsilon_0 > 0$,

$$\begin{aligned} P \left(\int_0^T \int_{\mathbb{R}^d} |D_{r,z} u(t, x)|^2 dz dr > 0 \right) &\geq \sup_{\varepsilon \in (0, \varepsilon_0)} P \left(C \left(J^{\varepsilon,1}(t, x) - J^{\varepsilon,2}(t, x) \right) > 0 \right) \\ &\geq \sup_{\varepsilon \in (0, \varepsilon_0)} P \left(J^{\varepsilon,2}(t, x) < C \varepsilon^{1-\sum_{i=1}^d (1/\alpha_i)} \right) \\ &\geq 1 - \inf_{\varepsilon \in (0, \varepsilon_0)} \left\{ \frac{1}{C \varepsilon^{1-\sum_{i=1}^d (1/\alpha_i)}} \mathbb{E} \left(J^{\varepsilon,2}(t, x) \right) \right\} \\ &\geq 1 - \inf_{\varepsilon \in (0, \varepsilon_0)} \frac{1}{C} \varepsilon^{1+\sum_{i=1}^d ((h_i-1)/\alpha_i)} = 1. \end{aligned} \quad (115)$$

This completes the proof of the theorem. \square

6. Conclusions and Discussions

We have obtained a unique solution of the stochastic fractional heat equation driven by fractional noise. The Lyapunov exponent of the solution has been estimated. In addition, we have discussed the Hölder continuity of the solution on both space and time parameters. Moreover, the absolute continuity of the solution has been obtained. Our future work will consider its application issues in computer networks, such as network traffic of fractional noise type passing through network servers.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

Xichao Sun is partially supported by Mathematical Tianyuan Foundation of China (no. 11426036) and Natural Science Foundation of Anhui Province (no. 1408085QA10). Ming Li acknowledges the supports in part by the National Natural Science Foundation of China under the Project Grant nos. 61272402, 61070214, and 60873264.

References

- [1] I. Podlubny, *Fractional Differential Equations*, Technical University of Kosice, Košice, Slovak Republic, 1999.
- [2] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives, Theory and Applications*, Gordon and Breach Science Publishers, 1993.
- [3] M. H. Heydari, M. R. Hooshmandasl, F. M. Ghaini, and C. Cattani, "A computational method for solving stochastic Itô-Volterra integral equations based on stochastic operational matrix for generalized hat basis functions," *Journal of Computational Physics*, vol. 270, pp. 402–415, 2014.
- [4] M. H. Heydari, M. R. Hooshmandasl, F. Mohammadi, and C. Cattani, "Wavelets method for solving systems of nonlinear singular fractional Volterra integro-differential equations," *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 1, pp. 37–48, 2014.
- [5] C. Cattani, G. Pierro, and G. Altieri, "Entropy and multifractality for the myeloma multiple TET 2 gene," *Mathematical Problems in Engineering*, vol. 2012, Article ID 193761, 14 pages, 2012.
- [6] C. Cattani and G. Pierro, "On the fractal geometry of DNA by the binary image analysis," *Bulletin of Mathematical Biology*, vol. 75, no. 9, pp. 1544–1570, 2013.
- [7] Z. W. Liao, "Low-dosed x-ray computed tomography imaging by regularized fully spatial fractional-order Perona-Malik diffusion," *Advances in Mathematical Physics*, vol. 2013, Article ID 371868, 9 pages, 2013.
- [8] S. X. Hu, "External fractional-order gradient vector Perona-Malik diffusion for sinogram restoration of low-dosed X-ray computed tomography," *Advances in Mathematical Physics*, vol. 2013, Article ID 516919, 10 pages, 2013.
- [9] K. Sobczyk, *Stochastic Differential Equations with Applications to Physics and Engineering*, Kluwer Academic, London, UK, 1991.
- [10] M. E. Gurtin and A. C. Pipkin, "A general theory of heat conduction with finite wave speeds," *Archive for Rational Mechanics and Analysis*, vol. 31, no. 2, pp. 113–126, 1968.
- [11] J. W. Nunziato, "On heat conduction in materials with memory," *Quarterly of Applied Mathematics*, vol. 29, pp. 187–204, 1971.
- [12] J. Droniou and C. Imbert, "Fractal first-order partial differential equations," *Archive for Rational Mechanics and Analysis*, vol. 182, no. 2, pp. 299–331, 2006.
- [13] V. V. Uchaikin and V. M. Zolotarev, *Chance and Stability, Stable Distributions and Their Applications*, Modern Probability and Statistics, De Gruyter, 1999.
- [14] C. Toma, "Wavelets-computational aspects of Sterian realistic approach to uncertainty principle in high energy physics: a transient approach," *Advances in High Energy Physics*, vol. 2013, Article ID 735452, 6 pages, 2013.
- [15] E. G. Bakhoun and C. Toma, "Transient aspects of wave propagation connected with spatial coherence," *Mathematical Problems in Engineering*, vol. 2013, Article ID 691257, 5 pages, 2013.
- [16] L. Li, H. Peng, J. Kurths, Y. Yang, and H. J. Schellnhuber, "Chaos-order transition in foraging behavior of ants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 23, pp. 8392–8397, 2014.
- [17] L. Li, J. Xiao, H. Peng, Y. Yang, and Y. Chen, "Improving synchronous ability between complex networks," *Nonlinear Dynamics*, vol. 69, no. 3, pp. 1105–1110, 2012.
- [18] C. Mueller, "The heat equation with Lévy noise," *Stochastic Processes and Their Applications*, vol. 74, no. 1, pp. 67–82, 1998.
- [19] J. Wu, "Fractal Burgers equation with stable Lévy noise," in *Proceedings of the International Conference SPDE and Applications*, vol. 2, January 2004.
- [20] L. Boulanba, M. Eddahbi, and M. Mellouk, "Fractional SPDEs driven by spatially correlated noise: existence of the solution and smoothness of its density," *Osaka Journal of Mathematics*, vol. 47, no. 1, pp. 41–65, 2010.
- [21] T. Chang and K. Lee, "On a stochastic partial differential equation with a fractional Laplacian operator," *Stochastic Processes and Their Applications*, vol. 122, no. 9, pp. 3288–3311, 2012.
- [22] L. Debbi and M. Dozzi, "On the solutions of nonlinear stochastic fractional partial differential equations in one spatial dimension," *Stochastic Processes and Their Applications*, vol. 115, no. 11, pp. 1764–1781, 2005.
- [23] X. Sun, Z. Wang, and J. Cui, "On a fractional SPDE driven by fractional noise and a pure jump Lévy noise in \mathbb{R}^d ," *Abstract and Applied Analysis*, vol. 2014, Article ID 758270, 10 pages, 2014.
- [24] A. Truman and J. Wu, "Fractal Burgers equation driven by Lévy noises," in *Stochastic Partial Differential Equations and Applications*, G. da Prato and L. Tubaro, Eds., vol. 245 of *Lecture Notes in Pure and Applied Mathematics*, pp. 295–310, Chapman Hall/CRC Taylor, 2006.
- [25] J. Liu, L. Yan, and Y. Cang, "On a jump-type stochastic fractional partial differential equation with fractional noises," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 75, no. 16, pp. 6060–6070, 2012.
- [26] D. Wu, "On the solution process for a stochastic fractional partial differential equation driven by space-time white noise," *Statistics & Probability Letters*, vol. 81, no. 8, pp. 1161–1172, 2011.

- [27] F. Biagini, Y. Hu, B. Øsendal, and T. Zhang, *Stochastic Calculus for fBm and Applications, Probability and Its Application*, Springer, Berlin, Germany, 2008.
- [28] Y. S. Mishura, *Stochastic Calculus for fBm and Related Processes*, vol. 1929 of *Lecture Notes in Mathematics*, 2008.
- [29] D. Nualart, *Malliavin Calculus and Related Topics*, Springer, Berlin, Germany, 2nd edition, 2006.
- [30] Y. Hu, "Heat equations with fractional white noise potentials," *Applied Mathematics and Optimization*, vol. 43, no. 3, pp. 221–243, 2001.
- [31] D. Nualart and Y. Ouknine, "Regularization of quasilinear heat equations by a fractional noise," *Stochastics and Dynamics*, vol. 4, no. 2, pp. 201–221, 2004.
- [32] S. Tindel, C. A. Tudor, and F. Viens, "Stochastic evolution equations with fractional Brownian motion," *Probability Theory and Related Fields*, vol. 127, no. 2, pp. 186–204, 2003.
- [33] Y. Hu and D. Nualart, "Stochastic heat equation driven by fractional noise and local time," *Probability Theory and Related Fields*, vol. 143, no. 1-2, pp. 285–328, 2009.
- [34] R. M. Balan, "The stochastic wave equation with multiplicative fractional noise: a Malliavin calculus approach," *Potential Analysis*, vol. 36, no. 1, pp. 1–34, 2012.
- [35] R. M. Balan and C. A. Tudor, "Stochastic heat equation with multiplicative fractional-colored noise," *Journal of Theoretical Probability*, vol. 23, no. 3, pp. 834–870, 2010.
- [36] R. M. Balan and C. A. Tudor, "The stochastic wave equation with fractional noise: a random field approach," *Stochastic Processes and Their Applications*, vol. 120, no. 12, pp. 2468–2494, 2010.
- [37] L. Bo, Y. Jiang, and Y. Wang, "On a class of stochastic Anderson models with fractional noises," *Stochastic Analysis and Applications*, vol. 26, no. 2, pp. 256–273, 2008.
- [38] L. Bo, Y. Jiang, and Y. Wang, "Stochastic Cahn-Hilliard equation with fractional noise," *Stochastics and Dynamics*, vol. 8, no. 4, pp. 643–665, 2008.
- [39] Y. Jiang, K. Shi, and Y. Wang, "Stochastic fractional anderson models with fractional noises," *Chinese Annals of Mathematics B*, vol. 31, no. 1, pp. 101–118, 2010.
- [40] Y. Jiang, X. Wang, and Y. Wang, "On a stochastic heat equation with first order fractional noises and applications to finance," *Journal of Mathematical Analysis and Applications*, vol. 396, no. 2, pp. 656–669, 2012.
- [41] Y. Jiang, T. Wei, and X. Zhou, "Stochastic generalized Burgers equations driven by fractional noises," *Journal of Differential Equations*, vol. 252, no. 2, pp. 1934–1961, 2012.
- [42] Y. Hu, "Chaos expansion of heat equations with white noise potentials," *Potential Analysis*, vol. 16, no. 1, pp. 45–66, 2002.
- [43] L. Debbi, "On some properties of a high order fractional differential operator which is not in general selfadjoint," *Applied Mathematical Sciences*, vol. 1, no. 25–28, pp. 1325–1339, 2007.
- [44] T. Komatsu, "On the martingale problem for generators of stable processes with perturbations," *Osaka Journal of Mathematics*, vol. 21, no. 1, pp. 113–132, 1984.
- [45] J. Mémin, Y. Mishura, and E. Valkeila, "Inequalities for the moments of Wiener integrals with respect to a fractional Brownian motion," *Statistics & Probability Letters*, vol. 51, no. 2, pp. 197–206, 2001.
- [46] T. T. Wei, "The high-order SPDEs driven by multi-parameter fractional noises," *Acta Mathematica Sinica*, vol. 26, no. 10, pp. 1943–1960, 2010.
- [47] R. C. Dalang, "Extending the martingale measure stochastic integral with applications to spatially homogeneous s.p.d.e.'s," *Electronic Journal of Probability*, vol. 4, article 6, pp. 1–29, 1999.

Research Article

A New Approach for Flexible Molecular Docking Based on Swarm Intelligence

Yi Fu,^{1,2} Xiaojun Wu,² Zhiguo Chen,² Jun Sun,² Ji Zhao,¹ and Wenbo Xu²

¹Department of Electronic and Information Engineering, Wuxi City College of Vocational Technology, Wuxi, Jiangsu 214153, China

²School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Correspondence should be addressed to Yi Fu; fuyi_bright@yahoo.com

Received 29 September 2014; Accepted 23 November 2014

Academic Editor: Ezzat G. Bakhoun

Copyright © 2015 Yi Fu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Molecular docking methods play an important role in the field of computer-aided drug design. In the work, on the basis of the molecular docking program AutoDock, we present QLDock as a tool for flexible molecular docking. For the energy evaluation, the algorithm uses the binding free energy function that is provided by the AutoDock 4.2 tool. The new search algorithm combines the features of a quantum-behaved particle swarm optimization (QPSO) algorithm and local search method of Solis and Wets for solving the highly flexible protein-ligand docking problem. We compute the interaction of 23 protein-ligand complexes and compare the results with those of the QDock and AutoDock programs. The experimental results show that our approach leads to substantially lower docking energy and higher docking precision in comparison to Lamarckian genetic algorithm and QPSO algorithm alone. QPSO-ls algorithm was able to identify the correct binding mode of 74% of the complexes. In comparison, the accuracy of QPSO and LGA is 52% and 61%, respectively. This difference in performance rises with increasing complexity of the ligand. Thus, the novel algorithm QPSO-ls may be used to dock ligand with many rotatable bonds with high accuracy.

1. Introduction

Computer-aided tools have wide applications in various fields, such as biotechnology engineering, medical engineering, mathematical modeling, and electronic information [1–4]. In the drug design field, molecular docking is a useful tool in drug discovery and rational drug design. New drugs can be designed efficiently by using computer-aided docking algorithm simulations to find highly affine components that bind well to the targeted protein [5, 6]. By doing so, the biochemical reaction that the target molecule catalyzes can be altered or prevented. In the past, drug discovery process was very expensive and time consuming by *in vitro* experiments. Now, structure-based *in silico* approach allows for a faster and cheaper identification of promising drug candidates by the virtual screening of compound databases [7].

Molecular docking has been developed with primarily two parts varying: scoring function and searching method [8, 9]. Docking methods typically use an energy-based scoring function to identify the energetically most favorable ligand

conformation when bound to the target. The general hypothesis is that lower energy scores represent better protein-ligand bindings compared to higher energy values. Therefore, molecular docking can be viewed as a complex combinatorial optimization problem. The aim of combinatorial optimization is to find the ligand-binding mode with the lowest energy [10].

During the last decade, many different tools have been used for the docking problem, such as DOCK [11], GOLD [12], FlexX [13], and AutoDock [14]. Among these programs, AutoDock is a well-known example of flexible docking technique. Simulated annealing [15] was originally used within AutoDock. Later, a genetic algorithm (GA) [14] and a Lamarckian genetic algorithm (LGA) [16], which incorporates a local search, have been added. Recently, also particle swarm optimization (PSO) [17–20] and ant colony optimization (ACO) [21] have been proposed as search methods in the AutoDock program. Although there exist many different algorithms and programs, the problem is far from being solved [22, 23]. None as of now can be viewed as offering

a robust and accurate solution to the docking problem, even in the context of a rigid protein receptor. In general, the complexity of the task rises with increasing flexibility of the molecules. For flexible docking, the number of rotatable bonds of the side chains increases the dimension of the search space excessively. And if parts of the receptor are also considered to be flexible, the number of rotatable bonds of the system will be the sum of the rotatable bonds of ligand and receptor region; this situation should become more complicated.

In this study, we propose a novel search method called QPSO-ls for solving highly flexible docking problem, which is a hybrid of quantum-behaved particle swarm optimization (QPSO) and a local search method. In 2004, inspired by quantum mechanics and trajectory analysis of PSO [24], Sun et al. [25] used a strategy based on a quantum delta potential well model to sample around the previous best points and later introduced the mean best position into the algorithm and proposed a new version of PSO, quantum-behaved particle swarm optimization. QPSO is a kind of probabilistic algorithm [26]; it needs no velocity vectors for particles and has fewer parameters to adjust. It has been proved that this iterative equation leads QPSO to be globally convergent [27]. It has been shown to successfully solve a wide range of continuous optimization problems. For example, it has been used to tackle the problems of multiobjective optimization [28], neural network training [29], bioinformatics [30], and so on. In addition to the applications, many efficient strategies have been proposed to improve the performance of QPSO [31, 32]. In the study, to improve search performance, an efficient variant of the Solis and Wets local search technique [33] is incorporated into QPSO algorithm such that both techniques cooperate fully with each other. Simulation results reveal that QPSO-ls has a better convergence and a lower docked energy than QPSO algorithm alone. The environment and scoring function that are used by the algorithm are provided by AutoDock version 4.2.

The remaining part of the paper is arranged as follows. First in Section 2 we describe the algorithms and scoring function employed by the program for the docking problem. Next in Section 3 we describe input data preparation and docking protocols. The docking experimental results are discussed in Section 4. Finally, conclusions are given in Section 5.

2. Algorithm and Scoring Function

2.1. Quantum-Behaved Particle Swarm Optimization (QPSO) Scheme. Quantum-behaved particle swarm optimization (QPSO) was motivated by concepts from quantum mechanics and particle swarm optimization (PSO); it is a probabilistic optimization algorithm. In QPSO, the state of a particle is depicted by wavefunction $\psi(\vec{x}, t)$ instead of \vec{x} and \vec{v} . The dynamic behavior of the particle is widely divergent from that of the particle in traditional PSO systems in that the exact values of \vec{x} and \vec{v} cannot be determined simultaneously. We can only learn the probability of the particle's appearing in position \vec{x} from probability density function $|\psi(\vec{x}, t)|^2$,

the form of which depends on the potential field the particle lies in. And then the probability distribution function of the particle's position can be calculated through the probability density function. By employing the Monte Carlo method, the particle's position is updated according to the following equation:

$$x(t) = p \pm \frac{L}{2} \ln\left(\frac{1}{u}\right) \quad u \sim U(0, 1). \quad (1)$$

In [25], parameter L is evaluated by

$$L(t+1) = 2 \cdot \beta \cdot |p - x(t)|, \quad (2)$$

where parameter β is called the contraction-expansion coefficient, which can be tuned to control the convergence speed of the algorithms. Then we get the position update equation as

$$x(t+1) = p \pm \beta \cdot |p - x(t)| \cdot \ln\left(\frac{1}{u}\right). \quad (3)$$

Parameter L is the only parameter in QPSO algorithm. The control method of L is vital to convergence rate and performance of the algorithm. Therefore, in the later work [34], Sun et al. proposed a mainstream thought point to evaluate parameter L , the creativity of a particle. The mainstream thought point or mean best position (*mbest*) is defined as the center of gravity *gbest* position of the particle swarm. That is,

$$\begin{aligned} mbest &= \frac{1}{M} \sum_{i=1}^M P_i \\ &= \left(\frac{1}{M} \sum_{i=1}^M P_{i1}, \frac{1}{M} \sum_{i=1}^M P_{i2}, \dots, \frac{1}{M} \sum_{i=1}^M P_{id} \right), \end{aligned} \quad (4)$$

where M is the population size and P_i is the *gbest* position of particle i . The value of L is given by

$$L(t+1) = 2 \cdot \beta \cdot |mbest - x(t)|. \quad (5)$$

Hence, the particle's position is updated according to the following equation:

$$X(t+1) = P \pm \beta \times |mbest - X(t)| \times \ln\left(\frac{1}{u}\right). \quad (6)$$

This equation is iterative equation of position in quantum-behaved particle swarm optimization (QPSO) algorithm.

The procedure for implementing the QPSO is given by Algorithm 1.

2.2. Local Search. The local search method of Solis and Wets was employed in this study. The method of Solis and Wets is a stochastic heuristic for continuous parameter spaces, which introduces a probabilistic element. Its primal purpose is the local optimization of functions that do not provide gradient information. Basically, the local optimization starts by exploring a random direction in search space and generally

```

Initialize population: random  $X_i$ 
do
  for  $i = 1$  to population size  $M$ 
    If  $f(X_i) < f(P_i)$  then  $P_i = X_i$ ;
     $P_g = \min(P_i)$ ;
    find out  $mbest$  using (4);
    for  $d = 1$  to dimension  $D$ 
       $f_{i1} = \text{rand}(0, 1)$ ;  $f_{i2} = \text{rand}(0, 1)$ ;
       $p = (f_{i1} * P_{id} + f_{i2} * P_{gd}) / (f_{i1} + f_{i2})$ ;
       $u = \text{rand}(0, 1)$ ;
      if  $(\text{rand}(0, 1) > 0.5)$ 
         $x_{id} = p - \beta * \text{abs}(mbest_d - x_{id}) * (\ln(1/u))$ ;
      else
         $x_{id} = p + \beta * \text{abs}(mbest_d - x_{id}) * (\ln(1/u))$ ;
      endif
    endfor
  endfor
Until termination criterion is met.

```

ALGORITHM 1

follows this direction with random movements as long as the objective function keeps improving. Continued improvements lead to an expansion of the random search steps, whereas continued failing narrows the search. AutoDock adopts a LGA which is a hybrid of GA and a variant of Solis and Wets local search, and the algorithm is illustrated in detail in [19, 33].

2.3. Hybrid Algorithm of QLDock. The landscape of energy function contains many local minima for flexible docking, so the performance of using only individual global search or local search may be not satisfactory. By analogy to combining GA with LS, we combined QPSO with LS and applied a hybrid algorithm to solve high flexible docking problem by proper design to enhance advantages of both algorithms as well as reducing their disadvantages to solve specific optimization problem.

In the hybrid search, QPSO can globally explore promising regions with low energy; the local search modifies the position parameters of a particle without referring to other particles and also maintains the diversity among particles. The local search may be applied to the particle according to a predefined probability. Hence, the algorithm QPSO-ls is beneficial to maintain diversity of particles and in turn prevent premature convergence. The global search method of QPSO and the Solis and Wets local search technique can efficiently compensate with each other in the new QPSO-ls hybrid algorithm. Simulation results show also that QPSO with local search is more efficient than QPSO without local search. A general structure of the QPSO-ls hybrid is as shown in Algorithm 2.

2.4. Scoring Function. Search algorithms are able to quickly generate large number of possible conformations. The purpose of a scoring function is to compare the “quality” of these possible solutions and then select best binding modes.

```

Initialize the population;
Random  $X_i(t)$ ;
Evaluate docking energy  $f(X_i)$ ;
 $P_i(t) = X_i(t)$ ;
 $P_g = \min P_i$ ;
do
  for  $i = 1$  to population size  $M$ 
    Evaluate docking energy of  $X_i(t + 1)$  using (6);
  Endfor;
  Apply the Solis and Wets local search algorithm;
  For  $i = 1$  to population size  $M$ 
    update  $P_i$  and  $P_g$ ;
  endfor;
   $t = t + 1$ ;
Until termination criterion is met;
Return the best position  $P_g$ .

```

ALGORITHM 2

The scoring functions in common use attempt to approximate the binding free energy for the ligand binding to the receptor; a low energy indicates stable system and thus a likely receptor-ligand binding interaction.

AutoDock 4.2 [35–37] adopts a semiempirical free energy scoring function to evaluate a docked conformation. The total docked energy is expressed as the sum of the intramolecular energy and the intermolecular energy of the ligand and protein. First, the intramolecular energetics are estimated for the transition from the unbound states to the conformation of the ligand and protein in the bound state. Then, the second step evaluates the intermolecular energetics of combining the ligand and protein in their bound conformation.

The force field includes six pairwise evaluations (V) and an estimate of the conformational entropy lost upon binding (ΔS_{conf}):

$$\Delta G = (V_{\text{bound}}^{L-L} - V_{\text{unbound}}^{L-L}) + (V_{\text{bound}}^{P-P} - V_{\text{unbound}}^{P-P}) + (V_{\text{bound}}^{P-L} - V_{\text{unbound}}^{P-L} + \Delta S_{\text{conf}}), \quad (7)$$

where P refers to the “protein” and L refers to the “ligand” in a protein-ligand docking calculation.

Each of the pairwise energetic terms is defined by the following energy:

$$\begin{aligned} V = & W_{\text{vdw}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{\text{hbond}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \\ & + W_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \\ & + W_{\text{sol}} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)}. \end{aligned} \quad (8)$$

The first term is a typical Lennard-Jones 12-6 dispersion repulsion interaction. The second term is a directional 12-10 hydrogen bond potential. The third term is a screened

TABLE 1: Protein-ligand complexes used for the experiments.

PDB	Ligand	Torsions	Dimension	PDB	Ligand	Torsions	Dimension
laaq	psi	20	27	lhvr	Xk2	10	17
lapt	iva	19	26	lnnb	dan	9	16
lepo	mor	17	24	lnsd	dan	9	16
lapu	iva	16	23	3tmn	val	7	14
licn	ola	15	22	2mcp	pc	4	11
4phv	vac	15	22	labf	fca	4	11
lhpv	478	14	21	ltnk	pra	4	11
2ifb	plm	14	21	ltnj	pea	3	10
lhtf	g26	13	20	lgsp	sgp	3	10
lphg	hem	12	19	ltng	amc	2	9
lcdg	mal	12	19	ltnl	tpa	2	9
lncs	sia	10	17	/	/	/	/

TABLE 2: Comparison of docking energy and RMSD values of QLDock, QDock, and AutoDock.

