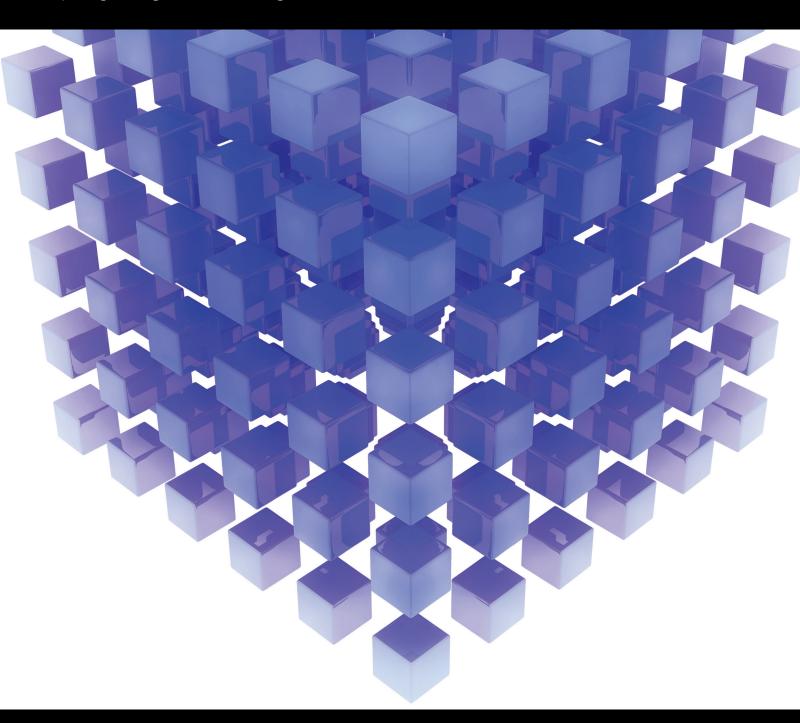