PDB	QLDock			QDock			AutoDock		
	E_{inter}	E_{intra}	RMSD	E_{inter}	E_{intra}	RMSD	E_{inter}	E_{intra}	RMSD
laaq	-13.18	-3.96	1.64	-7.88	-2.75	3.90	-5.98	-2.84	2.96
lapt	-11.35	-2.43	1.73	-3.72	-2.21	1.38	-7.74	-0.79	1.92
lepo	-10.67	-1.54	2.64	-6.27	-4.34	3.11	-5.97	-0.21	5.13
lapu	-10.25	-2.53	2.11	-6.79	-2.66	1.89	-3.03	-3.19	1.54
licn	-10.57	-0.82	1.40	-7.78	-0.48	1.94	-5.47	-0.26	3.49
4phv	-15.17	-2.86	1.37	-11.67	-1.65	2.88	-10.15	-1.88	2.17
lhpv	-11.25	-3.53	1.60	-5.9	-2.48	2.16	-6.03	-2.87	2.51
2ifb	-9.70	-0.55	1.06	-6.87	-0.41	1.61	-10.91	-0.49	1.40
lhtf	-9.4	-5.09	1.64	-10.25	-2.40	3.87	-4.87	-4.04	3.09
lphg	-19.23	-3.53	0.62	-16.26	-1.94	2.30	-13.5	-3.57	0.91
lcdg	-7.04	-2.82	3.67	-6.1	-2.28	3.63	-9.40	1.19	3.33
lncs	-2.54	-5.49	1.58	-4.5	-2.95	2.49	-1.04	-3.83	1.65
lhvr	-14.79	-3.64	0.95	-8.18	-2.72	1.81	-10.7	-4.28	0.78
lnnb	-3.68	-3.72	1.96	-3.66	-1.67	1.12	-3.82	-1.98	1.03
lnsd	-6.14	-2.97	1.75	-3.68	-2.82	1.16	-4.24	-2.28	2.21
3tmn	-7.69	-1.86	4.15	-2.40	-2.50	4.20	-6.52	-1.03	6.47
2mcp	-4.08	-1.00	2.38	-3.17	-1.04	1.50	-2.39	-0.74	1.92
labf	-6.91	-0.96	0.78	-5.62	-0.89	0.36	-6.85	-1.20	0.61
ltnk	-7.09	-0.09	0.63	-5.57	0.05	1.93	-7.38	-0.12	0.82
ltnj	-6.51	-0.06	1.10	-6.10	-0.09	1.31	-6.34	-0.08	0.91
lgsp	-6.99	-1.10	1.69	-3.52	-0.65	4.01	-7.35	-0.51	1.79
ltng	-6.95	0.04	0.96	-5.85	0.01	0.98	-6.62	0.02	0.99
ltnl	-6.90	-0.12	0.59	-4.14	-0.12	2.25	-6.95	-0.12	0.72

Coulombic electrostatic potential. The final term is a desolvation potential based on the volume of atoms (V) that surround a given atom and shelter it from solvent.

3. Data Preparation and Parameter Setting

3.1. Data Preparation. The prepared protein and ligand structures were saved in the PDBQT file format with the AutoDock version 4.2. All the following procedures were performed

using AutoDock Tools. First, the protein input files were obtained according to the AutoDock manual: (1) remove water molecules and metal ions not belonging to the binding site first; (2) repair residues with missing atoms; (3) add hydrogen atoms, assign partial charges, and merge nonpolar hydrogen atoms; and (4) assign solvent parameters. Then, the ligand input files also follow the AutoDock manual: (1) extract the coordinates of ligand atoms from the PDB file; (2) add hydrogen to all atoms and assign partial charges; and (3) define rigid root and torsions of the ligand. The program

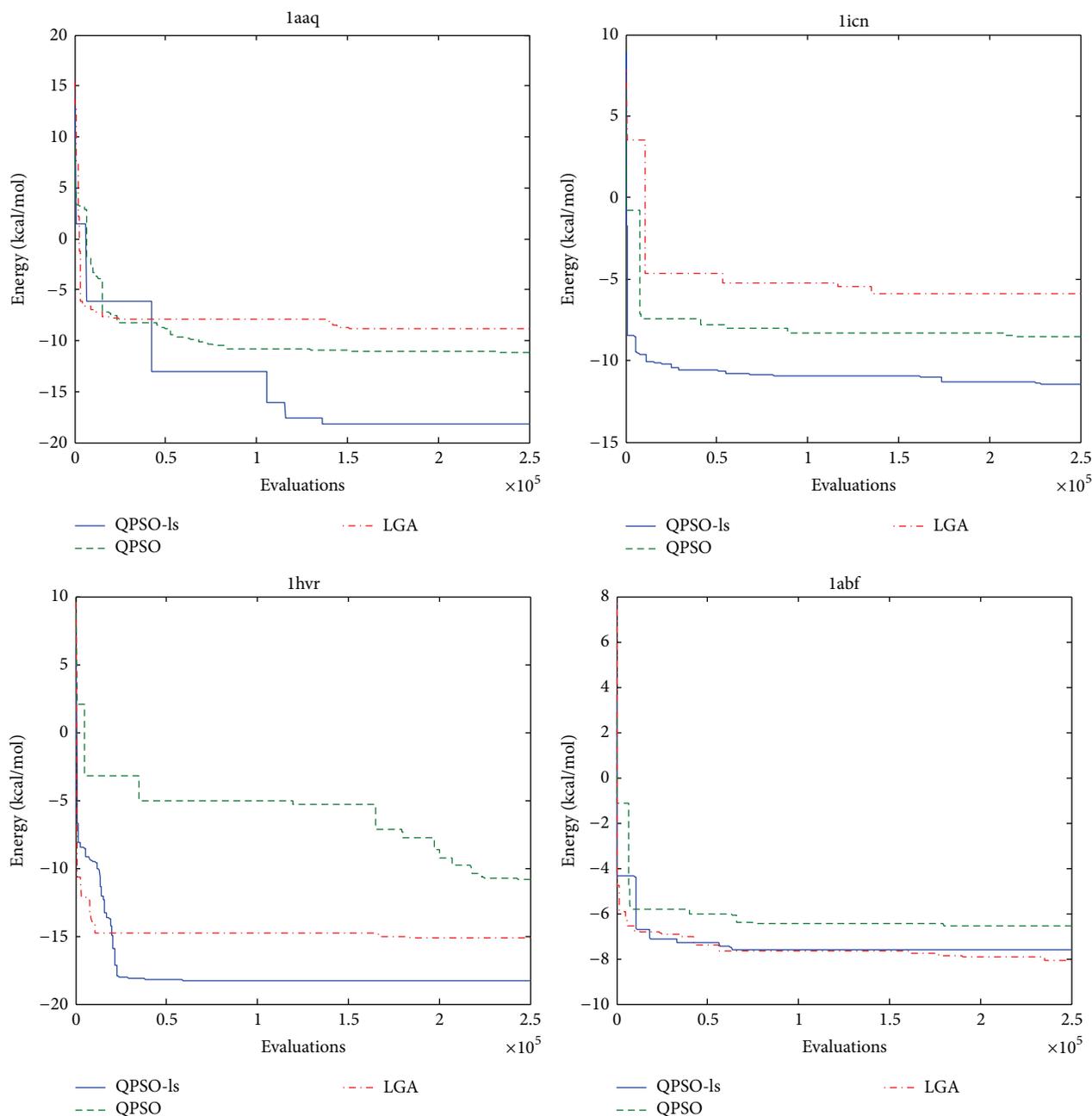


FIGURE 1: Convergence performance of the lowest docking energy for laaq, licn, lhvr, and labf during the optimization process.

AutoGrid was used to calculate energy grid maps. A default grid size of $60 \times 60 \times 60$ points with a spacing of 0.375 \AA was applied, which corresponds to a cube with an edge length of 22.5 \AA .

3.2. Experiment Parameter Setting. The 23 protein-ligand complexes that are used for the computational experiments are listed in Table 1. In AutoDock, the solution representation consists of a vector specifying the position and orientation of the ligand and the arrangement of its rotatable bonds. For describing the orientation of a ligand, a quaternion (defined

by a unit vector and a rotation around that vector) is used. Hence, a solution of docking needs to optimize $7 + n_{\text{tor}}$ parameters: 3 for the position, 4 for the orientation, and n_{tor} for the number of rotatable bonds in the ligand.

The docking experiments were performed on the same computer. All methods use the same protein coordinate and start with exactly the same initial location, conformation, and orientation of the ligand. Each docking experiment has been repeated 30 times. The initial population was set to 50 individuals; maximum number of energy evaluations was 2.5×10^5 ; maximum number of generations was 27,000. The local search was based on the Solis and Wets local search.

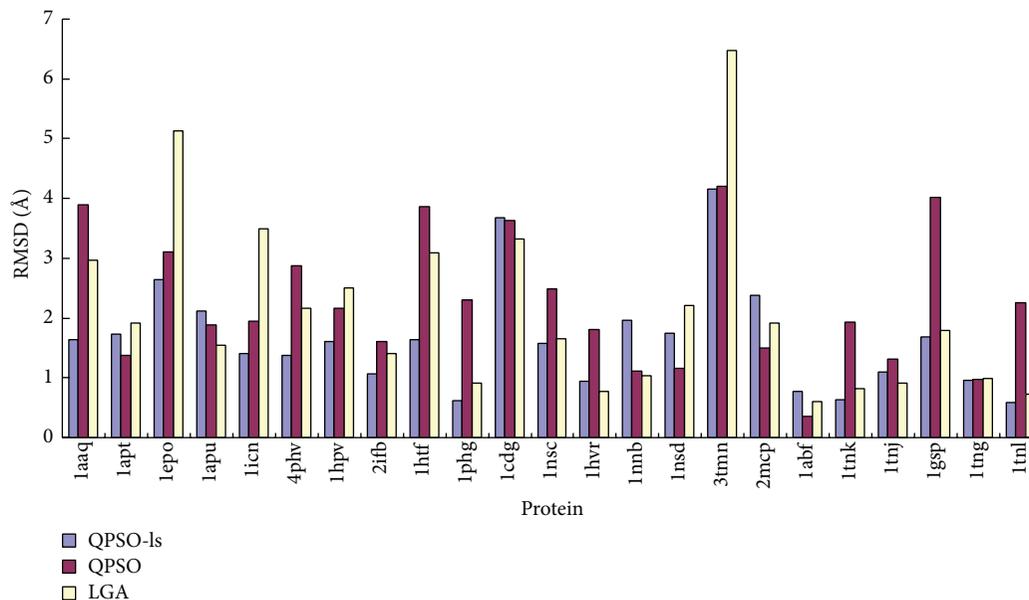


FIGURE 2: Comparison of the RMSD values with respect to three various algorithms for all instances.

The probability of performing local search on individual was 0.06; the iterations of Solis and Wets local search were 300. The other parameters provided by the default setting were the same as in AutoDock.

4. Results

To compare the performance of the docking problem, we applied the same parameter settings for all algorithms. For the evaluation of each complex, we recorded the docking energy, root mean square deviation (RMSD), running time, and the total number of scoring function evaluations in all docking runs. We rewrote the AutoDock program using C++ under Linux. The docking tests run on an Intel Pentium dual core 3.0 GHz PC with 2 G memory. The OS is Red Hat Enterprise Linux.

4.1. Evaluation of Docking Energy. To investigate the efficiency of the hybrid algorithm of QPSO-ls, 30 independent docking experiments were performed for QLDock, QDock, and AutoDock. Optimization algorithms are guided only by energy function; their search ability can be evaluated in terms of docked energy. Table 2 shows the results of the best docking energy and RMSD value of 30 docking runs using three algorithms. We chose 23 protein-ligand complexes to cover a wide range of ligands differing in their flexibility; the dimension of the search space ranges from 7 (2 rotatable bonds) to 27 (20 rotatable bonds). Since molecular docking simulation methods rely on the assumption that a lower binding energy of the predicted complex is closer to the native state, as can be seen in Table 2, QPSO-ls obtained the lowest energies for 19 out of 23 complexes. For highly flexible ligands with torsions >10, the energies obtained by QPSO-ls are much lower than those of LGA, except for the complex

2ifb with plm. The RMSD values obtained by QPSO-ls were also smaller than those of the LGAs, except for 1apu. For the complexes with less flexible ligands, the differences of both docked energy and RMSD between the three methods are not significant. The docking energy and RMSD obtained by QPSO-ls are considerably better than those returned by QPSO and LGA, for instance, especially for highly flexible docking. Also, it has been shown that the QPSO-ls performs very well compared to the common algorithm of QPSO.

In Figure 1, the docking energy of the best solution of 30 docking runs is shown for three algorithms. In the case of the structure laaq, the ligand psi has 20 rotatable bonds resulting in a dimension search space of 27 and represents a challenge for flexible docking. In QPSO-ls algorithm, the application of the local search strategy is advantageous resulting in a pose with binding energy = -17.14 kcal/mol and an RMSD value of 1.64 Å. In comparison, the value of QPSO's binding energy = -10.63 kcal/mol and RMSD = 3.90 Å and the value of LGA's binding energy = -8.82 kcal/mol and RMSD = 2.96 Å, respectively.

Moreover, QPSO-ls finds this ligand complex with lower binding energy after approximately 140,000 computing steps in comparison to 150,000 computing steps found by the LGA. It shows that QPSO-ls had better convergence performance than LGA, which also employs a local search strategy. The average execution time of QPSO-ls per independent run was 27.53 s, while that of LGA was 25.69 s. The execution time of QPSO-ls was slightly greater than that of LGA using the same number 250,000 of function evaluations.

For the complexes with less flexible ligands, the evolution of the lowest energy among the individuals through generations displays a similar behavior for three algorithms, and the algorithm converges very rapidly to lower energies. In most of the cases, the lowest energy is reached almost after 70000 computing steps and remains stable after that for

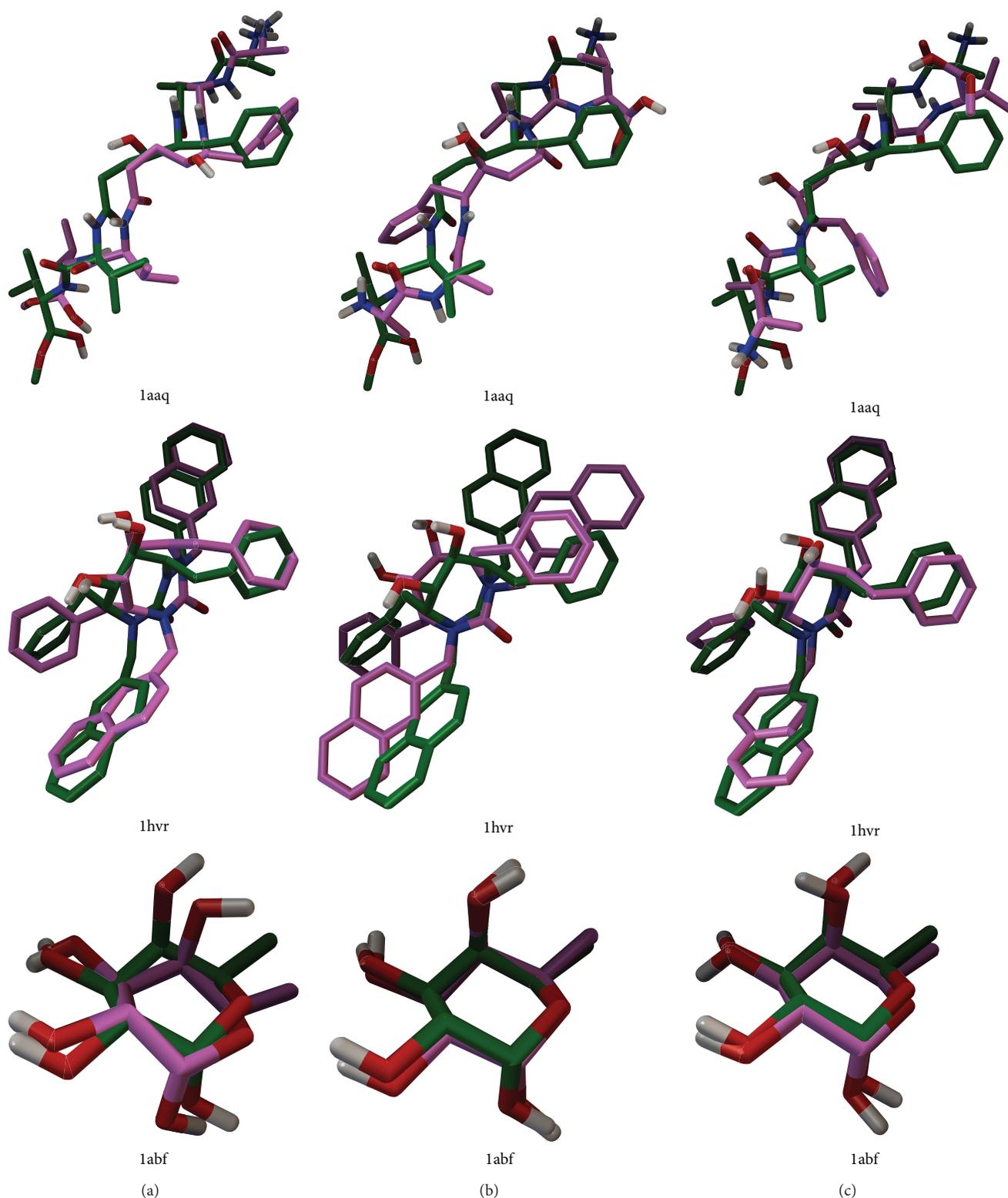


FIGURE 3: The docked ligand conformation with the lowest energies for 1aaq, 1hvr, and 1abf (from top to bottom). The native conformations (shown in green) and the predicted conformations (shown in magenta) are represented by sticks and balls. (a) QPSO-Is of QLDock. (b) QPSO of QDock. (c) LGA of AutoDock.

the labf/fca complexes. Similar results are obtained also for other less flexible ligands.

For the complexes with more than 15 rotatable bonds, the algorithm QPSO-ls is able to find conformations with much lower binding energies than QPSO and LGA. In this situation, QPSO and LGA lose the diversity in the population quickly and converge in some local optima. Moreover, the results show that the good performance of QPSO-ls with respect to QPSO is due to the concept of local search during the optimization, which can help to escape local minima. Although the local search procedure of QPSO-ls algorithm is computationally more demanding, the chance to find the optimum is higher. Consequently, this algorithm should be well suited for docking of ligands with many rotatable bonds.

4.2. Comparison of Docking Accuracy. To assess the docking accuracy of three programs, the root mean square deviation (RMSD) value with respect to the reference structure was chosen as a measure for the quality of the prediction. RMSD can be calculated using the following formula:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}, \quad (9)$$

where N is the total number of atoms in the molecule and δ is a distance between each pair of corresponding atoms. The RMSD value of 2 Å is used as a cutoff value. Poses closer to the experimentally determined structure (with RMSD lower than 2 Å) are considered sound.

Figure 2 presents the plots of the RMSD values derived by QLDock and two other docking programs versus corresponding protein-ligand complexes. A prediction of a binding mode is considered successful if the RMSD is below 2 Å. As can be seen from Figure 2, the QPSO-ls algorithm of QLDock predicts the protein-ligand complexes correctly in 74% of the cases, which is the highest percentage of all docking programs applied. The QPSO algorithm of QDOCK was successful in 52% of the cases and AutoDock in 61%. In 12 of 23 cases, QLDock provides also the lowest RMSD value of all docking programs investigated with an RMSD below 2.0 Å, as three programs employ the same docking function, which indicates the superior sampling efficiency of the QPSO-ls algorithm.

Comparing the performance of algorithm with respect to the flexibility of a ligand, we divided the investigated complexes into two classes depending on the number of rotatable bonds of the ligand: when the rotatable bond is more than five, ligand is regarded as highly flexible molecule. As listed in Table 1, QLDock predicts highly flexible ligands very well with an average RMSD value of 1.65 Å. The average RMSD values are 2.25 Å and 2.10 Å for QPSO and LGA docking programs, respectively. We can observe that better results are obtained for mean RMSD in the case of rotatable bond >10. On the contrary, this confirms again the applicability of the QPSO-ls algorithm for an efficient sampling in high-dimensional search spaces.

In Figure 3(a)laaq, we present results for the laaq/psi protein-ligand complex which show that the QLDock method

gives very useful insights into possible docking paths. The 3D visualization of the docking state as computed by QPSO and LGA for the laaq/psi suggests that the ligand could be undocked in suitable position; the aromatic ring of psi is misplaced as demonstrated in Figures 3(b)laaq and 3(c)laaq. Only QPSO-ls did find its suitable binding pose; its RMSD is 1.64 Å. In Figure 3 bottom fca was docked into the active site of labf by three algorithms, and the resultant binding poses were superposed with their binding poses in the X-ray crystal structures of fca complexes. For instance, the three results are all docked into the lhvr active site in an acceptable accuracy; the values of RMSD were 0.95 Å, 1.81 Å, and 0.78 Å for QPSO-ls, QPSO, and LGA, respectively. This demonstrates that the QPSO-ls results are much better than that of other docking programs in handling complex docking problems.

5. Conclusion

A QPSO-ls algorithm for the molecular docking problem was presented in this paper. We evaluated its performance by a series of experiments using 23 complexes. The proposed algorithm uses the energy evaluation function of the well-known tool AutoDock 4.2. To ensure the preservation of diversities of the particles and prevent the convergence procedure from prematurity, a local search strategy has been implemented. Although the number of computing steps needed to reach convergence increases, the QPSO-ls algorithm clearly outperforms the QPSO algorithm and the default LGA in terms of docked energy, convergence performance, robustness, and accuracy of the high-dimensional molecular docking problem. As demonstrated above, these results indicate also that successful identification of binding modes might be further improved by combining the results from multiple programs, and the QPSO based program seems to be excellent approach and alternative to solve the flexible docking problem in molecular docking.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors gratefully acknowledge financial support from “Qing Lan Project” of Jiangsu province and National Natural Science Foundation of China (60774079 and 61300149).

References

- [1] C. Toma, “Wavelets-computational aspects of Sterian realistic approach to uncertainty principle in high energy physics: a transient approach,” *Advances in High Energy Physics*, vol. 2013, Article ID 735452, 6 pages, 2013.
- [2] E. G. Bakhoun, “S-parameters model for data communications over 3-phase transmission lines,” *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 615–623, 2011.
- [3] L. Li, H. Peng, J. Kurths, Y. Yang, and H. J. Schellnhuber, “Chaos-order transition in foraging behavior of ants,” *Proceedings of the*

- National Academy of Sciences of the United States of America*, vol. 111, no. 23, pp. 8392–8397, 2014.
- [4] M. Li, “Approximating ideal filters by systems of fractional order,” *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 365054, 6 pages, 2012.
- [5] I. Halperin, B. Ma, H. Wolfson, and R. Nussinov, “Principles of docking: an overview of search algorithms and a guide to scoring functions,” *Proteins: Structure, Function and Genetics*, vol. 47, no. 4, pp. 409–443, 2002.
- [6] R. D. Taylor, P. J. Jewsbury, and J. W. Essex, “A review of protein-small molecule docking methods,” *Journal of Computer-Aided Molecular Design*, vol. 16, no. 3, pp. 151–166, 2002.
- [7] R. Thomsen and M. H. Christensen, “MolDock: a new technique for high-accuracy molecular docking,” *Journal of Medicinal Chemistry*, vol. 49, no. 11, pp. 3315–3321, 2006.
- [8] H. Gohlke, M. Hendlich, and G. Klebe, “Knowledge-based scoring function to predict protein-ligand interactions,” *Journal of Molecular Biology*, vol. 295, no. 2, pp. 337–356, 2000.
- [9] E. Perola, K. Xu, T. M. Kollmeyer, S. H. Kaufmann, F. G. Prendergast, and Y. P. Pang, “Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads,” *Journal of Medicinal Chemistry*, vol. 43, no. 3, pp. 401–408, 2000.
- [10] L. Guo, Z. Yan, X. Zheng, L. Hu, Y. Yang, and J. A. Wang, “A comparison of various optimization algorithms of protein–ligand docking programs by fitness accuracy,” *Journal of Molecular Modeling*, vol. 20, no. 7, pp. 2251–2261, 2014.
- [11] S. Makino and I. D. Kuntz, “Automated flexible ligand docking method and its application for database search,” *Journal of Computational Chemistry*, vol. 18, no. 14, pp. 1812–1825, 1997.
- [12] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, “Development and validation of a genetic algorithm for flexible docking,” *Journal of Molecular Biology*, vol. 267, no. 3, pp. 727–748, 1997.
- [13] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, “A fast flexible docking method using an incremental construction algorithm,” *Journal of Molecular Biology*, vol. 261, no. 3, pp. 470–489, 1996.
- [14] G. M. Morris, D. S. Goodsell, R. S. Halliday et al., “Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function,” *Journal of Computational Chemistry*, vol. 19, no. 14, pp. 1639–1662, 1998.
- [15] D. S. Goodsell and A. J. Olson, “Automated docking of substrates to proteins by simulated annealing,” *Proteins: Structure, Function and Genetics*, vol. 8, no. 3, pp. 195–202, 1990.
- [16] G. M. Morris, R. Huey, W. Lindstrom et al., “AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility,” *Journal of Computational Chemistry*, vol. 30, pp. 2785–2791, 2009.
- [17] S. Janson and D. Merkle, “A new multi-objective particle swarm optimization algorithm using clustering applied to automated docking,” in *Hybrid Metaheuristics*, vol. 3636 of *Lecture Notes in Computer Science*, pp. 128–141, 2005.
- [18] Y. Liu, L. Zhao, W. Li, D. Zhao, M. Song, and Y. Yang, “FIPS-Dock: a new molecular docking technique driven by fully informed swarm optimization algorithm,” *Journal of Computational Chemistry*, vol. 34, no. 1, pp. 67–75, 2013.
- [19] H.-M. Chen, B.-F. Liu, H.-L. Huang, S.-F. Hwang, and S.-Y. Ho, “SODOCK: swarm optimization for highly flexible protein–ligand docking,” *Journal of Computational Chemistry*, vol. 28, no. 2, pp. 612–623, 2007.
- [20] V. Namasivayam and R. Günther, “PSO@AUTODOCK: a fast flexible molecular docking program based on swarm intelligence,” *Chemical Biology and Drug Design*, vol. 70, no. 6, pp. 475–484, 2007.
- [21] O. Korb, T. Stützle, and T. E. Exner, “PLANTS: application of ant colony optimization to structure-based drug design,” in *Ant Colony Optimization and Swarm Intelligence: Proceedings of the 5th International Workshop, ANTS 2006, Brussels, Belgium, September 4–7, 2006*, vol. 4150 of *Lecture Notes in Computer Science*, pp. 247–258, Springer, Berlin, Germany, 2006.
- [22] J.-L. Chen, C.-W. Tsai, M.-C. Chiang, and C.-S. Yang, “A high performance cloud-based protein-ligand docking prediction algorithm,” *BioMed Research International*, vol. 2013, Article ID 909717, 8 pages, 2013.
- [23] A. N. Lima, E. A. Philot, D. Perahia, A. S. K. Braz, and L. P. B. Scott, “GANM: a protein-ligand docking approach based on genetic algorithm and normal modes,” *Applied Mathematics and Computation*, vol. 219, no. 2, pp. 511–520, 2012.
- [24] M. Clerc and J. Kennedy, “The particle swarm-explosion, stability, and convergence in a multidimensional complex space,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.
- [25] J. Sun, B. Feng, and W. Xu, “Particle swarm optimization with particles having quantum behavior,” in *Proceedings of the Congress on Evolutionary Computation (CEC’04)*, vol. 1, pp. 325–331, June 2004.
- [26] J. Kennedy, “Some issues and practices for particle swarms,” in *Proceedings of the IEEE Swarm Intelligence Symposium (SIS’07)*, pp. 162–169, Honolulu, Hawaii, USA, April 2007.
- [27] J. Sun, X. J. Wu, W. Fang, Y. R. Ding, H. X. Long, and W. B. Xu, “Multiple sequence alignment using the hidden Markov model trained by an improved quantum-behaved particle swarm optimization,” *Information Sciences*, vol. 182, pp. 93–114, 2012.
- [28] S. N. Omkar, R. Khandelwal, T. V. S. Ananth, G. N. Naik, and S. Gopalakrishnan, “Quantum behaved Particle Swarm Optimization (QPSO) for multi-objective design optimization of composite structures,” *Expert Systems with Applications*, vol. 36, no. 8, pp. 11312–11322, 2009.
- [29] S. Y. Li, R. G. Wang, W. W. Hu, and J. Q. Sun, “A new QPSO based BP neural network for face detection,” in *Fuzzy Information and Engineering*, B.-Y. Cao, Ed., vol. 40 of *Advances in Soft Computing*, pp. 355–363, Springer, Berlin, Germany, 2007.
- [30] Y. Cai, J. Sun, J. Wang et al., “Optimizing the codon usage of synthetic gene with QPSO algorithm,” *Journal of Theoretical Biology*, vol. 254, no. 1, pp. 123–127, 2008.
- [31] Z. Huang, Y. Wang, C. Yang, and C. Wu, “A new improved quantum-behaved particle swarm optimization model,” in *Proceedings of the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA’09)*, pp. 1560–1564, May 2009.
- [32] M. Pant, R. Thangaraj, and V. P. Singh Sobol, “Sobol mutated quantum particle swarm optimization,” *International Journal of Recent Trends in Engineering*, vol. 1, no. 1, pp. 95–99, 2009.
- [33] F. J. Solis and R. J.-B. Wets, “Minimization by random search techniques,” *Mathematics of Operations Research*, vol. 6, no. 1, pp. 19–30, 1981.
- [34] J. Sun, B. Feng, and W. Xu, “A global search strategy of quantum-behaved particle swarm optimization,” in *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems*, vol. 1, pp. 111–116, December 2004.
- [35] G. M. Morris, R. Huey, W. Lindstrom et al., “AutoDock4 and AutoDockTools4: automated docking with selective receptor

- flexibility,” *Journal of Computational Chemistry*, vol. 30, no. 16, pp. 2785–2791, 2009.
- [36] S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell, and A. J. Olson, “Virtual screening with AutoDock: theory and practice,” *Expert Opinion on Drug Discovery*, vol. 5, no. 6, pp. 597–607, 2010.
- [37] S. Forli and A. J. Olson, “A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking,” *Journal of Medicinal Chemistry*, vol. 55, no. 2, pp. 623–638, 2012.

Research Article

Obtaining Cross Modal Similarity Metric with Deep Neural Architecture

Ruifan Li,^{1,2} Fangxiang Feng,¹ Xiaojie Wang,^{1,2} Peng Lu,¹ and Bohan Li¹

¹School of Computers, Beijing University of Posts and Telecommunications, Beijing 100876, China

²Engineering Research Center of Information Networks, Ministry of Education, Beijing 100876, China

Correspondence should be addressed to Ruifan Li; rfl@bupt.edu.cn

Received 14 September 2014; Accepted 24 December 2014

Academic Editor: Florin Pop

Copyright © 2015 Ruifan Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Analyzing complex system with multimodal data, such as image and text, has recently received tremendous attention. Modeling the relationship between different modalities is the key to address this problem. Motivated by recent successful applications of deep neural learning in unimodal data, in this paper, we propose a computational deep neural architecture, bimodal deep architecture (BDA) for measuring the similarity between different modalities. Our proposed BDA architecture has three closely related consecutive components. For image and text modalities, the first component can be constructed using some popular feature extraction methods in their individual modalities. The second component has two types of stacked restricted Boltzmann machines (RBMs). Specifically, for image modality a binary-binary RBM is stacked over a Gaussian-binary RBM; for text modality a binary-binary RBM is stacked over a replicated softmax RBM. In the third component, we come up with a variant autoencoder with a predefined loss function for discriminatively learning the regularity between different modalities. We show experimentally the effectiveness of our approach to the task of classifying image tags on public available datasets.

1. Introduction

Recently, there is a growing demand for analyzing the complex systems with great number of variables [1, 2], such as multimodal data with image and text, due to the availability of computational power and massive storage. For one thing, information often naturally comes in multiple modalities of a large number of variables. For example, a travel photo shared on the website is usually tagged with some meaningful words. For another, analyzing those heterogeneous data of great number of variables from multiple sources could benefit different modalities. For instance, speaker's articulation and muscle movement can often aid in disambiguating between speeches with similar phones.

During the past few years, motivated by the biological propagation phenomena in distributed structure of human brain, deep neural learning has received considerable attention from the year of 2006. These deep neural learning methods are proposed to learn hierarchical and effective representations to facilitate various tasks with respect to recognizing and analyzing in complex artificial system. Even

with only a very short development, deep neural learning has achieved great success in some tasks of modeling the single modal data, such as speech recognition systems [3–6] and computer vision systems [7–12], to name a few.

Motivated by the progress in deep neural learning, in this paper, we endeavor to construct a computational deep architecture for measuring the similarity between modalities in complex multimodal system with a large number of variables. Our proposed framework, bimodal deep architecture (BDA), has three closely consecutive components. For images and text modalities, the first component can be constructed by some popular feature extraction methods in each individual modality. The second component has two types of stacked restricted Boltzmann machines (RBMs). Specifically, for image modality a Bernoulli-Bernoulli RBM (BB-RBM) is stacked over an RBM; for text modality a BB-RBM is stacked over a replicated softmax RBM (RS-RBM). In the third component, we come up with a variant autoencoder with a predefined loss function for discriminatively learning the regularity hidden within modalities.

It is worthwhile to highlight several aspects of the BDA proposed in this paper.

- (i) In the first component of the BDA, for image modality, three methods are utilized in our setting. However, we could explore more feature extraction methods.
- (ii) In the second component of the BDA, we stack two RBMs for each modality. In theory, we could stack more RBMs to make a more effective representation.
- (iii) In the third component of the BDA, motivated by the deep neural architecture, we come up with a loss function to keep small distance for semantically similar bimodal data and to generate large distance for semantically dissimilar ones.
- (iv) The work in this paper primarily focuses on the image and text bimodal data. However, the BDA presented here can be naturally extended to other different bimodal data.

The remainder of this paper is organized as follows. Section 2 describes and discusses the related work. Section 3 presents our deep architecture and its learning algorithm. Section 4 introduces the datasets, describes the other two methods for comparisons, and reports the experimental results. Finally Section 5 draws the conclusion and discusses the future work.

2. Related Work

There have been several approaches to learning from cross modal data with many variables. In particular, Blei and Jordan [13] extend latent Dirichlet allocation (LDA) by mining the topic-level relationship between images and text annotations. Xing et al. [14] build a joint model to integrate images and text, which can be viewed as an undirected extension to LDA. Jia et al. [15] proposes a combination of the undirected Markov random field and the directed LDA. However, this type of models with a single hidden layer is unable to obtain efficient representations because of the complexity of images and text.

Recently, motivated by deep neural learning, Chopra et al. [16] propose to learn a function such that the norm in the embedded space approximates the semantic distance. This learned network, however, keeps only half of the structure and only fits for unimodal data. Ngiam et al. [17] use a deep autoencoder for vision and speech fusion. Srivastava and Salakhutdinov [18] develop a deep Boltzmann machine as a generative model for images and text. However, these two works focus on cross modal retrieval but not the similarity metric.

Another line of research focuses on bimodal semantic hashing, which tries to represent data as binary codes. Subsequently, Hamming metric is applied for the learned codes as the measure of similarity. McFee and Lanckriet [19] propose a framework based on multimodal kernel learning approaches. However, this framework is limited to linear projections. Most similar to our work, Masci et al. [20] propose a framework based on the neural autoencoder

to merge multiple modalities into a single representation space. However, this framework can only be used for labeled bimodal data.

3. BDA for Cross Modal Similarity

The main idea of our deep framework is to construct hierarchical representations of bimodal data. This framework, as shown in Figure 1, has three closely consecutive components. In the first component, the low-level representations by classical single-modal methods for these two types of data are obtained, respectively. For images, the features are extracted by four descriptors in MPEG-7, and gist features are combined to form the low-level representations. For tag words, the typical bag-of-words (BOW) model is used for low-level representations.

In the second component, the low-level representations usually with different dimensions for image and tag words are distilled to form the mid-level representations using two stacked restricted Boltzmann machines (RBMs), respectively. The first layer RBMs, a Gaussian RBM for low-level representation of the images and a replicated softmax for those of text, are adopted mainly for normalizing these bimodal data with the same output units. The second layer RBMs, two binary RBMs, are used for expecting more abstract representations.

In the third component, we propose a variant of autoencoder for learning the high-level semantic similar/dissimilar representations of these bimodal data. The details of this network are described in Section 3.3. All numbers in the boxes are the number of neurons used in each layer. The detailed description of each component is presented in the following sections.

In the training stage, a collection of pairs of image and text is presented to the system. As a result, by our learning algorithm the system could learn the neural connection weights. In the test stage, a new pair of bimodal data is shown to the learned system such that we can obtain the similarity/dissimilarity between the pair of unseen data.

3.1. Obtaining Basic Representations. Different unimodal data, such as images or text, usually have different methods to extract the representative features. We use these extracted features as our basic representations. For image modality some popular methods, such as MPEG-7 and gist descriptors, can be used. Gist represents the dominant spatial structure of a scene by a set of perceptual dimensions, including naturalness, openness, roughness, expansion, and ruggedness. These perceptual dimensions can be estimated using spectral and coarsely localized information.

One part of MPEG-7 is a standard for visual descriptors. We use four different visual descriptors defined in MPEG-7 for image representations: color layout (CL), color structure (CS), edge histogram (EH), and scalable color (SC). CL is based on spatial distribution of colors and is obtained by applying the DCT transformation. CS is based on color distribution and local spatial structure of the color. EH is based on spatial distribution of edges. SC is based on the

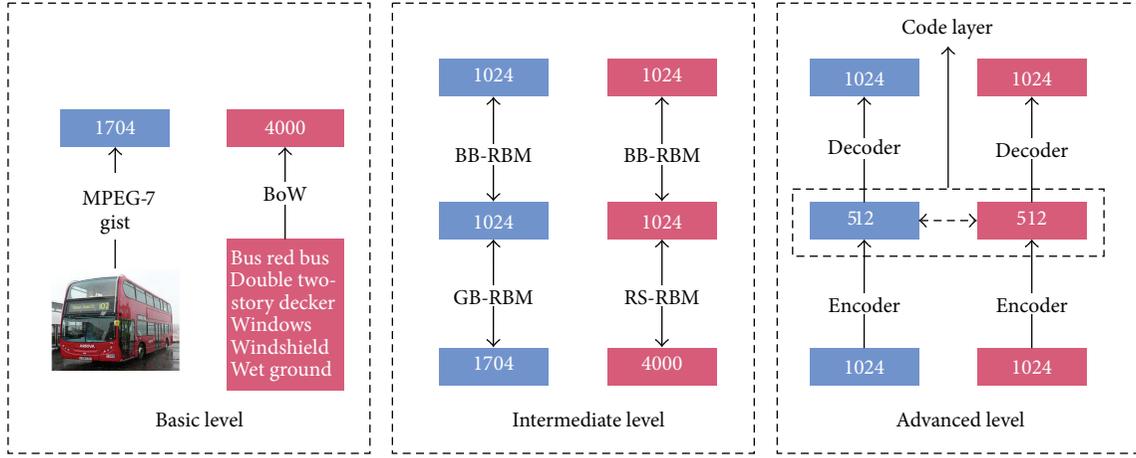


FIGURE 1: A deep framework is used for measuring the similarity of cross modal data such as images and text. From left to right, first, the classical methods for each modality could be used to extract basic modality-specific features. For example, we use MPEG-7, Gist, and some known features descriptors for images; we use bag-of-words model for tags. Second, for each modality two RBMs are stacked for extracting intermediate modality-specific features. For images, we stack a binary RBM over a Gaussian RBM; for text, we stack a binary RBM over a replicated softmax. Third, an autoencoder with similar constraint is used for extracting similar presentations. The number in each box is the neurons adopted in this layer.

color histogram in HSV color space encoded by a Haar transformation.

For text modality, we use the classical bag-of-words model for its basic representations. A dictionary of K high-frequency words is built from all the words in the database. Then, each tag can be represented as a multinomial variable. Conveniently, 1-of- K coding scheme is adopted. Thus, the tags of an image could be represented as a vector with K one/zero elements, in which each element stands for whether the tag is in the dictionary or not. For tags in the dictionary, they are encoded as 1s, and vice versa.

3.2. Learning Intermediate Representations

3.2.1. Modeling Binary Data. For binary data, as in the second layer of the second component in our framework, we use RBMs for modeling them. An RBM [21] is an undirected graphical model with stochastic binary units in the visible layer and the hidden layer but without connections between units within these two layers. Given that there are n visible units \mathbf{v} and m hidden units \mathbf{h} , and each unit is distributed by Bernoulli distribution with logistic activation function $\sigma(x) = 1/(1 + \exp(-x))$, we then define a joint probabilistic distribution p over the visible units \mathbf{v} and hidden units \mathbf{h}

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \exp(-E(\mathbf{v}, \mathbf{h})), \quad (1)$$

in which Z is the normalization constant and $E(\mathbf{v}, \mathbf{h})$ is the energy function defined by the configurations of all the units as

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^n \sum_{j=1}^m w_{ij} v_i h_j - \sum_{i=1}^n b_i v_i - \sum_{j=1}^m c_j h_j. \quad (2)$$

Here, w_{ij} is the connection weight between the i th visible unit with state v_i and the j th hidden unit h_j with state v_j . And the

two parameters, b_i and c_j , are biases for the i th visible unit and the j th hidden unit, respectively.

3.2.2. Modeling Real-Valued Data. We model the image features with real-valued data, as in the first layer of the second component in our framework, using Gaussian RBM. It is an extension of the binary RBM replacing the Bernoulli distribution with Gaussian distribution for the visible data [22]. The energy function of different configurations of visible units and hidden ones is

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^n \sum_{j=1}^m w_{ij} \frac{v_i}{\sigma_i} h_j + \sum_{i=1}^n \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=1}^m c_j h_j, \quad (3)$$

in which σ_i is the variance of the Gaussian distribution of i th visible unit. And we usually set the variances σ_i equal to one for all visible units. Except the variance σ_i , the other parameters in the above equation have the same meanings as those in (2).

3.2.3. Modeling Count Data. For the text features with count data, as in the first layer of the second component in our framework, we use replicated softmax model [23] for modeling these sparse vectors. The energy function of the date configurations is defined as follows:

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^n \sum_{j=1}^m w_{ij} v_i h_j - \sum_{i=1}^n b_i v_i - M \sum_{j=1}^m c_j h_j, \quad (4)$$

where M is the total number of words in a document. And the other parameters in the above equation have the same meanings as those in (2). Note that this replicated softmax model can be interpreted as an RBM model that uses a single visible multinomial unit with support $1, \dots, K$ which is sampled M times.

In the second component of our framework, for each modality we stack two RBMs to learn the intermediate representations. These two-layer stacked RBMs can be trained by the greedy layer-wise training method [24, 25]. And we can efficiently learn all the above three kinds of RBMs by using the contrastive divergence (CD) approximation [26].

3.3. Learning Advanced Representations. In the third component of our framework, we propose a special type of autoencoder for bimodal representations to learn the similarity. As shown in the rightmost of Figure 1, this autoencoder consists of two subnets, in which each subnet is fully connected perceptron. And these two networks are connected by some predefined similarity measure on the code layer. By designing a proper loss function from energy-based learning paradigm [27], we could learn the similar representations of these two bimodalities.

Formally, we denote the mapping from the inputs of two subnets to the code layers as $f(p; \mathcal{W}_f)$ and $g(q; \mathcal{W}_g)$, in which $f(\cdot)$ is for the image modality and $g(\cdot)$ is for the text modality. Here, p and q denote one intermediate representation for image and one intermediate representation for a given tag. And \mathcal{W} denotes the weight parameters in these two subnets. The subscripts f and g of the weight \mathcal{W} denote the corresponding two modalities, respectively. We define the compatibility measure between i th pair of image p_i and its given tags q_i as

$$C(p_i, q_i; \mathcal{W}_f, \mathcal{W}_g) = \|f(p_i; \mathcal{W}_f) - g(q_i; \mathcal{W}_g)\|_2, \quad (5)$$

where $\|\cdot\|_2$ is the \mathcal{L}_2 norm.

To learn the similar representations of these two modalities for one object, we come up with a loss function given input p_i, q_i , and a binary indicator δ with respect to the inputs. In our settings, the indicator variable δ equals one if the tag words q_i correspond to the image q_i , and the indicator variable δ equals zero otherwise. To simplify the notation we group the network parameters \mathcal{W}_f and \mathcal{W}_g into one parameter Θ . As a result, we define the loss function ℓ on any pair of inputs p_i and q_i as

$$\begin{aligned} \ell(p_i, q_i, \delta; \Theta) &= \alpha(\ell_f(p_i; \mathcal{W}_f) + \ell_g(q_i; \mathcal{W}_g)) \\ &+ (1 - \alpha)\ell_c(p_i, q_i, \delta; \Theta), \end{aligned} \quad (6)$$

where

$$\ell_f(p_i; \mathcal{W}_f) = \|p_i - \hat{p}_i\|_2^2, \quad (7a)$$

$$\ell_g(q_i; \mathcal{W}_g) = \|q_i - \hat{q}_i\|_2^2, \quad (7b)$$

$$\ell_c(p_i, q_i, \delta; \Theta) = \delta C^2 + (1 - \delta) \exp(-\lambda C). \quad (7c)$$

Here, the total loss comprises of three parts. The first two losses ℓ_f and ℓ_g are caused by data reconstruction errors for the given inputs (an image and its tags) of two subnets, while the third loss $\ell_c(p_i, q_i, \delta; \Theta)$, called the contrastive loss, is incurred by whether the image and tags are compatibility or not in two different situations indicated by δ . And $\|\cdot\|_2$

in (7a) and (7b) denotes the \mathcal{L}_2 norm. The \hat{p}_i in (7a) is the construction representation of the input image p_i , and the \hat{q}_i in (7b) is the construction representation of the input text q_i . λ in (7c) is a constant, which depends on the upper bound of $C(p_i, q_i; \Theta)$ on all training data. The constant α ($0 < \alpha < 1$) in the total loss function (6) is a parameter used to trade off between two groups of losses, the reconstruction losses, and the compatibility loss.

By the standard backpropagation algorithm [28] we can learn the connections weights among neurons in this autoencoder. During learning, the two subnets are coupled at their code layer with the similarity measure. After being learned, the two subnetworks will have different parameters even if they have the same architecture. As a result, the codes for new inputs can be obtained using the learned network parameters.

To summarize, by the above three consecutive components we can learn the similarity metric for bimodal data.

4. Experiments and Results

We evaluate our proposed method for the task of image annotation selection compared with multilayer perceptron (MLP) with two hidden layers and canonical correlation analysis (CCA) with RBMs as a benchmark method on two publicly available datasets.

In the following sections, we will describe the two datasets used, our experimental settings, and the evaluation criteria. Moreover, we report and discuss our experimental results.

4.1. Datasets and Preprocessing. The two datasets used in our experiments are the Small ESP Game dataset [29] and the multimodal learning challenge 2013 (MLC-2013) dataset [10]. The ESP dataset, briefly known as ESP, was created by Luis von Ahn using crowdsourcing efforts of ESP online players from different locations. Specifically, the ESP consists of 100,000 labeled images with the corresponding labeled tags. The images in the ESP have a variety of sizes; the tags are influenced by the game format. Note that each image in the ESP only has the correct tag words. Some examples from the ESP are shown in Figure 2. The ESP dataset is available at <http://www.cs.cmu.edu/~biglou/resources/>.

The MLC-2013 dataset, briefly known as MLC, was created by Ian J. Goodfellow for the workshop on representation learning at International Conference on Machine Learning, ICML 2013. Specifically, the MLC consists of 1,000 manually labeled images, which were obtained by Google image search queries for some of the most commonly used words in the ESP. And for each image, two labels are given; one fits better than the other. The labels were intended to resemble those in the EPS. For example, they include incorrect spellings that were common in the EPS. Some examples from the MLC are shown in Figure 3. The MLC dataset is available at <http://www-etud.iro.umontreal.ca/~goodfeli/mlc2013.html>.

In all our experiments, we use the ESP as the training set and the MLC as the test set. In addition, we note that each image in the ESP involves a unique group of tags; we must firstly construct the counterpart group of tags for each

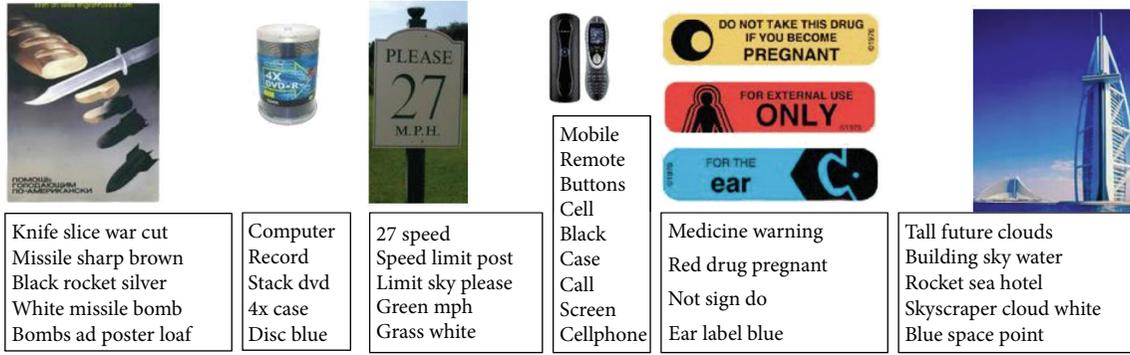


FIGURE 2: Examples from the ESP dataset. The top is images and the bottom is their corresponding group of tag words. See text for more details.



FIGURE 3: Two examples from the MLC dataset. Each one shows the image and its two tags. The words in sky blue box are from its original tag, and the words in carnation box are from the generated one.

image in the ESP. Therefore, in order to train our system, we automatically generate an incorrect counterpart for word tags of each image in the ESP. This can be achieved by randomly choosing one from all the correct tag words of the rest of the images, while ensuring that each of the tag words occurs only one time. That is our preprocessing procedure.

4.2. Settings of Our BDA Method. In this section, we describe the settings in our experiments, including the details of feature extraction methods in the first component and the neurons configurations in the second and third components.

In the first component, for image modality, three popular methods are adopted. The first group of features is obtained by the following steps: (1) preprocess images using local contrast normalization [30]; (2) train a K -means dictionary on patches of ESP; (3) extract soft threshold features [31] from all image patches and then downsample using local max pooling; (4) train a K -means dictionary on patches of these K -means feature maps; (5) extract a second layer of soft threshold features, using global max pooling to make a “bag of visual words” (BOVW) feature vector for each image.

The second group of features is obtained by MPEG-7 visual descriptors. We use 192 coefficients of DCT transformation for Cl, 256 coefficients form for CS, fixed 80 coefficients for EH, and 256 coefficients form for CS. The software module based on the MPEG-7 Reference Software (available at <http://www.cs.bilkent.edu.tr/~bilmdg/bilvideo-7/Software.html>) permits obtaining all four different descriptors. Thus, we obtain the features of MPEG-7 descriptors with the size of 784.

The third group of features is obtained by gist descriptor. The package used in our paper is available at <http://people.csail.mit.edu/torralba/code/spatialenvelope/> for downloading. Through all the three groups of features, each image is represented as a vector of 1704 dimensions. More specifically, this image representation is composed of the leading 408 dimensions from BOVW, the middle 784 dimensions from MPEG-7, and the last 512 dimensions from gist features.

For text representation we use bag-of-words model. In our experiments, a dictionary of 4000 high-frequency words is built from all the tag words of ESP. Thus, each tag is represented as a vector with 4000 one/zero elements.

In the second component, we use the neurons configuration of 1704-1024-1024 for image modality and of 4000-1024-1024 for text modality. That is, Gaussian-Bernoulli RBM has the 1704 visible neurons and 1024 hidden neurons. The replicated softmax has 4000 visible neurons and 1024 hidden neurons. And the two Bernoulli RBMs in the above layers in the intermediate representations have 1024 visible neurons and 1024 hidden neurons.

In the third component, we use the neurons configuration of 1024-512 both for image modality and text modality. That

is, the autoencoders in the advanced representations for both modalities have 1024 visible neurons and 512 hidden neurons.

As for the parameters, we illustrate the method for setting the parameter α in the following section. We have mentioned that the parameter λ depends on the upper bound of $C(p_i, q_j; \Theta)$ on all training data in the previous section. Practically, we repeated several experiments with different values for choosing this parameter. Then, the value of 0.1 is set for this parameter.

4.3. Settings of Benchmark Methods. For comparison of two benchmark methods, MLP with two hidden layers and CCA with RBMs are used in our experiments. Next, we describe the details of these two methods.

The MLP is a popular supervised learning model in artificial neural networks system [32]. In our experiments, we use the neural structure proposed in [10]. This network has the following structure. The input layer of this network is the basic representation for the image modality. Two hidden layers with the same number of neurons are utilized. The top layer is the basic representation for the text modality. To be specific, we set the input layer of the MLP system to have 1704 neurons and the output layer to have 4000 neurons. And each of two hidden layers has 1024 neurons. And all the neurons in this MLP have logistic activation function. This MLP system is graphically depicted in Figure 4. To test an unseen image, we first use the preprocessing modules outside the dashed box to obtain the bimodal representations. Then, using the learned system, we can obtain the output as the basis for judging the better fitting tag words. The idea behind this design is to model the nonlinear mapping directly from the image modality to the text modality. Note that, unlike the traditional MLP, the MLP used in our experiments has two hidden layers. And this MLP with multiple hidden layers can be seen as a kind of deep neural architecture, which would benefit the nonlinearly semantic mapping from one modality to the other.

The CCA [33, 34] is a multivariate statistical model that facilitates the study of linear interrelationships between two groups of variables. More specifically, given a pair of data p and q , the CCA attempts to discover matrices U and V such that Up and Vq have maximum correlations. The first d canonical components of U and V could be used for projecting an unseen pair of data into a d -dimensional space. The CCA-based system for modeling the relationship between multimodal data achieves the state-of-art performance in analyzing numerous cross modal tasks. To name a few, recent work [17, 35, 36] shows the strength of CCA system. Therefore, we adopt a CCA system as one of our benchmark systems.

In the CCA system, the two groups of variables are modal representations. One group of input of the CCA system is the basic representation for the image modality. The other is the replicated softmax RBM representation based on the BOW model. Then, the CCA is applied for these two quantities. To be specific, we set the image representation to have 1,704 neurons. And a replicated softmax RBM is used with an input of 4,000 neurons and output of 1,024 neurons. The canonical

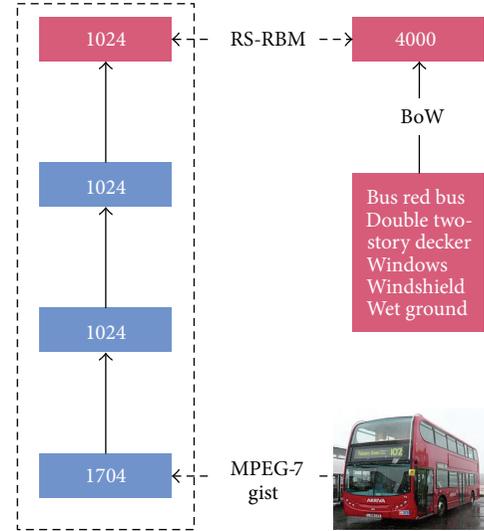


FIGURE 4: The MLP system used in our experiments. In the lower right corner, an image is represented using MPEG-7 and gist descriptors forming a vector with 1,704 neurons. And in the upper right corner, the corresponding tag words are first represented using the BOW model. Furthermore, a replicated softmax RBM with 4,000 visible neurons and hidden 1,024 neurons is adopted to learn a text representation. Finally, in the dashed box, from the bottom to the top, an MLP with two hidden layers is designed to learn the mapping from image modality to text modality.

components d in CCA is set as 1,024. This CCA system is graphically depicted in Figure 5. To test an unseen image, we first use the preprocessing modules outside the dashed box to obtain the bimodal representations. Then, using the learned mapping, we can obtain the output as the basis for judging the better fitting tag words. The CCA code is available at <http://davidroihardoon.com/Professional/Code.html>.

4.4. Evaluation Criteria. The performance of predictions of a classifier is evaluated based on the accuracy. Here, the accuracy is, for simplicity, defined as the area under the receiving operating curve (ROC). The adoption of this accuracy is motivated by two facts. One is its successful application on evaluating binary classifiers [37]. The other is the encouragement by the builder of the MLC dataset [10]. Next, we will describe the definition and its computation of the accuracy used for evaluating these models.

Often used for evaluating binary classifiers, the ROC is a plot of false positive rate $U_1(t)$ at a decision threshold t on the x -axis against true positive rate $U_0(t)$ at t on the y -axis. And the subscript for U denotes the corresponding class. Given the score output of a classifier, an instance is classified as the positive class if the score is greater than the decision threshold t or as the negative otherwise. Note that these two quantities $U_1(t)$ and $U_0(t)$ will be monotonically nondecreasing with increasing t . We then define the accuracy as follows:

$$\text{Accuracy} \triangleq \int_0^1 U_0(s) dU_1(s) = \int_{-\infty}^{+\infty} U_0(s) u_1(s) ds, \quad (8)$$

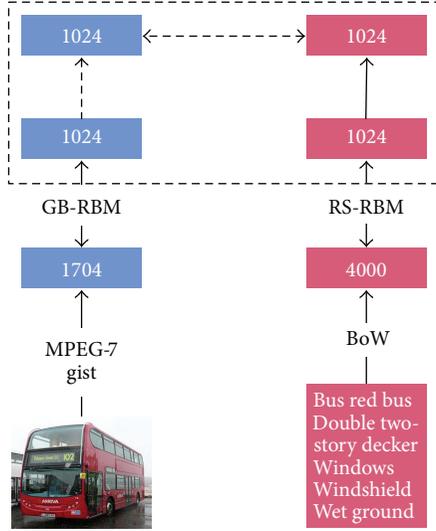


FIGURE 5: The CCA system used in our experiments. From the lower left corner, the image modality is represented using MPEG-7 and gist descriptors forming a vector with size 1,704. Then, a Gaussian RBM with 1,704 visible neurons and 1,024 hidden neurons is used to learn the image representation. From the lower right corner, the text modality is represented using BoW model forming a vector with size 4000. Then, a replicated softmax RBM with 4,000 visible neurons and hidden 1,024 neurons is adopted to learn a text representation. Finally, a CCA model with 1,024 twin inputs and 1,024 twin outputs is built using these bimodal representations.

in which $u_1(\cdot)$ is the density function for class 1. Thus, $U_1(t) = \int_{-\infty}^t u_1(s) ds = P(s \leq t | 1)$ is the false positive rate at threshold t . When dealing with empirical distributions the integral is replaced by the sum. In [37], an efficient algorithm is proposed for calculating the accuracy.

In order to fulfill the computation, the three models in our experiments should be able to produce a continuous valued output that can be used for ranking its predictions. Therefore, we define the continuous valued output $P(p_i)$ for the positive class as the ratio of the squared compatibility metric of an image p_i with its true tag words q_i to the sum of the squared compatibility metrics of an image p_i with its two tag words q_i and \tilde{q}_i . To be specific, the value $P(p_i)$ is defined as follows:

$$P(p_i) \triangleq \frac{C^2(p_i, q_i)}{C^2(p_i, q_i) + C^2(p_i, \tilde{q}_i)}. \quad (9)$$

4.5. Results. We compare our deep neural architecture with an MLP-based system with two hidden layers and a CCA-based system with two RBMs used. The experimental results of different methods are shown in Table 1. Our BDA achieves the highest accuracy of 88.96% among the three models. And the CCA-based system gains better performance than that of the MLP-based system.

Besides, we investigate the effect of hyperparameter α in loss function (6). Figure 6 shows the accuracies achieved of different values of α , by setting α from 0.0 to 1.0 with 0.1 as the step size. As shown in this figure, the best performance

TABLE 1: Accuracies achieved by BDA and benchmark methods.

Method	Accuracy (%)
BDA	88.96
CCA [17]	85.54
MLP [10]	81.54

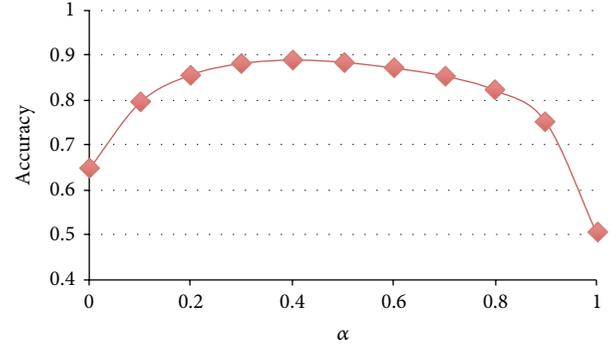


FIGURE 6: The effect of hyperparameter α on our system.

is achieved when the value of α is equal to 0.4. Note that, when the value of α is set to 1, the BDA degenerates to two autoencoders resulting in an accuracy with random guess; when the value of α is set to 0, the BDA degenerates to an only one autoencoder, with a higher accuracy than that of a random guess.

4.6. Discussion. In this section we will make a discussion on the three methods and the experimental results. The models used in our experiments have something in common. Firstly, for image modality, some popular descriptors are adopted. And for the text modality, we adopt the classical BOW model. These low-level modal specific representations are substantially built for further more abstract representations. Secondly, all the three models obtain a cross modal metric enabling these heterogeneous data from different sources to be comparable.

Note that the experimental results are largely affected by the differences among these models. In the MLP-based system, the nonlinear mapping function is obtained directly from the low-level representation of image modality to that of text modality. The assumption behind the MLP-based system is that the text representation is abstract whilst the image representation is concrete, relatively. In contrast, the other two systems, the CCA-based and our BDA-based system, treat the bimodal data as the same starting point. The assumption behind these systems is that there exists a common representation space for bimodal data of one object. Recent neurobiological research [38] has a strong support for the latter. Therefore, the BDA-based and CCA-based systems outperform the MLP-based system. Furthermore, note that the CCA only considers the linear coefficient, whilst our BDA uses the nonlinear reconstruction and compatibility constraints. Some important information related to these two modalities cannot be captured by the CCA method. Overall, the accuracies achieved by the three models are expected.

The experiments on the hyperparameter α show that the BDA relates to the unimodal autoencoder, as already shown in the above experimental results. And also it is interesting to note that the highest accuracy is not achieved at the medium value of 0.5. It means that the bimodal autoencoder in the BDA system primarily keeps the reconstructions of modality-specific data and secondly considers the compatibility between modalities. This could be used as a guideline for heuristically setting the hyperparameter α .

5. Conclusion

To conclude, we propose a computational deep neural architecture for measuring the similarity between different modalities in complex system with image and text. Our proposed framework closely combines feature extraction methods for individual modality and deep neural networks using stacked RBMs and a variant of neural autoencoder architecture. We show experimentally the effectiveness of our approach to the task of classifying image tags on public available datasets.

Our computational framework is flexible and could be extended in several ways. For example, we could explore more feature extraction methods in the first component. As another example, complex neural based representation could be exploited. Moreover, this architecture presented here can be naturally extended to other different modalities. In our future work we will investigate this flexibility in other complex systems.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was partially supported by National Natural Science Foundation of China (nos. 61273365, 61100120, 61202247, and 61202248), National High Technology Research and Development Program of China (no. 2012AA011103), Discipline Building Plan in 111 Base (no. B08004), and Fundamental Research Funds for the Central Universities (no. 2013RC0304). The authors would also like to thank the editor and the anonymous reviewers for their useful comments and suggestions that allowed them to improve the final version of this paper.

References

- [1] L. Li, H. Peng, J. Kurths, Y. Yang, and H. J. Schellnhuber, "Chaos-order transition in foraging behavior of ants," *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8392–8397, 2014.
- [2] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–27, 2009.
- [3] K. Chen and A. Salman, "Learning speaker-specific characteristics with a deep neural architecture," *IEEE Transactions on Neural Networks*, vol. 22, no. 11, pp. 1744–1756, 2011.
- [4] A.-R. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 1, pp. 14–22, 2012.
- [5] G. Heigold, V. Vanhoucke, A. Senior et al., "Multilingual acoustic models using distributed deep neural networks," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '13)*, pp. 8619–8623, IEEE Computer Society, Vancouver, Canada, May 2013.
- [6] D. Yu, L. Deng, and F. Seide, "The deep tensor neural network with applications to large vocabulary speech recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 2, pp. 388–396, 2013.
- [7] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [8] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Communications of the ACM*, vol. 54, no. 10, pp. 95–103, 2011.
- [9] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, pp. 1106–1114, Morgan Kaufmann, Lake Tahoe, Nev, USA, 2012.
- [10] I. J. Goodfellow, D. Erhan, P. L. Carrier et al., "Challenges in representation learning: a report on three machine learning contests," in *Proceedings of the 20th International Conference on Neural Information Processing*, pp. 117–124, IEEE Computer Society, Daegu, Korea, 2013.
- [11] C. Farabet, C. Couprie, L. Najman, and Y. Lecun, "Learning hierarchical features for scene labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1915–1929, 2013.
- [12] F. Feng, X. Wang, and R. Li, "Cross-modal retrieval with correspondence autoencoder," in *Proceedings of the 22nd ACM International Conference on Multimedia*, pp. 7–16, ACM, Orlando, Fla, USA, 2014.
- [13] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 127–134, ACM, New York, NY, USA, 2003.
- [14] E. P. Xing, R. Yan, and A. G. Hauptmann, "Mining associated text and images with dual-wing harmoniums," in *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI '05)*, pp. 633–641, AUAI Press, Arlington, Va, USA, July 2005.
- [15] Y. Jia, M. Salzmann, and T. Darrell, "Learning cross-modality similarity for multinomial data," in *Proceedings of ACM the International Conference on Multimedia Information Retrieval*, pp. 2407–2414, IEEE, Washington, DC, USA, November 2011.
- [16] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 539–546, IEEE, Washington, DC, USA, June 2005.
- [17] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th International Conference on Machine Learning*, pp. 689–696, Omnipress, Bellevue, Wash, USA, July 2011.
- [18] N. Srivastava and R. Salakhutdinov, "Multimodal learning with deep Boltzmann machines," in *Advances in Neural Information Processing Systems*, vol. 25, pp. 2231–2239, Morgan Kaufmann, Lake Tahoe, Nev, USA, 2012.

- [19] B. McFee and G. Lanckriet, "Learning multi-modal similarity," *Journal of Machine Learning Research*, vol. 12, no. 8, pp. 491–523, 2011.
- [20] J. Masci, M. M. Bronstein, A. M. Bronstein, and J. Schmidhuber, "Multimodal similarity-preserving hashing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 824–830, 2014.
- [21] P. Smolensky, "Information processing in dynamical systems: foundations of harmony theory," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group, Eds., vol. 1, pp. 194–281, MIT Press, Cambridge, Mass, USA, 1986.
- [22] M. Welling, M. Rosen-Zvi, and G. Hinton, "Exponential family harmoniums with an application to information retrieval," in *Advances in Neural Information Processing Systems 17*, pp. 501–508, Morgan Kaufmann, Vancouver, Canada, 2004.
- [23] R. Salakhutdinov and G. Hinton, "Replicated softmax: an undirected topic model," in *Advances in Neural Information Processing Systems*, vol. 22, pp. 1607–1614, Morgan Kaufmann, Vancouver, Canada, 2009.
- [24] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [25] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems 19*, pp. 153–160, MIT Press, Vancouver, Canada, 2007.
- [26] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.
- [27] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F.-J. Huang, "A tutorial on energy-based learning," in *Predicting Structured Data*, G. Bakir, T. Hofman, B. Schölkopf, A. Smola, and B. Taskar, Eds., pp. 1–59, MIT Press, Cambridge, Mass, USA, 2006.
- [28] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [29] L. von Ahn and L. Dabbish, "Labeling images with a computer game," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 319–326, ACM Press, Vienna, Austria, 2004.
- [30] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is real-world visual object recognition hard?" *PLoS Computational Biology*, vol. 4, no. 1, article e27, 2008.
- [31] A. Coates and A. Y. Ng, "The importance of encoding versus training with sparse coding and vector quantization," in *Proceedings of the 28th International Conference on Machine Learning (ICML '11)*, pp. 921–928, Bellevue, Wash, USA, July 2011.
- [32] F. M. Ham and I. Kostanic, *Principles of Neurocomputing for Science and Engineering*, McGraw-Hill Higher Education, 1st edition, 2000.
- [33] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, NY, USA, 3rd edition, 2003.
- [34] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: an overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
- [35] N. Rasiwasia, J. C. Pereira, E. Coviello et al., "A new approach to cross-modal multimedia retrieval," in *Proceedings of the 18th ACM International Conference on Multimedia*, pp. 251–260, ACM, New York, NY, USA, 2010.
- [36] J. Kim, J. Nam, and I. Gurevych, "Learning semantics with deep belief network for cross-language information retrieval," in *Proceedings of the 24th International Conference on Computational Linguistics*, pp. 579–588, ACL Press, IIT Bombay, India, 2012.
- [37] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [38] S. V. Shinkareva, V. L. Malave, R. A. Mason, T. M. Mitchell, and M. A. Just, "Commonality of neural representations of words and pictures," *NeuroImage*, vol. 54, no. 3, pp. 2418–2425, 2011.

Research Article

Prevention and Trust Evaluation Scheme Based on Interpersonal Relationships for Large-Scale Peer-To-Peer Networks

Lixiang Li,^{1,2,3,4} Jürgen Kurths,⁵ Yixian Yang,² and Guole Liu²

¹ Institute of Network Coding, The Chinese University of Hong Kong, Shatin, NT, Hong Kong

² Information Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, P.O. Box 145, Beijing 100876, China

³ Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518057, China

⁴ Shenzhen Key Laboratory of Network Coding Key Technology and Application, Shenzhen 518057, China

⁵ Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany

Correspondence should be addressed to Lixiang Li; li_lixiang2006@163.com

Received 21 July 2014; Accepted 24 August 2014; Published 29 September 2014

Academic Editor: Haipeng Peng

Copyright © 2014 Lixiang Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, the complex network as the frontier of complex system has received more and more attention. Peer-to-peer (P2P) networks with openness, anonymity, and dynamic nature are vulnerable and are easily attacked by peers with malicious behaviors. Building trusted relationships among peers in a large-scale distributed P2P system is a fundamental and challenging research topic. Based on interpersonal relationships among peers of large-scale P2P networks, we present prevention and trust evaluation scheme, called IRTrust. The framework incorporates a strategy of identity authentication and a global trust of peers to improve the ability of resisting the malicious behaviors. It uses the quality of service (QoS), quality of recommendation (QoR), and comprehensive risk factor to evaluate the trustworthiness of a peer, which is applicable for large-scale unstructured P2P networks. The proposed IRTrust can defend against several kinds of malicious attacks, such as simple malicious attacks, collusive attacks, strategic attacks, and sybil attacks. Our simulation results show that the proposed scheme provides greater accuracy and stronger resistance compared with existing global trust schemes. The proposed scheme has potential application in secure P2P network coding.

1. Introduction

In P2P networks, due to their characteristics of openness, anonymity, and dynamic nature of P2P networks without verifications, peers can freely join in and leave the systems, which leads to a P2P system being vulnerable and easily attacked by malicious peers [1–16]. In P2P systems about 50% of the network peers perform malicious behaviors, which are providing false services, spreading malicious codes or viruses, and so forth [1, 4, 5, 10, 11]. In order to encourage and stimulate peers to participate in the system, it is very important to ensure the authenticity of shared resources and to resist malicious peers. One approach is building trust and reputation management to promote a good collaborative relationship among peers in the P2P systems. In these trust and reputation systems, the two matters of prime importance

are ensuring the highly accurate trustworthiness of calculating trust and being robust to malicious peers. Currently, most trust and reputation systems focus on evaluating the credibility of resource providers [1–16], but in the absence of any central authority, repository, and global information there is no silver bullet for securing P2P networks [5, 10, 11].

In this paper, we present a prevention and trust evaluation scheme based on interpersonal relationships for large-scale P2P networks. The major contributions of this paper are as follows.

First, although some existing schemes consider the weights of peers, they believe that the weights of those peers with high credibility are greater than those of the peers with low credibility in the trust computing. From the point of view of interpersonal relationships, this view may not be completely reasonable. We put forward that the weights of

the peers who are familiar with the request peer are greater than those of the peers who are not. We also believe that the early experiences have a small impact on the trust evaluation, whereas the recent experiences contribute more.

Second, when it comes to the trust of peers, the existing schemes take it for granted that the QoS is equal to credibility. As a consequence, a request peer evaluates only the QoS of the provider. Considering the recommendation of other peers, the request peer merely considers the credibility of these peers, and how to calculate the credibility is very complex. In response to this problem, we distinguish QoS and QoR among peers. For one specific transaction, we propose that the request peer evaluates the QoS of the service provider and the QoR of the recommenders as well, which can easily improve the ability of resisting malicious behaviors.

Third, in order to prevent and describe the unpredictable and uncertain behaviors from malicious peers, we take all risk factors into consideration such as peer's trust value, context, transactions, and the accumulated speed of trust included. The risk value is used to prevent and measure various malicious behaviors, such as fluctuating behavior and misuse of trust. In our design, the weights of the risk factor are adjustable so that they can be effectively applied to different environments with different requirements.

2. Related Works

The main problem of reputation systems is how to deal with trust networks [13]. A trust network is a virtual network on top of a P2P system as shown in Figure 1, in which the bottom layer is the physical entities composed of the terminal machines, the middle layer is the logic layer of P2P systems, and the top layer is the trust layer composed of the trust ratings after peers' transactions.

The EigenTrust scheme [1] computes a global trust value for a peer by performing a distributed algorithm approaching the eigenvector of the trust matrix over the peers. The scheme relies on a reliable selection of some pretrust peers, who are supposed to be trusted by all peers and have higher trust. This assumption may not always be matched in a distributed computing environment. Because the pretrust peers may not exist lastly, and once they behave any malicious actions and score badly after some transactions, the system may not work reliably.

In a PeerTrust scheme [2], Xiong and Liu proposed three basic trust parameters and two adaptive factors to compute the trustworthiness of peers. They incorporated the concepts of a trust value and similarity with itself to compute credibility and satisfaction. The trustworthiness of a peer is considered to be a mean value of the evaluation of peers' behaviors, while the evaluation is given by the nearby peers. However, the limitation of this approach is that the computation convergence rate in large-scale P2P systems is not provided. The five factors used in their trust scheme may be retrieved with a heavy overhead.

In a R2Trust scheme [10], peers' trust values are evaluated from direct interactions and peers referrals. The scheme

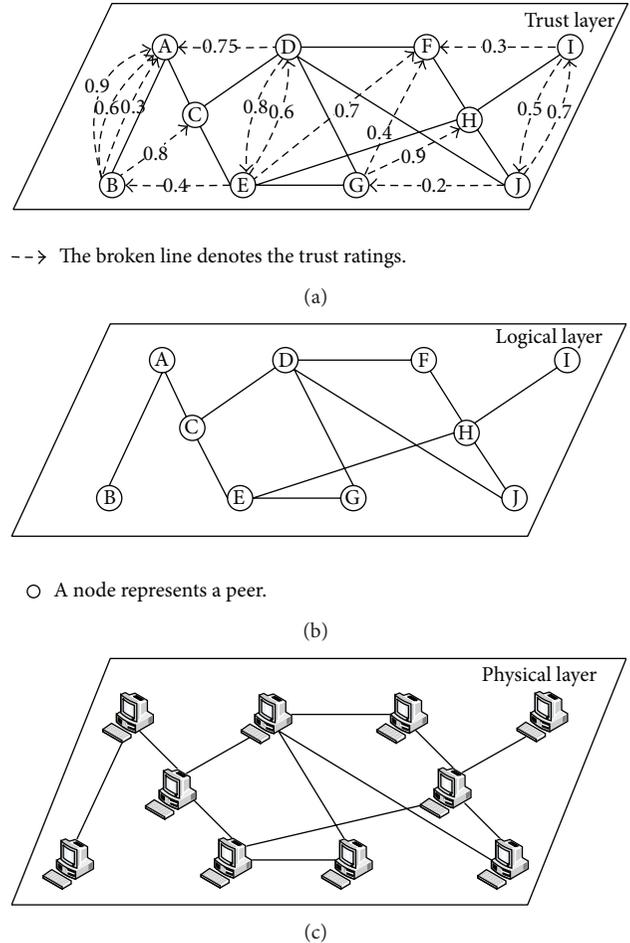


FIGURE 1: Trust overlay networks for a P2P system (color online).

distinguishes the credibility of peers. As a result, the aggregated trust value will filter out the noises and provide more accurate trust values. It can defend against several malicious attacks, such as simple malicious attacks, collusive attacks, and strategic attacks. But it has not taken measures to restrain fake and sybil attacks [17–19]. Li et al. [11] proposed multiple factors to be incorporated to reflect the complexity of trust. The weighted moving average and ordered weighted averaging combination algorithms can dynamically assign the properties (weights) of the multiple factors and calculate a more accurate trust value. But they did not give detailed analysis about resisting against peers' malicious behaviors of the scheme.

The above P2P trust and reputation systems are based on the assumption that the better the peer's QoS is, the higher the reputation of the peer is. These schemes only use a one-sided trust factor to quantify and predict trustworthiness among peers, which leads to lower resistance to malicious behaviors. In this paper, considering the interpersonal relationships among peers, the reputation of the peer is divided into service trust and recommendation trust. In addition, we put forward the comprehensive risk factors to prevent the malicious behaviors and propose a robust and efficient P2P reputation

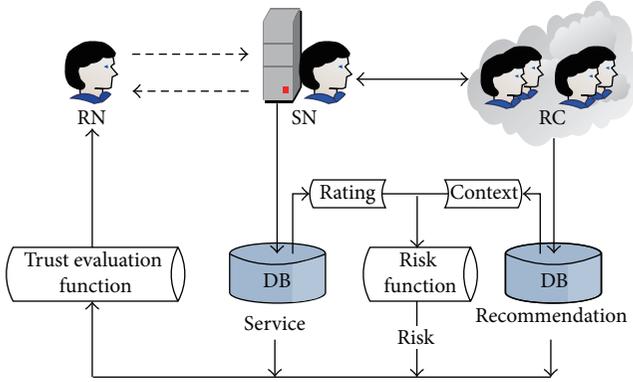


FIGURE 2: System structure of the prevention and trust evaluation scheme (color online).

system. We construct a mathematic scheme to discuss the capacity of resisting against malicious behaviors. Finally, we conduct simulations to illustrate them.

3. The Prevention and Trust Evaluation Scheme

In this section, we will present the basic framework of the prevention and trust evaluation scheme in detail.

3.1. Basic Framework. The behaviors of the network peers are determined by the operators and the peers should reflect the characteristics of operators. Based on the various roles in P2P systems, peers can be divided into three types: service node (SN), request node (RN), and recommendation node (RC). SN represents the peers who provide services for others in the networks, RN is the peers who request services, and RC denotes the peers who had transactions with SN in the past and recommend feedback to others. The detailed framework is shown in Figure 2.

Before a request node deals with a service node, the request node needs to consider three factors: the history records, the recommendation information, and the comprehensive risk factor. The decision information formed by the three factors through a trust evaluation function is sent to the request node. Then the request node determines whether the transaction is with the service node or not.

A peer enters the system with a unique identity, and it needs to be authenticated to communicate with other peers. When a new peer joins the P2P system, it cannot gain all the authority. It can only provide services to other peers, and when the number of the services is greater than the initial threshold (ϕ), the new peer can enter into the normal trust accumulation. With an the increase of the trust, the peer will have more authority. The accumulation of the trust is very slow, and once the peer has malicious behaviors, the accumulated trust is quickly reduced. Once a peer turns into a malicious one, it will lose all privileges.

TABLE 1: The transaction ratings of quality of service and quality of recommendation.

QoS	QoR	Description	Value
<i>Good</i>	<i>Good</i>	The peer is good.	(0.5, 1]
<i>Common</i>	<i>Common</i>	The peer is correct.	(0, 0.5]
<i>Dishonest</i>	<i>Dishonest</i>	The peer is dishonest.	[-0.5, 0)
<i>Malicious</i>	<i>Malicious</i>	The peer is malicious.	[-1, -0.5)

3.2. Trust Evaluation. Peers' trust evaluation consists of three parts: QoS trust value, QoR trust value, and risk factor value. All these three parts are calculated based on the history records of transaction ratings. We first classify the transaction ratings based on the quality of the provided by the SN and the RC, as shown in Table 1. We define the quality set $QS = \{Good, Common, Dishonest, Malicious\}$. The rating *Good* represents the good and honest peers, *Common* represents the peers who cause some damage, *Dishonest* represents the peers who provide false services or false recommendation, and *Malicious* represents the peers who spread the viruses or malicious codes. Note that the classification is flexible, and more classes or subclasses can be introduced if it is necessary [11, 13].

How to calculate these three parts' trust values will be discussed in Sections 4.1–4.3 in detail. The overall trust value of a peer is the maximum value of the three parts, and let $TV_{i,j}$ denote the overall trust evaluation of the service provider j from the view point of the request peer i . Let $TS_{i,j}$ denote the QoS trust value of peer j , $TR_{i,j}$ is the QoR trust value of these peers having transactions with peer j in the past, and $R_{i,j}$ is the risk factor value of the service provider j . Therefore, the overall trust value for the service provider j at request peer i is defined as

$$TV_{i,j} = \max(\alpha TS_{i,j}, \beta TR_{i,j}, \gamma R_{i,j}), \quad (0 \leq \alpha, \beta, \gamma \leq 1), \quad (1)$$

where α , β , γ are the weights of the related trust value and $\alpha + \beta + \gamma = 1$.

4. Trust Calculation and Analysis of the Resistance to Malicious Behaviors

In this section, we will present the trust calculation of our IRTrust scheme and analyze how to resist peers' malicious behaviors.

4.1. Service Trust Value. Let us assume that the request peer i finds the resource located in the peer j . The peer i has k transactions with peer j in the current time cycle m . The ratings that peer j got from peer i is the rating sequence $\{r_1, r_2, \dots, r_k\}$. Peer j accumulated a service trust value of the

k th transaction in the current time cycle m which is defined as

$$V_{i,j}(k) = \begin{cases} V_{i,j}(k-1) + e^{-a_1 * |r_k - 0.5|}, & r_k \in (0.5, 1], \\ V_{i,j}(k-1) + e^{-a_2 * |r_k + 0.5|}, & r_k \in (0, 0.5], \\ V_{i,j}(k-1) - e^{-a_3 * |r_k - 0.5|}, & r_k \in [-0.5, 0), \\ V_{i,j}(k-1) - e^{-a_4 * |r_k + 0.5|}, & r_k \in [-1, -0.5), \end{cases} \quad (2)$$

where a_1, a_2, a_3, a_4 are the adjustment factor of the accumulated reputation. The function e^{-x} decreases with the increasing of x , the accumulated reputation of *Good* rating should be better than that of *Common* rating, and the lost reputation of *Dishonest* rating should be less than that of *Malicious* rating. Thus, the adjustment factor should meet the following conditions $a_1 < a_2, a_3 > a_4$, which merge the accumulation function and the penalty function. The accumulated service trust value of the current time cycle m is defined as

$$AS_{i,j}(m) = \frac{\sum_{i=0}^k V_{i,j}(k)}{k}. \quad (3)$$

If there are n time cycles $\{t_1, t_2, \dots, t_n\}$ from the beginning T_{start} to the end T_{end} and peer i has the number of transactions with peer j in each time cycle, the direct service trust value $TS_{i,j}$ is computed directly from peer i 's historical ratings for peer j . We define $TS_{i,j}$ as follows:

$$TS_{i,j} = \begin{cases} \frac{\sum_{i=1}^n AS_{i,j}(i1)}{n}, & AS_{i,j}(i1) > 0, \\ 0, & AS_{i,j}(i1) = 0. \end{cases} \quad (4)$$

4.2. Recommendation Trust Value. In a fully distributed P2P system involving numerous nodes, it is often not possible for a request peer to directly connect the service peer and to assess the trust value of the service provider. Instead, the request peer needs to resort to other peers in the P2P system and rely on the collective opinions to assess the trust. Although some trust schemes based on recommendation have already been proposed [2, 4, 5, 8, 9], they also lead to new challenges, such as how to determine the accuracy of collected opinions and how to efficiently aggregate referrals from diverse recommenders with different trustworthiness.

The recommenders' reputations are different from each other. The referrals from peers with high reputation are more trustworthy than those from peers with low reputation [2]. However, the referrals are treated equally and their credibility are not considered. In order to neutralize different reputation reports, Tian and Yang [10] proposed the credibility to weigh the feedback of the referrals. The recommenders from peers with high credibility are more trustworthy than those from low credibility peers. In this paper, we consider the peers in the P2P system as reflecting the characteristics of the interpersonal relationships. Due to that the peers with high reputation or credibility may not be willing to give recommendations to new peers. For example, a respected

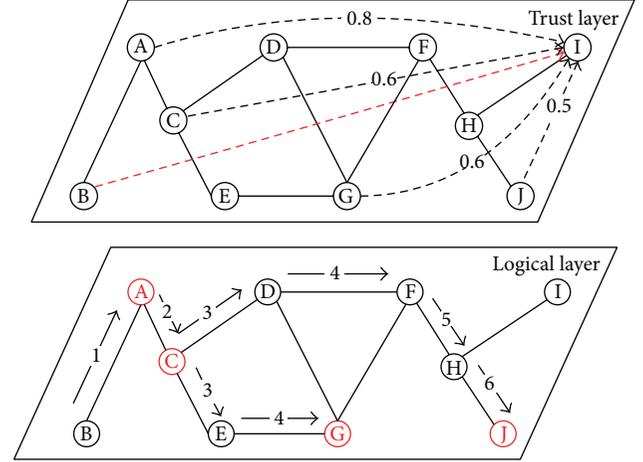


FIGURE 3: Computing the jumps of referrals (color online).

famous person may probably not write a recommendation for a stranger whom he (she) never knows. Therefore, the weight of the peers with high reputation or credibility may not be the greatest for the request peer. We deem that the acquainted peers of the request peer should have higher recommending weights.

In IRTrust scheme, we use the jumps among peers of the logical layer to describe the relationships. When the jumps value is 1, the neighbors are the most acquainted ones to the request peer. The more the jumps are, the more remoter the relationships between the request peer and the recommenders are. For example, as shown in Figure 3, the service provider is peer I , the request peer is B , and the referral peers $\{A, C, G, J\}$ have transactions with the service provider in the trust layer. In the logical layer, the jumps with peer B are $\{1, 2, 4, 6\}$. Maybe the reputation or credibility of the peer J is the highest, but the recommending weight of peer A is the highest.

Suppose that the peer i has had k transactions with peer j in the current time cycle m . The referral peer P got the recommended rating sequence $\{c_1, c_2, \dots, c_k\}$ from peer i . In the current time cycle m , we define the P accumulated recommendation trust value of the k th transaction as

$$C_{i,j}(k) = \begin{cases} C_{i,j}(k-1) + e^{-b_1 * |r_k - 0.5|}, & r_k \in (0.5, 1], \\ C_{i,j}(k-1) + e^{-b_2 * |r_k + 0.5|}, & r_k \in (0, 0.5], \\ C_{i,j}(k-1) - e^{-b_3 * |r_k - 0.5|}, & r_k \in [-0.5, 0), \\ C_{i,j}(k-1) - e^{-b_4 * |r_k + 0.5|}, & r_k \in [-1, -0.5), \end{cases} \quad (5)$$

where $C_{i,j}$ is the recommended rating sequence from i to j , and b_1, b_2, b_3, b_4 are the adjustment factors of the accumulated recommendation trust and they should meet $b_1 < b_2, b_3 > b_4$. The accumulated recommendation trust value of peer P in the current time cycle m is defined as

$$AR_{i,j}(P, m) = \frac{\sum_{i=0}^k C_{i,j}(k)}{k}. \quad (6)$$

If there are n time cycle $\{t_1, t_2, \dots, t_n\}$ from the beginning T_{start} to the end T_{end} and the peer i has the number of transactions with the peer j in each time cycle, the recommendation

trust value $AR_{i,j}(P)$ is computed from the peer i 's historical ratings for the peer P . We define $AR_{i,j}(P)$ as follows:

$$AR_{i,j}(P) = \frac{\sum_{i2=1}^n AR_{i,j}(P, i2)}{n}. \quad (7)$$

Assume that peers $\{P_1, P_2, \dots, P_n\}$ are referrals to request peer i and that the relationship distances with peer i are $\{h_1, h_2, \dots, h_n\}$. The inverse of h is the weight of referrals. If the h value is too large, it means that the request node needs to traverse the whole network, which will very strongly increase the computational effort. Here, we use the flooding algorithm with TTL to search the referrals. (The maximum of TTL value is 7 [20–22].) If $h > 7$ or the referral peer has left the network, we conform its h value to 10. Thus, the indirect recommendation trust value $TR_{i,j}$ is defined as

$$TR_{i,j} = \frac{\sum_{i3=1}^n (1/h_{i3}) AR_{i,j}(P_{i3})}{n}. \quad (8)$$

4.3. Risk Factor Value. The risk factor is used to describe the probability of the service provider's being a malicious peer. The reputation of service trust is an accumulative value for the past behaviors and reflects the overall evaluation got from the responding peers. However, it is not sensitive enough to perceive suddenly malicious behaviors of peers, because the value is posterior, and it will be decreased after the peer is spoiled. Thus, the risk factor is used to portray the unpredictable and uncertain behaviors of those potential malicious peers.

Li et al. [11] proposed the risk window to calculate the risk. The smaller of risk window size, the more accurate the risk assessment. But this will decrease the availability of the resources, because less risk for cooperation is requested and less peers are qualified to be cooperative. In R2Trust, the risk value is computed by applying the concept of information entropy which has been proven to be applicable in dealing with uncertain problems [10]. In this paper, the risk factor includes three parts: the rating risk factor (rr), the context risk factor (rt), and the cumulative speed factor (rs).

The rating risk factor rr can be used to restrain the simple and collusive malicious attacks. The rating risk factor is defined as

$$rr = \frac{N_0}{N}, \quad (9)$$

where N is the number of ratings and N_0 is the number of ratings which are less than zero.

The contexts are complete transaction records including size, category, and time stamp. Here, we uniform the turnover of transactions. The context risk factor is used to restrain the strategic attack of peers. We define it as

$$rt = \frac{\text{turnover}_c}{\sum_{\text{past}} \text{turnover}}, \quad (10)$$

where turnover_c is the current transaction turnover.

The cumulative speed factor is the increased speed of trust of a service provider. It is also used to restrain the strategic attack which is defined as

$$rs(m, m-1) = \frac{AS(m) - AS(m-1)}{T}, \quad (11)$$

where m is the current time cycle, $m-1$ is the pretime cycle, and T is the length of time cycle. If there are n time cycles $\{t_1, t_2, \dots, t_n\}$ from the beginning T_{start} to the end T_{end} ,

$$rs = \frac{\sum_{i4=1}^n rs_{i4}}{n}. \quad (12)$$

The risk factor value is the overall value of the three parts and is defined as

$$R_{i,j} = \max(rr, rt, rs). \quad (13)$$

The larger the risk factor value, the higher the probability that the service provider is the malicious peer.

4.4. Simple Attack Analysis. When this kind of malicious peers attacks, the vicious behaviors are isolated. They always provide the inauthentic services, maliciously slander the QoS of good peers, exaggerate the QoS of malicious peers, or fake the peers with high reputation [1–16]. For this kind of malicious peers attack, some literatures have already proposed corresponding measures. In PeerTrust, the feedback of a peer and the credibility factor for the feedback are used to resist the exaggerating and slandering [2]. In R2Trust, a reputation evaluation factor is proposed to resist the malicious behaviors [10]. In [23], digital signature is used to restrain the fake behaviors. In this paper, we described our measures which can restrain these malicious behaviors.

If a peer provides inauthentic services, such as false files, malicious code, and Trojan virus, the penalty function as is given in (2) will reduce its service reputation quickly, and this will seriously weaken the role of the peer in the network. When its service reputation is low enough, the peer will lose its effectiveness, that is, downloading and recommending. If a peer maliciously slanders the QoS of good peers or exaggerates the service quality of malicious peers, its corresponding reputation will be reduced when one of the others is communicated with the peer which is maliciously slandered or exaggerated. Then the peer will lose the corresponding effectiveness. To avoid a peer faking other peers with high reputation, we adopt the authentication technology based on identification in the process of peers' communication.

4.5. Collusion Attack Analysis. Collusion attack is broadly defined as any malicious coordinated behavior of a group of users aimed at gaining undeserved benefits or at damaging well behaved users. Thus, this kind of attack behavior is a kind of joint attack of a group of peers. In the group, each peer is well behaved and never provides inauthentic services or defames the reputation of other members. But this type of malicious peers can form a malicious cycle by assigning a high trust value to other malicious peers in the network.

Collusive peers provide inauthentic services to outsiders when selected as download sources and provide denigrated ratings for those noncollusive peers [2, 7, 24–30].

The collusive peers can hide their malicious intentions in an unstructured P2P system by assigning high trust value to each other. To solve this problem, Xiong and Liu assumed that there is no similarity in the evaluation to peers mutually, when a collusive peer does a good rating to the target peer [2]. The tolerance to the collusive attack has been improved to put weight on the evaluation value by using this similarity. Sato proposed the trust estimation method to give a defense against the collusive attack: the maximum likelihood estimation method [7]. Some literatures use encryption algorithms, game theory, and economics to resist a collusion attack [27–30].

In this paper, being different with previous literatures, the reputation of peers is divided into service reputation and recommendation reputation. When peers communicate or trade with each other, the request peer needs to evaluate the QoS of the service peer, and at the same time, it is required to evaluate the QoR of the peers who have connections with the service peer. This measure can effectively resist collusion attack. Once one of the collusion peers maliciously attacks the request peer, not only the service reputation of the peer but also all its neighbors' recommendation reputation will be punished by the penalty function as is given in (5). Namely, if a peer maliciously attacks, the whole group will be punished.

4.6. Strategy Attack Analysis. This kind of malicious attack behavior has a clear purpose. In the beginning, the malicious peers are well behaved and accumulated the reputation by honest services. Once these peers build good reputation, then they will start to abuse their credibility to mislead other peers [1]. They always adopt low cost to increase their reputation when they do a strategy attack. Taking the overdraft of credit card as an example, one person wants to overdraw the credit card maliciously. First, he/she needs to increase his/her credit in order to get more amounts. The strategy is to pay the small overdraft back promptly. The credit is accumulated by the frequent consumption-repayment process. When the credit reaches to a certain degree, the person will own more overdraft amounts. Eventually the person can overdraw all the amounts and does not pay the debts again.

Strategy attack is also called onCoff attack in some literatures. To address such potential dynamic behaviors of peers, Xiong and Liu proposed a simple adaptive time window-based algorithm to better react to the above behaviors [2]. Tian and Yang introduced a risk factor value and information entropy to the trust computation in order to portray the unpredictable and uncertain behaviors of these malicious peers [10]. In this paper, we retain the parameter of the context risk factor, and at the same time, we define the reputation accumulative speed as is given in (10) and (12).

4.7. Sybil Attack Analysis. A sybil attack is the one in which a malicious attacker subverts the reputation system of P2P network by creating a large number of pseudonymous entities and uses them to gain a disproportionately large influence.

The reputation system's vulnerability to a sybil attack depends on (i) how cheaply identities can be generated, (ii) the degree to which the reputation system accepts inputs from entities that do not have a chain of trust linking them to a trusted entity, and (iii) whether the reputation system treats all entities identically.

A sybil attack is a powerful threat faced by any decentralized distributed P2P system that has no central, trusted authority to vouch for a one-to-one correspondence between users and identities [1]. Generally, the users only have a rating describing how well the user behaves in a reputation system. For example, eBay ratings are based on users' previous transactions with other users. Buyers and sellers in eBay rate each other after every transaction, and the overall reputation is the sum of these ratings over the last six months. Sybil attacks can create a large number of sybil nodes that collude to artificially increase a user's rating.

The existing literatures have done some researches on this type attack. Cheng and Friedman [17] surveyed and found that many existing reputation mechanisms were not resistant to this type of attack behavior. They used a static graph formulation of reputation and formalized the notion of sybilproofness. Yu et al. [18] proposed a detection mechanism (called SybilGuard) that relies on social networks of P2P users to limit the corruptive influences of sybil attacks. Quercia and Hailes [19] proposed an effective way of identifying sybil attackers for in-range portable devices (MobID) to reduce the number of interactions with sybil attackers and consequently enable collaborative applications.

In this paper, identity-based authentication techniques are used to prevent sybil attacks and dismiss masquerading hostile entities. When a peer joins the reputation system, it can only be one identity, and it is not allowed to create multiple identities. We use the SHA-1 hash function to realize the uniqueness and the anonymity. In addition, we set the transaction threshold ϕ (such as 50) to limit the permission of the new intransit peers. Namely, the new peers only provide service and have no permission to other operations. Once the number of service is more than the threshold, the peers may enter into the normal trust assessment. These measures can effectively resist sybil attacks.

5. Experiments and Comparisons

In this section, we will present results of our experiments which will show the effectiveness of our trust scheme. Firstly, we repeat the EigenTrust [1], the PeerTrust [2], and the R2Trust [10] simulator experiments using the Query Cycle Simulator and Matlab 2008. In our evaluation, we assess the performance of our scheme and compare it with the EigenTrust [1], the PeerTrust [2], and the R2Trust [10] schemes. We study their performance under a variety of malicious threat behaviors (discussed in Sections 4.2–4.7). We perform 100 query cycles in our simulations and the simulation results are the average expectations.

5.1. Simulation Environment. Considering the characteristics of P2P networks, such as the node degree distribution with

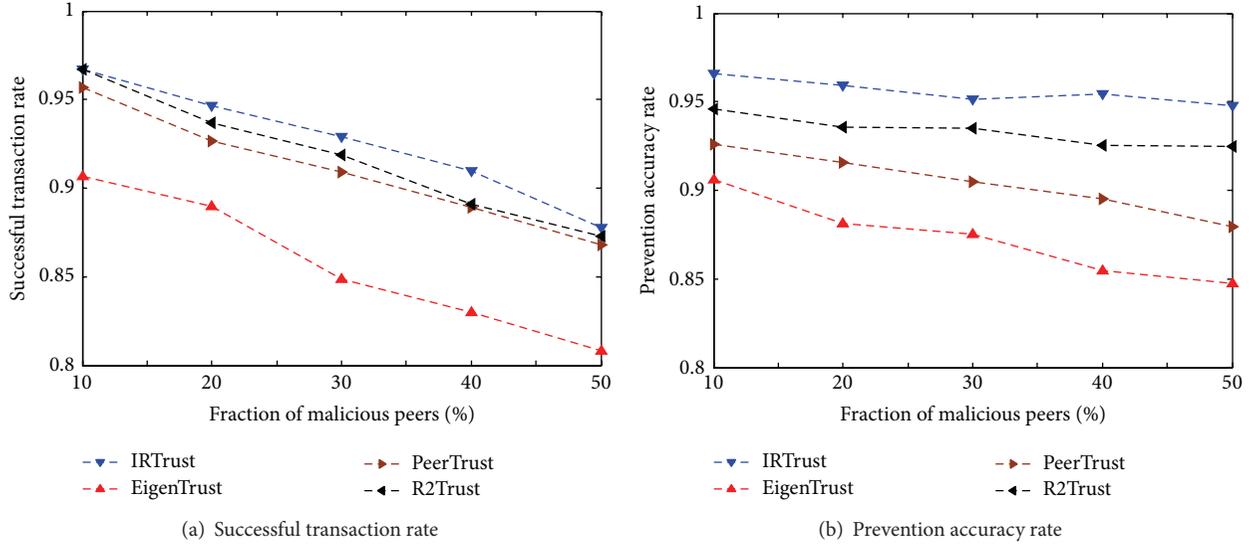


FIGURE 4: Performances of IRTrust, EigenTrust, PeerTrust, and R2Trust in simple attack (color online).

a power-law distribution, we construct the P2P network based on a BA scale-free network to approach to real-world networks. There are two types of peers: honest and malicious ones. The honest peers always give the correct service and feedback, but the malicious peers always give the opposite opinion to others when they own the corresponding power. In our experiments, we use the flooding algorithm to compute the interpersonal distance between peers; the detailed algorithm is given in the literature [20].

In the simulation we assume that there are 5000 peers in the network. Among them, there are 500–2500 malicious peers and the query message is flooded with TTL = 7. In the experiment, the peer's changing is in the uptime stage and is uniformly distributed over [0, 1]. The change of issuing queries in the uptime is uniformly distributed over [0, 0.5]. In addition, different types of peers also vary in their behaviors when responding to queries and providing files. For good peers, the probability of providing authentic files is 100%. Simple malicious peers will respond to all queries when they have received and provide inauthentic files with a probability of 100%. Collusive peers provide with a probability of 100% malicious files to other peers. Other parameters in the experiments are given in Table 2.

We compare the successful transaction rate and the prevention accuracy rate of our scheme with the EigenTrust, the PeerTrust, and the R2Trust under the conditions of simple, collusive, strategic, and sybil attacks. The metrics, successful transaction rate, is the ratio of the number of the successful transactions over the total number of transactions. It is typically used to evaluate the efficiency of a trust scheme [2, 10]. The prevention accuracy rate is the ratio of the number of successful transactions with forecasting the peer's status correctly over the total number of transactions.

5.2. Simple Attack. Figure 4 depicts our simulations under the condition of simple malicious peers. In Figure 4(a),

TABLE 2: The parameters and their values in the simulations.

Parameters	Description	Value
(α, β, γ)	Weight of direct trust value, indirect recommendation value, and peer's risk value	(0.3, 0.2, 0.5)
(a_1, a_2, a_3, a_4)	Adjustment factors	(3, 6, 4, 8)
(b_1, b_2, b_3, b_4)	Adjustment factors	(3, 6, 4, 8)
η	The ratio of malicious peers	(0.1–0.5)
ϕ	Threshold of the number of services	50
N	The size of P2P networks	5000

when the ratio of malicious peers is low in the system, all the four schemes perform high successful transaction rates. With the fraction of malicious peers increasing, the successful transaction ratio of all the schemes decreases, but it decreases most intensely in the EigenTrust scheme. However as a whole, the four schemes keep high efficiency, because they punish the malicious peers. The successful transaction ratio of EigenTrust scheme drops more quickly than the other schemes because the EigenTrust scheme does not differentiate the reliability of all the referrals. In comparison with the PeerTrust scheme and the R2Trust scheme, the proposed IRTrust scheme retains a very high successful transaction rate, and it even retains about 90% when the fraction of malicious peers is 50%.

Figure 4(b) depicts the prevention accuracy rate of these four schemes. Compared to all the four schemes, the rate of the EigenTrust scheme is the least and decreases with the increasing of the fraction of malicious peers. Although the experimental results show some fluctuations, the rate of the IRTrust scheme is the best and can reach 96%. But as a whole, the four schemes keep high efficiency, because they have punishment measures to malicious peers. In addition, the

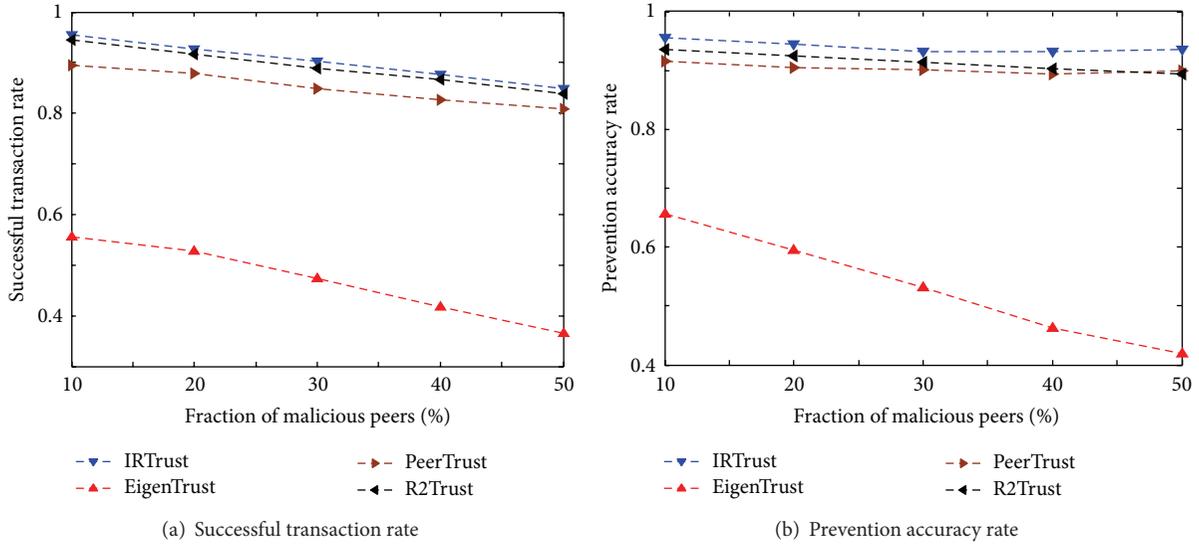


FIGURE 5: Performances of IRTrust, EigenTrust, PeerTrust, and R2Trust in collusion attack (color online).

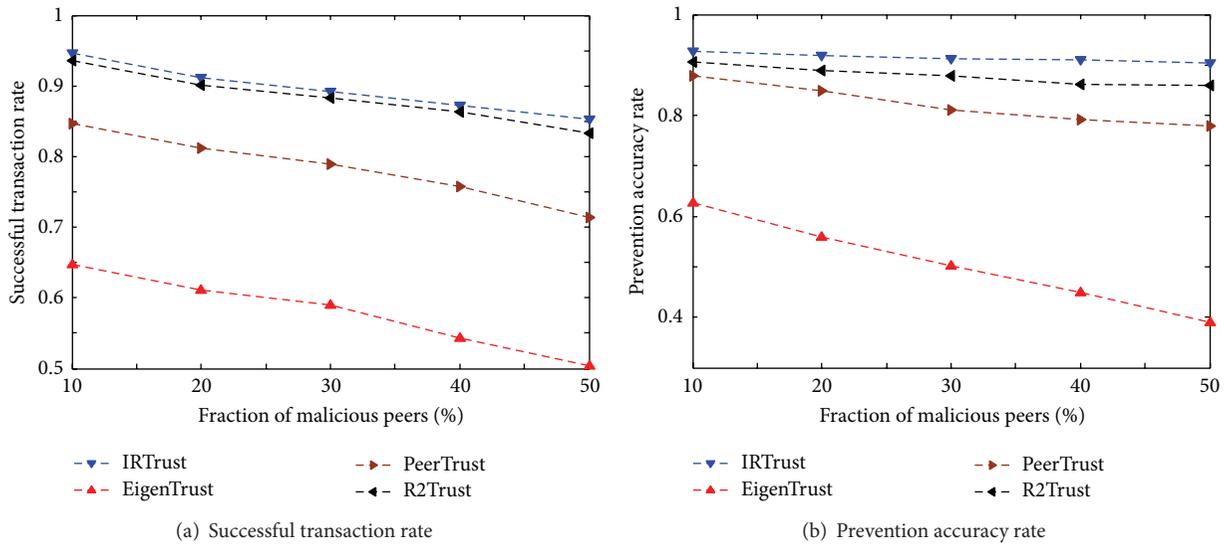


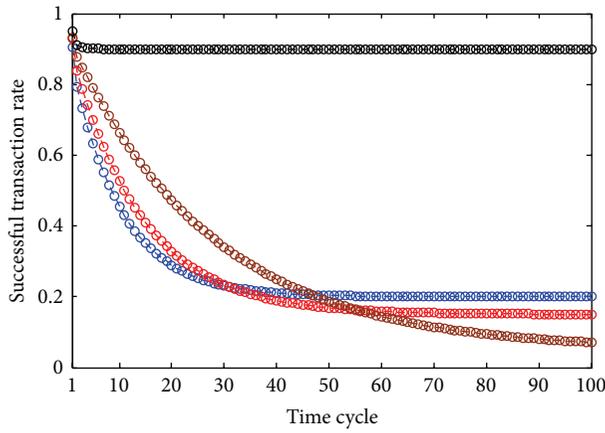
FIGURE 6: Performances of IRTrust, EigenTrust, PeerTrust, and R2Trust in strategy attack (color online).

R2Trust scheme and the IRTrust scheme have risk assessment measures.

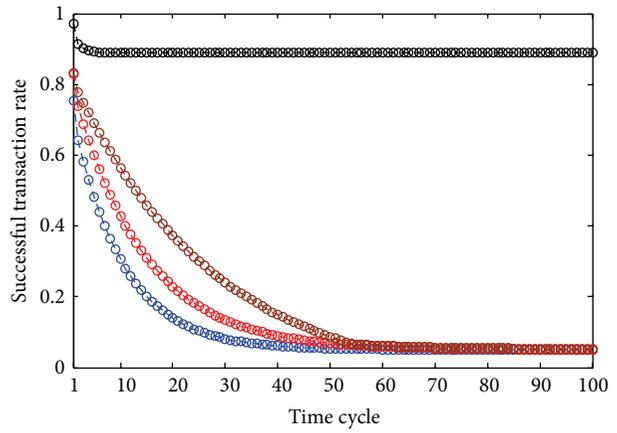
5.3. Collusion Attack. Figure 5 shows the simulations under the condition of collusive malicious peers. Each peer in the colluding group provides inauthentic services to the peers outward, boosts the trust value of their accomplices regardless of their behaviors, and downplays the trust value of good providers. In Figure 5(a), we can see that the successful transaction rate of the EigenTrust scheme descends obviously when the number of malicious peers increases. Because it does not give clear differentiations about the reliability of all the referrals, namely, more malicious peers will lead to more computations. Although the PeerTrust scheme owns high efficiency, it has no risk factor of peer, and it is inferior

to the two other schemes. Compared to the R2Trust scheme, the IRTrust scheme is designed to tackle collusive attacks with differentiating the service and recommendation, and therefore it is proved more robust against collusive attacks.

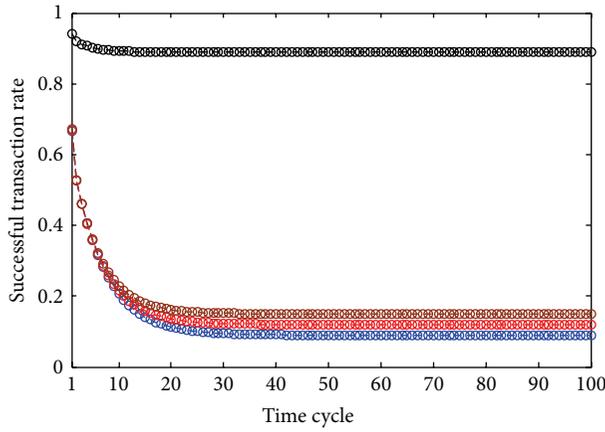
In Figure 5(b), we can see that the prevention accuracy rate of the EigenTrust scheme descends evidently when the number of malicious peers increases. Since the scheme can not differentiate the reliability of the referrals, the more malicious peers, the more transactions with the malicious peers. In the PeerTrust scheme, it has community factor to resist the collusive attack, and the R2Trust scheme uses the relative reputation difference to identify the collusive attack. Therefore, they can maintain a higher prevention accuracy rate to this type of malicious attack. Compared to the rest schemes, the IRTrust scheme is designed to tackle collusive



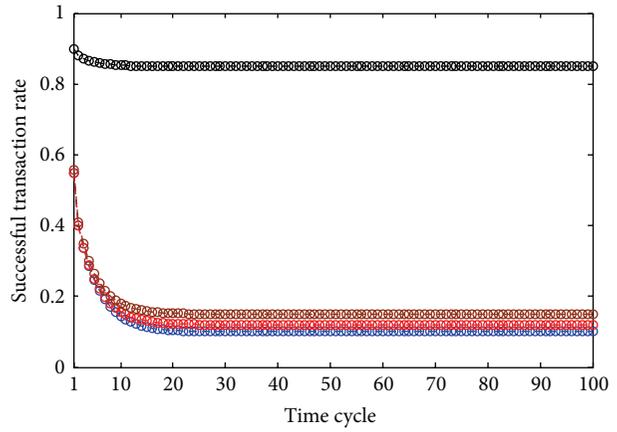
(a) 10% of malicious peers



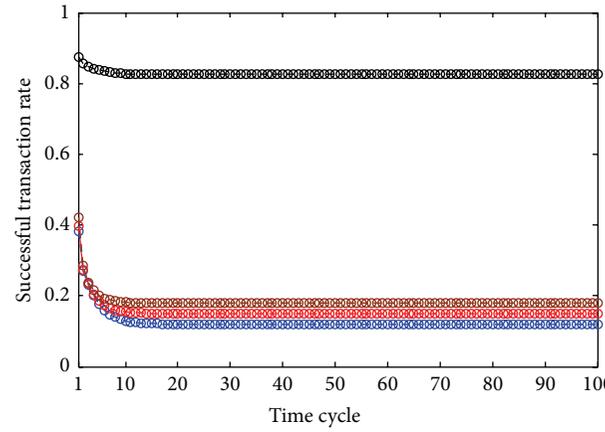
(b) 20% of malicious peers



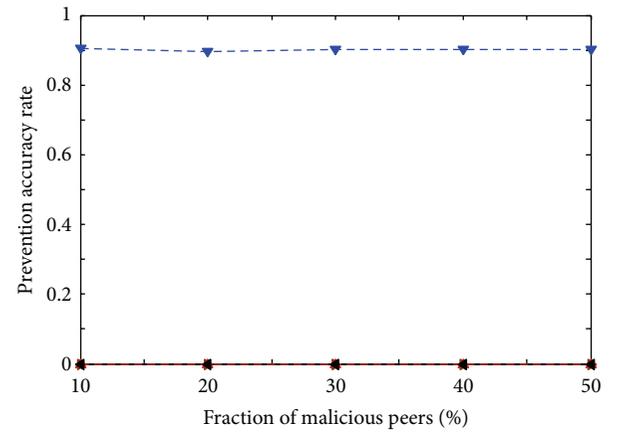
(c) 30% of malicious peers



(d) 40% of malicious peers



(e) 50% of malicious peers



(f) Prevention accuracy rate

FIGURE 7: Performances of IRTrust, EigenTrust, PeerTrust, and R2Trust in sybil attack (color online).

attacks with differentiating the service and recommendation, and the efficiency is stable and the highest.

5.4. Strategy Attack. Figure 6 depicts the simulations under the condition of strategic malicious peers. Figure 6(a) depicts the transaction rate of these four schemes. Although the EigenTrust scheme uses the global trust value to resist this type of attack, the demand of close cooperation of peers and time synchronization is higher. The more peers and more malicious ratio, the lower successful transaction ratio. The PeerTrust can effectively restrain the strategic attack, but the trust accuracy of PSM algorithm is affected by the width of the sliding window. Its performance is inferior to the R2Trust scheme and IRTrust scheme. Compared to PeerTrust scheme and R2Trust scheme, the proposed IRTrust scheme retains higher successful transaction rates. Even the fraction of malicious peers is 50% and the successful transaction rate maintains about 86%.

Figure 6(b) depicts the prevention accuracy rate of them. To this type of malicious attack, a good idea is to keep all the transaction records and the contexts. The EigenTrust scheme needs to keep closer cooperation of peers and higher time synchronization, which defends against this kind of attack. Thus, its prevention accuracy ratio is the least which decreases quickly with the malicious peers increasing. The PeerTrust scheme can effectively restrain the strategic attack by adopting the PSM algorithm, but the performance of PSM algorithm seriously relies on the width of the sliding window. As a result, this scheme is inferior to the R2Trust scheme and IRTrust scheme. Compared to PeerTrust scheme and R2Trust scheme, the proposed IRTrust scheme keeps higher and more stable prevention accuracy rates. Even the fraction of malicious peers is 50% and the prevention accuracy rate is about 90%.

5.5. Sybil Attack. Figure 7 shows the simulations under the condition of sybil malicious peers. In Figures 7(a)–7(e), these simulations are the successful transaction ratios in the different size of malicious peers. When the ratio of malicious nodes is small, at the initial stage, the transaction success rate is relatively high, but as the increase of time period, the malicious node will attack scope which will be larger and larger and the transaction success rate will rapidly decline. Because these three models (EigenTrust, PeerTrust, and R2Trust) can not distinguish sybil attacks, the more the malicious nodes, the lower the transaction success rate. The successful transaction ratio of the IRTrust scheme decreases with the increasing of the fraction of malicious peers, and it is about 85% when the fraction of malicious peers is 50%.

Figure 7(f) depicts the prevention accuracy ratio of these four schemes. A peer joins the system with multiple identities, and it provides inauthentic services or malicious attack to the others using one of the identities. When its reputation is reduced by the system, the peer can exit the system and rejoin the system with a new identity again. Although the EigenTrust, PeerTrust, and R2Trust schemes adopt the global trust value to restrain the malicious attack of peers, they can not address the problem of sybil attack. In IRTrust scheme, we

use identity-based authentication techniques, set the service threshold, and restrict the reputation accumulative ratio to put off and restrain the sybil attack. Therefore, our trust scheme can effectively prevent the peers' malicious behaviors, and the prevention accuracy ratio is stable and can keep about 90% in the simulations.

6. Conclusion

To encourage resource sharing among peers and resist malicious behaviors, trust management is essential for peers to assess the trustworthiness of others and to interact selectively with more reputable ones. In this paper, we fully consider the interpersonal relationships among peers and present an IRTrust scheme. To improve the resistance of malicious peers, the proposed scheme adopts the identity authentication, distinguishes the service and recommendation, and uses the comprehensive risk factor to prevent malicious attack and evaluate the trustworthiness of a peer. The IRTrust scheme is highly resistant to the malicious behavior attacks of peers, such as simple attacks, collusive attacks, strategic attacks, and sybil attacks. The simulation results prove that our trust scheme performs well even when the fraction of malicious peers in the system reaches 50%.

The study of secure network coding scheme suitable for P2P networks is a hot topic, and the scheme and analysis method of different attacks in this paper have potential application in designing secure P2P network coding.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper is supported by the AoE Grant E-02/08 from the University Grants Committee of the Hong Kong Special Administration Region, China, the Hong Kong Scholars Program (Grant no. HJ2012005), the China Postdoctoral Science Foundation Funded Project (Grant no. 2012T50209), the Shenzhen Key Laboratory of Network Coding Key Technology and Application (Grant no. ZSDY20120619151314964), the Peacock Scheme (Grant no. KQCX20130628164008004), the Fundamental Research Fund (Grant no. JCYJ2013401171935815), Shenzhen, China, and the National Natural Science Foundation of China (Grant nos. 61170269, 61121061).

References

- [1] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in p2p networks," in *Proceedings of the 12th International Conference on World Wide Web (WWW '03)*, pp. 640–651, May 2003.
- [2] L. Xiong and L. Liu, "PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 7, pp. 843–857, 2004.

- [3] Z. Liang and W. Shi, "PET: a personalized trust model with reputation and risk evaluation for P2P resource sharing," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, January 2005.
- [4] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, 2007.
- [5] R. Zhou and K. Hwang, "PowerTrust: a robust and scalable reputation system for trusted peer-to-peer computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 4, pp. 460–473, 2007.
- [6] K. Watanabe, Y. Nakajima, T. Enokido, and M. Takizawa, "Ranking factors in peer-to-peer overlay networks," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 2, no. 3, article no. 11, 2007.
- [7] F. Sato, "Estimation of trustworthiness for P2P systems in a collusive attack," *International Journal of Web and Grid Services*, vol. 4, no. 1, pp. 24–34, 2008.
- [8] K. Chen, K. Hwang, and G. Chen, "Heuristic discovery of role-based trust chains in peer-to-peer networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 1, pp. 83–96, 2009.
- [9] M. Jamali and M. Ester, "TrustWalker: a random walk model for combining trust-based and item-based recommendation," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 397–405, July 2009.
- [10] C. Tian and B. Yang, "R2Trust, a reputation and risk based trust management framework for large-scale, fully decentralized overlay networks," *Future Generation Computer Systems*, vol. 27, no. 8, pp. 1135–1141, 2011.
- [11] X. Li, F. Zhou, and X. Yang, "A multi-dimensional trust evaluation model for large-scale P2P computing," *Journal of Parallel and Distributed Computing*, vol. 71, no. 6, pp. 837–847, 2011.
- [12] G. Ciccarelli and R. Lo Cigno, "Collusion in peer-to-peer systems," *Computer Networks*, vol. 55, no. 15, pp. 3517–3532, 2011.
- [13] Z. Despotovic and K. Aberer, "P2P reputation management: probabilistic estimation vs. Social networks," *Computer Networks*, vol. 50, no. 4, pp. 485–500, 2006.
- [14] P. Dewan and P. Dasgupta, "P2P reputation management using distributed identities and decentralized recommendation chains," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 7, pp. 1000–1013, 2010.
- [15] A. Satsiou and L. Tassiulas, "Reputation-based resource allocation in P2P systems of rational users," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 4, pp. 466–479, 2010.
- [16] Y. Wang and A. Nakao, "Poisonedwater: an improved approach for accurate reputation ranking in P2P networks," *Future Generation Computer Systems*, vol. 26, no. 8, pp. 1317–1326, 2010.
- [17] A. Cheng and E. Friedman, "Sybilproof reputation mechanisms," in *Proceedings of the ACM SIGCOMM 3rd Workshop on the Economics of Peer-to-Peer Systems (P2PECON '05)*, pp. 128–132, August 2005.
- [18] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, "Sybilguard: defending against sybil attacks via social networks," in *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, 2006.
- [19] D. Quercia and S. Hailes, "Sybil attacks against mobile users: friends and foes to the rescue," in *Proceedings of the INFOCOM*, 2010.
- [20] A. Iamnitchi, M. Ripeanu, and I. Foster, "Small-world file-sharing communities," in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '04)*, vol. 2, pp. 952–963, March 2004.
- [21] E. Meshkova, J. Riihijärvi, M. Petrova, and P. Mähönen, "A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks," *Computer Networks*, vol. 52, no. 11, pp. 2097–2128, 2008.
- [22] M. Yang and Z. Fei, "A novel approach to improving search efficiency in unstructured peer-to-peer networks," *Journal of Parallel and Distributed Computing*, vol. 69, no. 11, pp. 877–884, 2009.
- [23] J. Liu, R. Sun, and K. Kwak, "Fair exchange signature schemes," *Science China Information Sciences*, vol. 53, no. 5, pp. 945–953, 2010.
- [24] T. Gamer, "Collaborative anomaly-based detection of large-scale internet attacks," *Computer Networks*, vol. 56, no. 1, pp. 169–185, 2012.
- [25] S. Marti and H. Garcia-Molina, "Taxonomy of trust: categorizing P2P reputation systems," *Computer Networks*, vol. 50, no. 4, pp. 472–484, 2006.
- [26] L. Mekouar, Y. Iraqi, and R. Boutaba, "Peer-to-peer's most wanted: malicious peers," *Computer Networks*, vol. 50, no. 4, pp. 545–562, 2006.
- [27] E. J. Anderson and T. D. H. Cau, "Implicit collusion and individual market power in electricity markets," *European Journal of Operational Research*, vol. 211, no. 2, pp. 403–414, 2011.
- [28] Y. Chen, W. S. Lin, and K. J. R. Liu, "Risk-distortion analysis for video collusion attacks: a mouse-and-cat game," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1798–1807, 2010.
- [29] J. H. Park and D. H. Lee, "Fully collusion-resistant traitor tracing scheme with shorter ciphertexts," *Designs, Codes and Cryptography*, vol. 60, no. 3, pp. 255–276, 2011.
- [30] Y. Zhu, L. Huang, W. Yang, and X. Yuan, "Efficient collusion-resisting secure sum protocol," *Chinese Journal of Electronics*, vol. 20, no. 3, pp. 407–413, 2011.

Research Article

Topological Embedding Feature Based Resource Allocation in Network Virtualization

Hongyan Cui,^{1,2} Shaohua Tang,^{1,2} Fangfang Sun,^{1,2} Yue Xu,^{1,2} and Xiaoli Yang^{1,2}

¹ Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China

² State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Hongyan Cui; yan555cui@163.com

Received 22 July 2014; Accepted 4 August 2014; Published 31 August 2014

Academic Editor: Haipeng Peng

Copyright © 2014 Hongyan Cui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Virtualization provides a powerful way to run multiple virtual networks on a shared substrate network, which needs accurate and efficient mathematical models. Virtual network embedding is a challenge in network virtualization. In this paper, considering the degree of convergence when mapping a virtual network onto substrate network, we propose a new embedding algorithm based on topology mapping convergence-degree. Convergence-degree means the adjacent degree of virtual network's nodes when they are mapped onto a substrate network. The contributions of our method are as below. Firstly, we map virtual nodes onto the substrate nodes with the maximum convergence-degree. The simulation results show that our proposed algorithm largely enhances the network utilization efficiency and decreases the complexity of the embedding problem. Secondly, we define the load balance rate to reflect the load balance of substrate links. The simulation results show our proposed algorithm achieves better load balance. Finally, based on the feature of star topology, we further improve our embedding algorithm and make it suitable for application in the star topology. The test result shows it gets better performance than previous works.

1. Introduction

The concept of network virtualization is put forward for the first time by Anderson et al. [1]. Researches in network virtualization involve many aspects [2–5]. Because of topological diversity of virtual networks, node and link resource constraints, online request, and access control, the embedding problem becomes the core problem of the network virtualization. In this paper, we mainly study the virtual network embedding problem.

In the environment of network virtualization, infrastructure providers (InPs) and service providers (SPs) play important roles [6, 7], which correspond to the substrate network (SN) and virtual network requests (VNs), respectively. Infrastructure providers focus on physical network construction and maintenance, while service providers concentrate on receiving users' virtual network requests and providing services. The main goal of virtual network embedding is

how to assign virtual network requests to a shared substrate network optimally with node and link resource constraints being satisfied.

Previous work does not consider the adjacent degree when the virtual nodes connecting directly to each other are mapped onto the substrate network. The result is that the virtual nodes connecting directly to each other may be mapped on the substrate network far away from (multiple hops) each other; thus, one virtual link occupies many substrate links, and the utilization efficiency of substrate resource is extremely low. In this paper, considering the degree of convergence when mapping virtual networks to the substrate network, we propose a new algorithm based on maximum topology mapping convergence-degree. Convergence-degree means the adjacent degree of virtual network's nodes when they are mapped onto the substrate network. The maximum convergence-degree not only ensures virtual nodes are mapped onto the substrate nodes with abundant resources,

but also ensures the virtual nodes connecting directly to each other are mapped onto the substrate nodes nearby, so the topology of a virtual network gathers together when the virtual network is mapped onto the substrate network. Therefore, the cost and complexity of link embedding reduced and the efficiency increases significantly; especially, the load balance of substrate links is improved.

The paper's main contributions are as follows.

- (1) Defining topology mapping convergence-degree, we realize a new algorithm. The algorithm largely decreases the complexity of embedding problem and improves the network utilization efficiency.
- (2) The proposed algorithm defines the balance rate of link load to reflect the load balance of substrate links and improves it largely. To the best of our knowledge, there are few articles researching it.
- (3) Based on the feature of star topology, we further propose a star topology embedding algorithm.

The remainder of this paper is organized as follows. In Section 2, we present the related work. Section 3 introduces the model and objective descriptions. Section 4 presents the proposed algorithm. In Section 5, we describe the performance simulation and analysis. Section 6 discusses star topology embedding algorithm. Section 7 gives conclusion.

2. Related Work

Because the virtual network embedding is an NP-hard problem [8], researchers have proposed many virtual network embedding algorithms through limiting problem space, which are mostly heuristic algorithms, such as ant swarm algorithm [9]. According to different control modes of virtual network embedding, virtual network embedding algorithms can be divided into centralized embedding algorithms and distributed embedding algorithms. Centralized embedding algorithms allocate resources by the central decision-making body, such as [10–14]. Distributed embedding algorithms are usually performed by the substrate nodes coordinately, such as [15]. According to the different embedding sequences of virtual nodes and virtual links, algorithms can be divided into first-order embedding algorithms and second-order embedding algorithms. In first-order embedding algorithms, node embedding and link embedding are completed at the same stage, such as [12–14], while node embedding and link embedding in second-order embedding algorithms are completed at the different stage, such as [10, 11, 15–18].

Yu et al. [11] present a second-order embedding algorithm supporting path splitting and migration. The algorithm can be divided into node embedding and link embedding stage. Firstly, authors use greedy way to embed all virtual nodes. Then, authors embed virtual links using K -shortest paths algorithm. Further, authors introduce path splitting and migration features and employ multicommodity flow algorithm in link embedding stage to improve algorithm's performance. Chowdhury et al. [10] propose an embedding algorithm with coordinated node and link mapping. The authors introduce the concept of metanode and transform

the node and link resource constraints problem into mixed integer programming problem. Each virtual node belongs to one metanode. Each metanode contains a subset of physical nodes. Then, virtual link embedding problem with bandwidth constraints can be regarded as commodity flow problem between metanodes. At last, the authors propose two kinds of algorithms: D-ViNE and R-ViNE, while a generalized window-based embedding algorithm (WiNE) is further discussed. Inspired by Markov random walk model, Cheng et al. [13] propose a measurement of node resource.

NodeRank. The node's NodeRank value not only reflects the resource of the node itself but also reflects other nodes' resources in substrate network topology. Based on the idea, the authors put forward two kinds of virtual network embedding algorithms: RW-MaxMatch and RW-BFS.

Although there have been a large number of previous researches, the problem of virtual network embedding still has a lot of research space; especially there are few references in the load balancing of substrate links and algorithm's efficiency. So in this paper, we take into account the load balancing of substrate links and algorithm's efficiency while considering the network utilization efficiency.

3. Embedding Model and Objective Descriptions

In this part, we will give the model and objective descriptions of virtual network embedding.

3.1. Substrate Network. The substrate network which InPs provide will be expressed as an undirected graph $G^s = (N^s, L^s, A_N^s, A_L^s)$, where N^s and L^s are sets of substrate node and link, respectively, and A_N^s and A_L^s are attribute sets of substrate node and link, respectively. The typical attributes of the node and link are the node's CPU capacity and the link's bandwidth.

3.2. Virtual Network Requests. The virtual network users, such as service providers and researchers, may need different virtual network requests. Similarly, each virtual network request will also be expressed as an undirected graph $G^v = (N^v, L^v, A_N^v, A_L^v)$, where N^v and L^v are sets of request node and link, respectively, and A_N^v and A_L^v are attribute sets of request node and link, respectively. Here, we also only consider the node's CPU requirement and the link's bandwidth requirement, which is estimated as long-range dependent traffic [19].

3.3. Embedding Model. The process of virtual network embedding M can be defined as finding a subgraph in undirected graph G^s for G^v satisfying the nodes' and links' attributes constraints (A_N^v, A_L^v) . This process is expressed as follows:

$$M : G^v(N^v, L^v, A_N^v, A_L^v) \longrightarrow G^s(N^s, L^s, A_N^s, A_L^s). \quad (1)$$

The process can be further divided into node embedding process M_N and link embedding process M_L :

$$\begin{aligned} M_N : G^v(N^v, A_N^v) &\longrightarrow G^s(N^s, A_N^s), \\ M_L : G^v(L^v, A_L^v) &\longrightarrow G^s(L^s, A_L^s). \end{aligned} \quad (2)$$

In the process of embedding, the substrate resources have to meet request's node and link resources requirements, namely,

$$\begin{aligned} N^v(i) \xrightarrow{M_N} N^s(j) \\ \sum_{i \in N^v} A_N^v(i) \leq A_N^s(j), \quad j \in N^s, \\ L^v(i) \xrightarrow{M_L} L^s(j) \\ \sum_{i \in L^v} A_L^v(i) \leq A_L^s(j), \quad j \in L^s, \end{aligned} \quad (3)$$

where $N^v(i)$ and $N^s(j)$ represent the i th node of the virtual network request and the j th node of substrate network, respectively, $L^v(i)$ and $L^s(j)$ represent the i th link of the virtual network request and the j th link of substrate network, respectively, $A_N^v(i)$ and $A_N^s(j)$ represent the required resource of the i th virtual node and the available resource of the j th substrate node, respectively, and $A_L^v(i)$ and $A_L^s(j)$ represent the required resource of the i th virtual link and the available resource of the j th substrate link, respectively.

In order to make our work more practical, we adopt the time window model in Figure 1. We assume that virtual network requests arrive in a Poisson process and each request's lifetime obeys the exponential distribution. In one time window, the requests will be mapped according to their revenues. If one virtual network is embedded unsuccessfully, the request will be postponed into the waiting queue temporarily for subsequent embedding, while the virtual networks which lifetimes come to an end will release their resources in the time window.

3.4. Objective Descriptions

3.4.1. R/C Ratio. The revenue is the reward when virtual networks are being embedded successfully. For a virtual network which has been embedded successfully, we define the revenue $R(G^v)$ as follows:

$$R(G^v) = \sum_{l^v \in L^v} \text{BW}(l^v) + \alpha \sum_{n^v \in N^v} \text{CPU}(n^v), \quad (4)$$

where BW is the required bandwidth of virtual link, the CPU is the required CPU of virtual node n^v , and α is a weight coefficient balancing the influence of CPU and BW to revenue.

Cost is the consumption of substrate network resources for finishing virtual network embedding. We define the cost $C(G^v)$ as follows:

$$C(G^v) = \sum_{l^v \in L^v} \sum_{l^s \in L^s} \text{BW}(F_{l^s}^{l^v}, l^v) + \beta \sum_{n^v \in N^v} \text{CPU}(n^v), \quad (5)$$

where $F_{l^s}^{l^v} \in \{0, 1\}$ and $F_{l^s}^{l^v} = 1$ if substrate link l^s allocates bandwidth to virtual link l^v ; otherwise $F_{l^s}^{l^v} = 0$. $\text{BW}(F_{l^s}^{l^v}, l^v)$ is the bandwidth which the substrate link l^s allocates to the virtual link l^v . Similarly, β is a weight coefficient balancing the influence of CPU and BW to cost.

When we have defined $R(G^v)$ and $C(G^v)$, revenue/cost ratio (R/C) can be defined as follows:

$$\begin{aligned} \frac{R}{C} &= \frac{R(G^v)}{C(G^v)} \\ &= \frac{\sum_{l^v \in L^v} \text{BW}(l^v) + \alpha \sum_{n^v \in N^v} \text{CPU}(n^v)}{\sum_{l^v \in L^v} \sum_{l^s \in L^s} \text{BW}(F_{l^s}^{l^v}, l^v) + \beta \sum_{n^v \in N^v} \text{CPU}(F_{n^s}^{n^v})}. \end{aligned} \quad (6)$$

R/C is directly related to net profit of InPs and reflects the resource utilization of substrate network. In this paper, we set $\alpha = \beta = 1$ but do not break generality.

3.4.2. The Balance Rate of Link Load. The usage of substrate links may overuse partly, while other parts are idle. Former researches rarely evaluate whether the usage of substrate links is balanced or not. In this paper, we define the standard deviation of link load as load balance rate to reflect the load balance of substrate links. The balance rate of link load has important practical significance for improving the reliability of the substrate network.

First, we define link load $L_l(t, l^s)$ as the stress of substrate link l^s in time t :

$$L_l(t, l^s) = \frac{\sum_{v^v \uparrow l^s} \text{BW}(l^v)}{\text{BW}(l^s)}, \quad (7)$$

where $\sum_{v^v \uparrow l^s} \text{BW}(l^v)$ means the occupied bandwidth by the virtual links which are mapped onto the substrate link l^s .

Then, we define the standard deviation of link load as balance rate of link load *LinkLoadRate*:

$$\text{LinkLoadRate} = \sqrt{\frac{1}{|L^s|} \sum_{l^s \in L^s} [L_l(t, l^s) - L_a(t)]^2}, \quad (8)$$

where $|L^s|$ is the number of links in the set of L^s . $L_a(t) = \sum_{l^s \in L^s} L_l(t, l^s) / |L^s|$ is the average stress of links L^s in time t .

3.4.3. RunTime. The average running time consumed by a time window can reflect the efficiency of an algorithm. We define *RunTime* as follows:

$$\text{RunTime} = \frac{\text{Total run time}}{\text{Number of time windows}}. \quad (9)$$

4. The Maximum Convergence-Degree Algorithm

In this section, we will give the motivation of our proposed algorithm. Then we present the algorithm in detail. Finally, the complexity of the algorithm is discussed briefly.

4.1. Motivation. In previous research work, such as [10, 11], each virtual node is mapped in isolation in node embedding

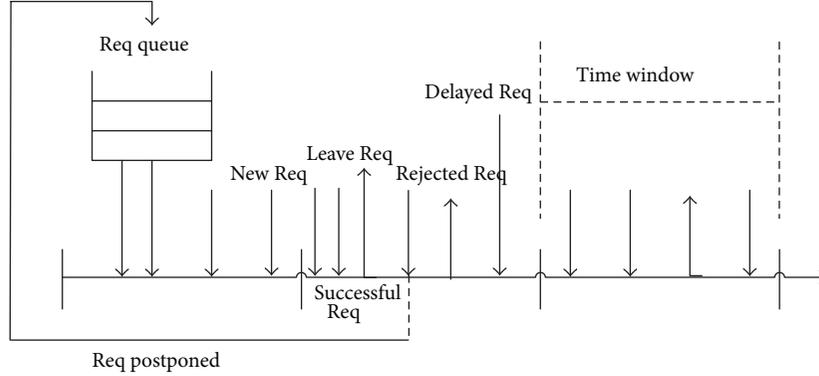


FIGURE 1: Model of time window.

stage. Although paper [13, 16] considers the connection relationship of substrate nodes, paper [16] only considers the substrate nodes which directly connect to the occupied substrate nodes, while the treatment for other substrate nodes is the same as paper [11]. Paper [13] defines the transfer probability matrix for all the virtual and substrate nodes, so the algorithm is fairly complex.

In order to overcome the above problems, we propose a new algorithm based on the maximum topology mapping convergence-degree. The proposed algorithm considers the topologies for both virtual network and substrate network. The maximum convergence-degree ensures the virtual nodes connecting directly to each other in a virtual network are embedded onto the substrate nodes nearby. The following example presents our algorithm's superiority.

As shown in Figure 2, (a) is the substrate network. Request1 and request2 represent two virtual network requests. The embedding result of the algorithm in paper [11] is (b). (c) is the embedding result by the algorithm proposed in paper [16]. The embedding result of our proposed algorithm is (d). For our proposed algorithm based on maximum topology mapping convergence-degree, when embedding request1, we choose the substrate node (5) to map the virtual node (c) after virtual nodes (a) and (b) being mapped onto the substrate nodes (1) and (6), because node (5) not only has the abundant resource (node's CPU capacity and link's bandwidth), but also the sum of the shortest paths between node (5) and nodes (1) and (6) is minimum (the sum of shortest paths is 2). The reason we consider the sum of the shortest paths between node (5) and nodes (1) and (6) is that node (c) connects directly to nodes (a) and (b) in the virtual network (if node (c) only connects directly to node (a), we will only consider the shortest path between the candidate substrate node and node (1)). This mapping way ensures that the virtual nodes connecting directly to each other in a virtual network are mapped onto the substrate nodes nearby, so the topology of a virtual network gathers together when the virtual network is mapped onto the substrate network. From the results (a) and (b), we can see the substrate resource is used inefficiently since one virtual link occupies many substrate links, while some substrate nodes and links are overloaded. From the intuitional view,

we can see the embedding result of our algorithm is more reasonable. We will prove the superiority of our proposed algorithm in Section 5.

4.2. Details of Maximum-Convergence-Degree Algorithm. In this part, we will introduce our algorithm in detail. Our proposed algorithm belongs to the second-order embedding algorithm, which can be divided into node embedding stage and link embedding stage. We employ the time window model described in Figure 1.

In node embedding stage, the virtual networks will be sorted according to their revenues in a time window. We select a virtual network with the maximum revenue to map. Then, we sort the nodes of the virtual network in descending order according to their required resources VR. VR is defined as follows:

$$VR(n^v) = CPU(n^v) \sum_{l^v \in \text{Neib}(n^v)} BW(l^v), \quad (10)$$

where $\text{Neib}(n^v)$ is the link set connecting directly to the virtual node n^v . Then we select a virtual node with the largest VR in this virtual network and embed it onto the substrate node. Considering the degree of convergence when embedding virtual networks onto the substrate network, the substrate node n^s we choose not only meets virtual node's CPU requirement, but also has the maximum convergence-degree CSR. CSR is defined as follows:

$$CSR(n^s) = \frac{CPU(n^s) \sum_{l^s \in \text{Neib}(n^s)} BW(l^s)}{N_{\text{path}} + \varepsilon}, \quad (11)$$

where $\text{Neib}(n^s)$ is the link set connecting directly to the substrate node n^s . $CPU(n^s)$ is the residual resource of node n^s . $BW(l^s)$ is the residual bandwidth of link l^s . N_{path} is the sum of the shortest paths between the candidate substrate node n^s and substrate nodes n^i ($i = 1, 2, \dots, k, k \neq s$), where n^i ($i = 1, 2, \dots, k, k \neq s$) are substrate nodes which have been assigned to the virtual network's nodes connecting directly to the node n^v . ε is an infinitesimal number in order to avoid the condition that N_{path} is equal to 0. $CPU(n^s) \sum_{l^s \in \text{Neib}(n^s)} BW(l^s)$ reflects the abundant degree of the node resource including

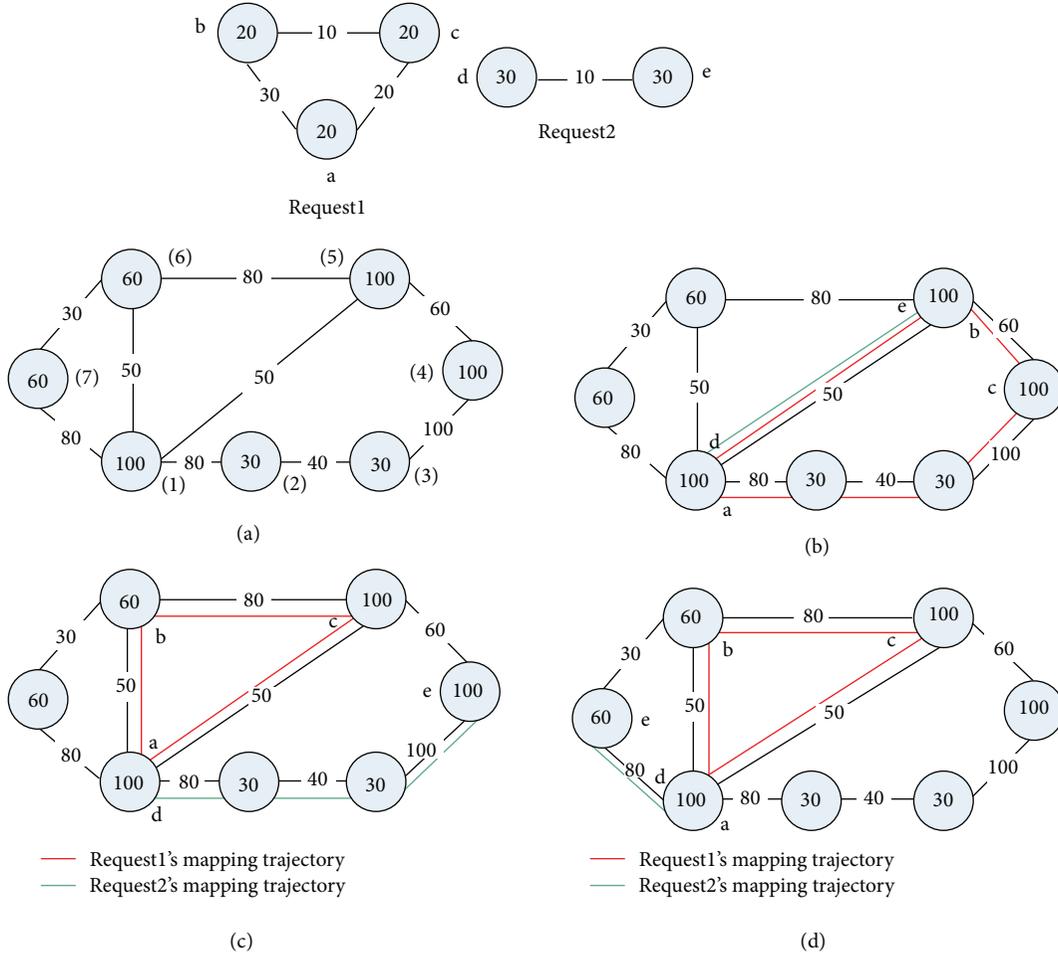


FIGURE 2: (a) Substrate network. (b) Embedding result of baseline algorithm proposed in paper [9, equation (11)]. (c) Embedding result of proximity algorithm proposed in paper [11, equation (16)]. (d) Embedding result of our proposed algorithm.

CPU and link resources. N_{path} reflects the adjacent degree of virtual nodes connecting directly to each other when they are mapped onto the substrate network; thus, CSR reflects the degree of convergence when mapping virtual nodes onto the substrate network. The bigger the value of CSR, the higher the degree of convergence.

The procedure of node embedding stage is executed in Algorithm 1. When all virtual nodes in this virtual network are embedded successfully according to Algorithm 1, we conduct link embedding. We use K -shortest paths algorithm when substrate links do not support path splitting, while we employ multicommodity flow algorithm when substrate links support path splitting. The link embedding stage is executed in Algorithm 2.

4.3. Discussion of Complexity. Compared with algorithms proposed in paper [11, 13, 16], our algorithm defines the topology mapping convergence-degree and just adds calculation factor $N_{path} + \epsilon$ in node embedding stage. Therefore, the increased complexity in node embedding stage can be ignored.

Because our algorithm considers the degree of convergence when mapping virtual network onto the substrate network, this mapping way ensures that the topology of a virtual network gathers together in substrate network when it is embedded successfully. So the complexity of link embedding stage is greatly reduced, while the running time of our proposed algorithm also reduced significantly. In Section 5, the simulation result will prove our algorithm's superiority.

5. Performance Evaluation

In this section, we will present simulation settings and compare our algorithm (maximum-convergence-degree algorithm) with baseline algorithm proposed in paper [11], proximity algorithm proposed in paper [16], and RW-MaxMatch algorithm proposed in paper [13] in the performance of revenue/cost ratio (R/C), balance rate of link load $LinkLoadRate$, and algorithm's running time $RunTime$. Finally we present the simulation results and analysis.

```

(1) Sort the virtual networks in descending order
according to their revenues  $R(G^v)$  in a time
window.
(2) for each virtual network with maximum revenue
do
sort the virtual nodes according to their required
resources VR in descending order
(3)   for each node with the largest VR do
assign the node onto the substrate node which not
only meets its' CPU requirement, but also has the
maximum convergence-degree CSR
(4)   end for
(5)   if all nodes are embedded successfully then
return NODE_MAPPING_SUCCESS
(6)   else
postpone the virtual network into the waiting
Queue
return NODE_MAPPING_FAILED
(7)   end if
(8) end for

```

ALGORITHM 1: Maximum-convergence-degree algorithm node embedding stage.

```

(1) sort the virtual networks mapped successfully
in the node stage according to their revenues
 $R(G^v)$  in descending order
(2) for each virtual network with maximum revenue
do
(3)   for each virtual link of the request do
(4)     if substrate network does not support
path splitting then
map virtual link using the  $k$ -shortest paths
algorithm
(5)     else
map virtual link using the multi-commodity flow
algorithm
(6)     end if
(7)   end for
(8)   if all links are embedded successfully then
return LINK_MAPPING_SUCCESS
(9)   else
postpone the virtual network into the waiting
queue
return LINK_MAPPING_FAILED
(10)  end if
(11) end for

```

ALGORITHM 2: Maximum-convergence-degree algorithm link embedding stage.

5.1. Simulation Environment. We generate substrate network topology adopting the GT-IMT tool [20]. The substrate network is set to have 100 nodes and about 570 links. The computing resources of nodes' CPU and the bandwidth of links are configured to real numbers uniformly distributed between 50 and 100.

We set 100 time units as one time window. The virtual networks arrive in a Poisson process with an average arrival rate λ per time window. Each virtual network's lifetime obeys

exponential distribution with an average of 10 time windows. For each virtual network, the number of virtual nodes follows a uniform distribution between 2 and 10. The connection probability among virtual nodes is 0.5. The required resources of virtual link's bandwidth BW and virtual node's CPU are uniform distribution between 0 and Max ($10 \leq \text{Max} \leq 100$); namely, $E(\text{BW}) = E(\text{CPU})$ is increased from 5 to 50, where $E(\cdot)$ is the mean. The DELAY which each virtual network waits for subsequent mapping is set to 2 time windows. For reaching a

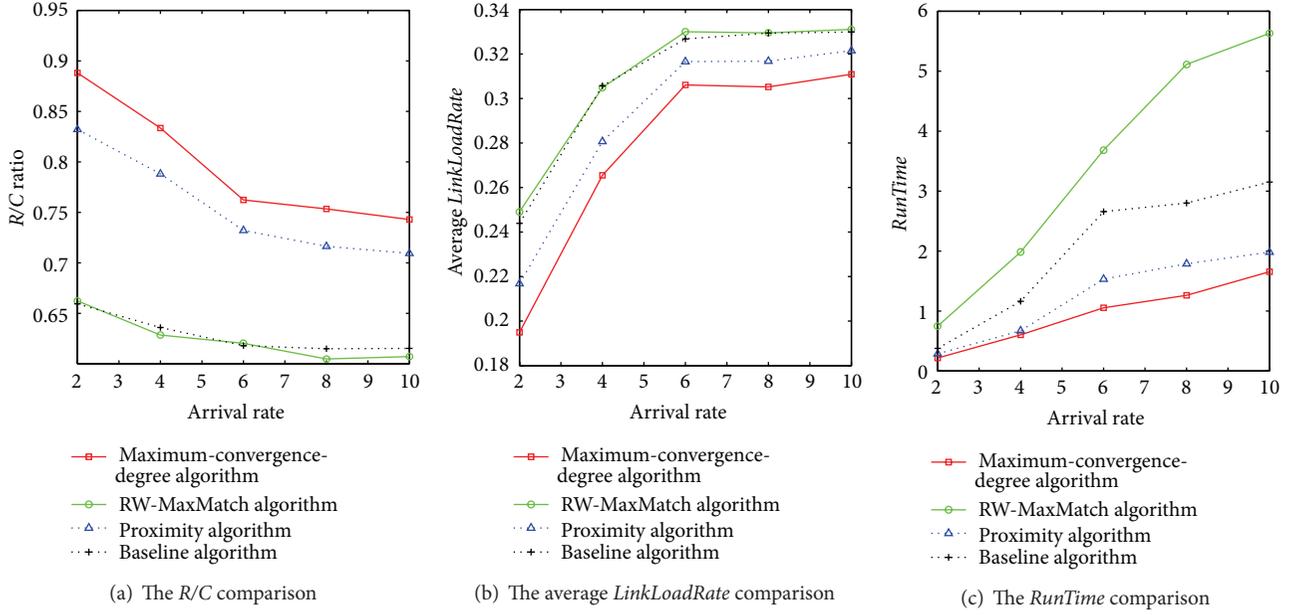


FIGURE 3: Performance comparison under increasing arrival rate λ . $E(BW) = E(CPU) = 25$. DELAY = 2. Splitting ratio is 0.

steady state, we simulate 500 time windows, corresponding to about 2500 virtual networks.

5.2. Simulation Results and Analysis

5.2.1. Maximum-Convergence-Degree Algorithm Produces Higher R/C Ratio. From Figures 3(a) and 4(a), we can see our proposed algorithm produces higher R/C under different VNs arrival rate and VNs required resources. Because we consider the topology mapping convergence-degree in the node embedding stage, it ensures the topology of a virtual network gathers together when the virtual network is mapped onto the substrate network. Each virtual link is mapped onto the substrate links with fewer hops; thus, the consumption of substrate bandwidth resources reduces greatly in link embedding stage, and R/C is improved greatly.

5.2.2. Maximum-Convergence-Degree Algorithm Makes Use of the Substrate Links More Balanced. From Figures 3(b) and 4(b), we can see that the load balance of substrate links in our proposed algorithm is always better than baseline algorithm, RW-MaxMatch algorithm, and proximity algorithm under different VNs arrival rate and VNs required resources. This indicates that our proposed algorithm uses the substrate more evenly and reasonably, and it will not show that parts of substrate links are idle, while other parts of substrate links are overused. The reason is when virtual nodes in a virtual network connecting directly to each other are mapped onto the substrate nodes nearby, the probabilities which one virtual link is mapped onto multiple substrate links and one substrate link is occupied by several virtual links will decrease, so it will not show the loads of some substrate links are lighter, while some parts are heavier. The load balance

of substrate links has important practical significance for improving the reliability of the substrate network.

5.2.3. Maximum-Convergence-Degree Algorithm Reduces the Running Time Greatly. From Figures 3(c) and 4(c), we can see our proposed algorithm reduces the running time greatly and improves the algorithm efficiency under different VNs arrival rate and VNs required resources. As stated above, because we consider the topology mapping convergence-degree in node embedding stage, it ensures the virtual nodes in a virtual network connecting directly to each other are mapped onto the substrate nodes nearby; thus, each virtual link is mapped onto the substrate links with fewer hops. On one hand, the running time of searching shortest paths for virtual links decreases, and on the other hand, with a higher success ratio in the link embedding stage, the running time is saved by avoiding backtracking. So the running time of the algorithm is reduced greatly, and the efficiency is improved greatly.

5.2.4. The Effect of Path Splitting Feature. Path splitting is a subsidiary feature of substrate network brought up by previous work [11]. It splits one virtual link into small pieces to assign more virtual networks to the substrate network. In order to demonstrate the superiority of our algorithm in different path splitting ratios, we also compare the algorithms performance with the splitting ratio increasing from 0 to 100%, while other parameters are set as follows: $E(BW) = E(CPU) = 25$. Arrival rate = 5. DELAY = 2.

As shown in Figure 5, our proposed algorithm is always better than baseline algorithm, proximity algorithm, and RW-MaxMatch algorithm in the performances of revenue/cost (R/C), balance rate of link load $LinkLoadRate$, and running time $RunTime$. Therefore, we can conclude that our proposed

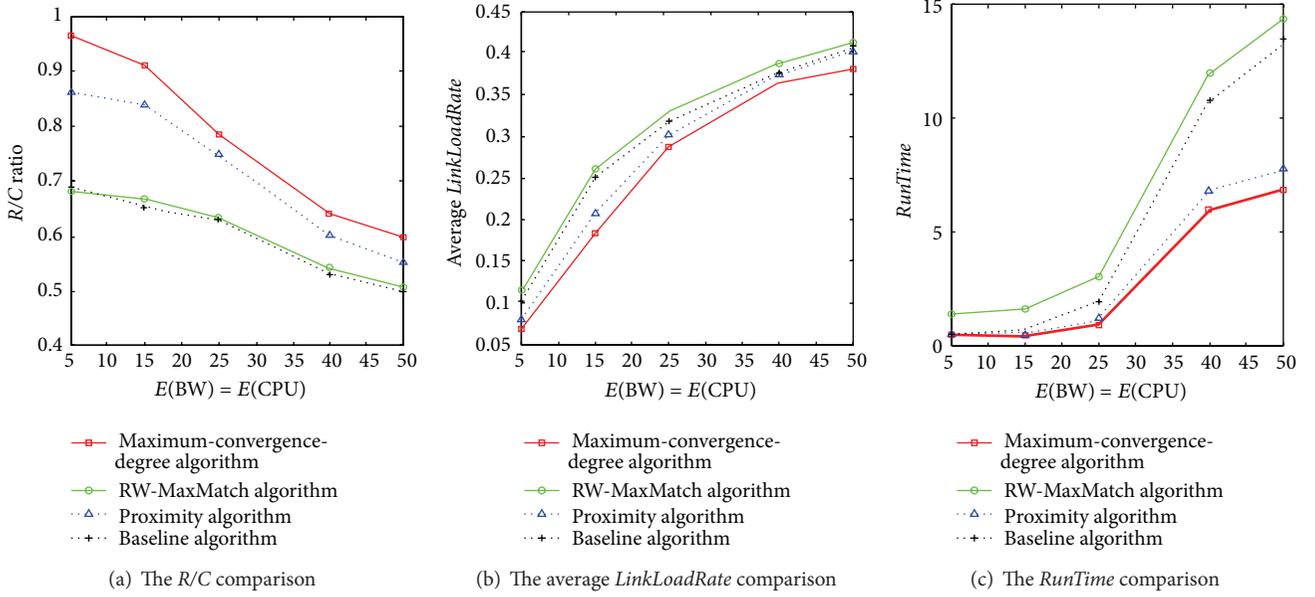


FIGURE 4: Performance comparison under increasing $E(BW) = E(CPU)$. Arrival rate = 5. DELAY = 2. Splitting ratio is 0.

algorithm is also suitable for the substrate network supporting any link splitting ratio.

6. The Star Topology Embedding Algorithm

There are many different kinds of network topologies, and common topological types are star, ring, tree, and mesh, and so forth. Different topologies are suitable for different types of applications. Among them, the star topology is the most commonly used, because it is easy to extend and convenient for management, such as Ethernet and CDN (content distribution network). We have proposed the embedding algorithm based on topology mapping convergence-degree in the above without considering topology feature. Of course, the proposed algorithm also applies to specific topology types. However, we believe there are more efficient embedding algorithms for specific topology types. Therefore, considering the importance of star topology, we further discuss star topology embedding algorithm on the basis of our previous work.

We know the feature of star topology type is that all nodes in the topology connect directly to a central node. Data transmission and exchange are controlled by the central node. The central node in the star topology is the most important node, so we think the central node should be mapped firstly in the process of mapping a star topology. Due to the fact that other nodes in star topology connect directly to the central node, if the central node is mapped onto a substrate node n^s with a large degree $d(n^s)$, the probability of the star topology being mapped successfully will increase, and the cost of link embedding stage will be reduced. Thus, the substrate node n^s we choose to map the central node not only

has enough available resources, but also has the maximum $DSR(n^s)$. $DSR(n^s)$ is defined as follows:

$$DSR(n^s) = CPU(n^s) d(n^s) \sum_{l^s \in Neib(n^s)} BW(l^s), \quad (12)$$

where $d(n^s)$ represents node's degree, namely, the number of links associated to the node n^s . After the central node is mapped successfully, the mapping methods of other virtual nodes and virtual links are the same as maximum-convergence-degree algorithm we have proposed.

In order to verify the effect of our improved star topology embedding algorithm, we compare the simulation results with previous works in the simulation environment of splitting ratio as 0, $E(BW) = E(CPU) = 25$, Arrival rate = 5, DELAY = 2, while other simulation parameters are the same as Section 5. The result is shown in Figure 6, where we can see the performances of star topology embedding algorithm are further improved. The result confirms that there are more efficient embedding algorithms for specific topology types, and we will research specific embedding algorithms according to the features of other different topology types in the future.

7. Conclusions

In this paper, we propose a new algorithm based on maximum topology mapping convergence-degree. The proposed algorithm ensures the virtual nodes connecting directly to each other in a virtual network are embedded onto the substrate nodes nearby; thus, the topology of a virtual network gathers together when the virtual network is mapped onto the substrate network, and the efficiency of link embedding

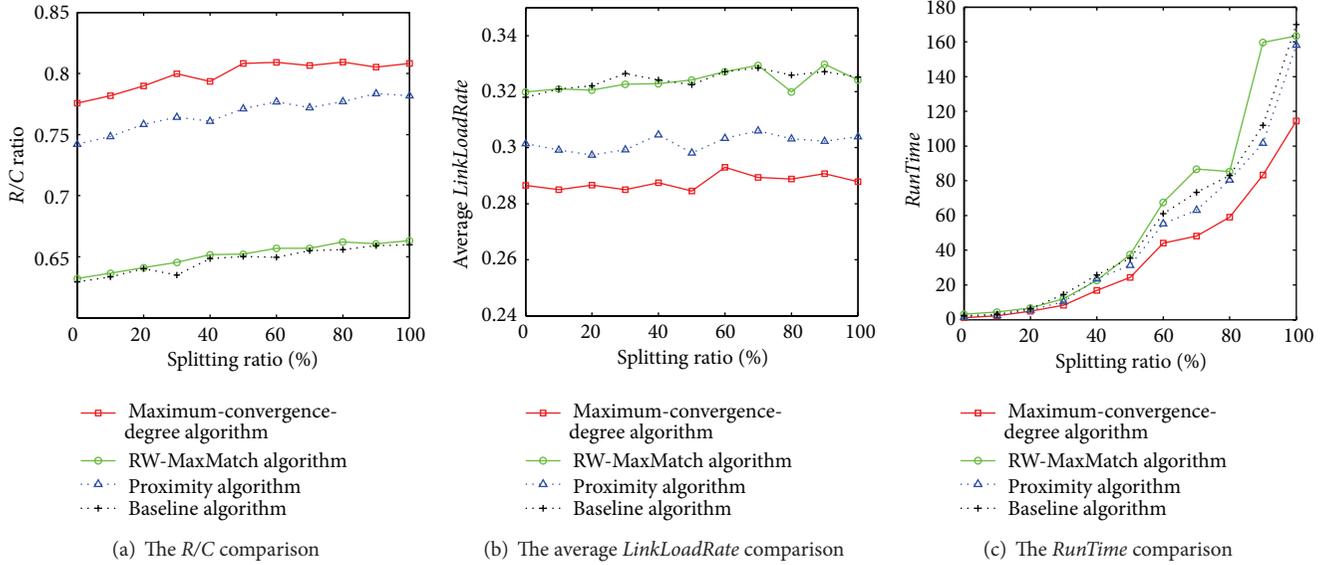


FIGURE 5: Performance comparison under different splitting ratios. $E(BW) = E(CPU) = 25$. Arrival rate = 5. DELAY = 2.

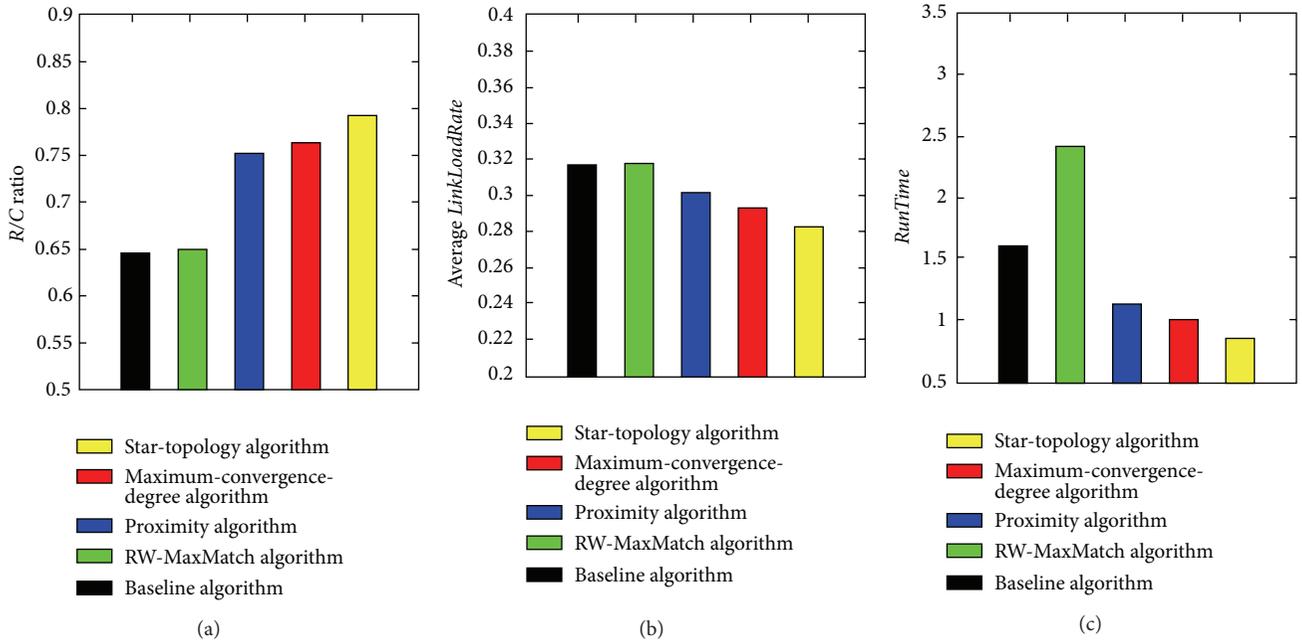


FIGURE 6: Performance comparison under $E(BW) = E(CPU) = 25$. Arrival rate = 5. DELAY = 2. Splitting ratio is 0.

increases significantly. The simulation results show, under a wide range of virtual networks and substrate network conditions, our proposed algorithm largely enhances the network utilization efficiency and decreases the complexity of the embedding problem; especially, the proposed algorithm improves the load balance of substrate links. Furthermore, based on the feature of star topology and its importance, we propose star topology embedding algorithm, and the performances of star topology embedding algorithm are further improved.

In our future research, we will further discuss how to realize the algorithm based on software defined network and how to optimize the method according to the actual restrictions such as networks node's position, reliability, and so forth.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61201153), the National 973 Program of China under Grant (2012CB315805), the Prospective Research Project on Future Networks in the Jiangsu Future Networks Innovation Institute (BY2013095-2-16), the National Basic Research Program 973 of China (2012CB315801), and the Fundamental Research Funds for the Central Universities (2013RC0113).

References

- [1] T. Anderson, L. Peterson, S. Shenker, and J. Turner, "Overcoming the internet impasse through virtualization," *IEEE Computer Magazine*, vol. 38, no. 4, pp. 34–41, 2005.
- [2] W. Yeow, C. Westphal, and U. Kozat, "Designing and embedding reliable virtual infrastructures," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 2, 2011.
- [3] Y. Zhou, Y. Li, G. Sun, D. Jin, L. Su, and L. Zeng, "Game theory based bandwidth allocation scheme for network virtualization," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–5, December 2010.
- [4] I. Houidi, W. Louati, D. Zeghlache, and S. Baucke, "Virtual resource description and clustering for virtual network discovery," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC '09)*, pp. 1–6, Dresden, Germany, June 2009.
- [5] A. Belbekkouche, M. M. Hasan, and A. Karmouch, "Resource discovery and allocation in network virtualization," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 4, pp. 1114–1128, 2012.
- [6] N. M. Mosharaf Kabir Chowdhury and R. Boutaba, "Network virtualization: state of the art and research challenges," *IEEE Communications Magazine*, vol. 47, no. 7, pp. 20–26, 2009.
- [7] N. M. M. K. Chowdhury and R. Boutaba, "A survey of network virtualization," *Computer Networks*, vol. 54, no. 5, pp. 862–876, 2010.
- [8] D. Andersen, "Theoretical approaches to node assignment," Working Paper, 2002.
- [9] L. Li, J. Kurths, H. Peng, Y. Yang, and Q. Luo, "Exponentially asymptotic synchronization of uncertain complex time-delay dynamical networks," *The European Physical Journal B*, vol. 86, no. 4, article 125, 9 pages, 2013.
- [10] M. Chowdhury, M. R. Rahman, and R. Boutaba, "ViNEYard: virtual network embedding algorithms with coordinated node and link mapping," *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 206–219, 2012.
- [11] M. Yu, Y. Yi, J. Rexford et al., "Rethinking virtual network embedding: substrate support for path splitting and migration," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 17–29, 2008.
- [12] J. Lischka and H. Karl, "A virtual network mapping algorithm based on subgraph isomorphism detection," in *Proceedings of the 1st ACM Workshop on Virtualized Infrastructure Systems and Architectures (VISA '09)*, pp. 81–88, 2009.
- [13] X. Cheng, S. Su, Z. Zhang et al., "Virtual network embedding through topology-aware node ranking," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 2, 2011.
- [14] J. Shamsi and M. Brockmeyer, "QoSMap: QoS aware mapping of virtual networks for resiliency and efficiency," in *Proceedings of the IEEE GLOBECOM Workshop*, pp. 1–6, Washington, DC, USA, November 2007.
- [15] I. Houidi, W. Louati, and D. Zeghlache, "A distributed virtual network mapping algorithm," in *Proceeding of the IEEE International Conference on Communications (ICC '08)*, pp. 5634–5640, Beijing, China, May 2008.
- [16] J. Liu, T. Huang, J. Chen, and Y. Liu, "A new algorithm based on the proximity principle for the virtual network embedding problem," *Journal of Zhejiang University: Science C*, vol. 12, no. 11, pp. 910–918, 2011.
- [17] H. Cui, W. Gao, Y. Liu et al., "A virtual network embedding algorithm based on virtual topology connection feature," in *Proceedings of the IEEE 16th International Symposium on Wireless Personal Multimedia Communications (WPMC '13)*, pp. 1–5, 2013.
- [18] X. Chen, Y. Luo, and J. Wang, "Virtual network embedding with border matching," in *Proceedings of the 4th International Conference on Communication Systems and Networks (COMSNETS '12)*, pp. 1–8, January 2012.
- [19] M. Li, W. Chen, and L. Han, "Correlation matching method for the weak stationarity test of LRD traffic," *Telecommunication Systems*, vol. 43, no. 3–4, pp. 181–195, 2010.
- [20] E. W. Zegura, K. L. Calvert, and S. Bhattacharjee, "How to model an internetwork," in *Proceedings of the 15th Annual Joint Conference of the IEEE Computer Societies (INFOCOM '96)*, vol. 2, pp. 594–602, San Francisco, Calif, USA, March 1996.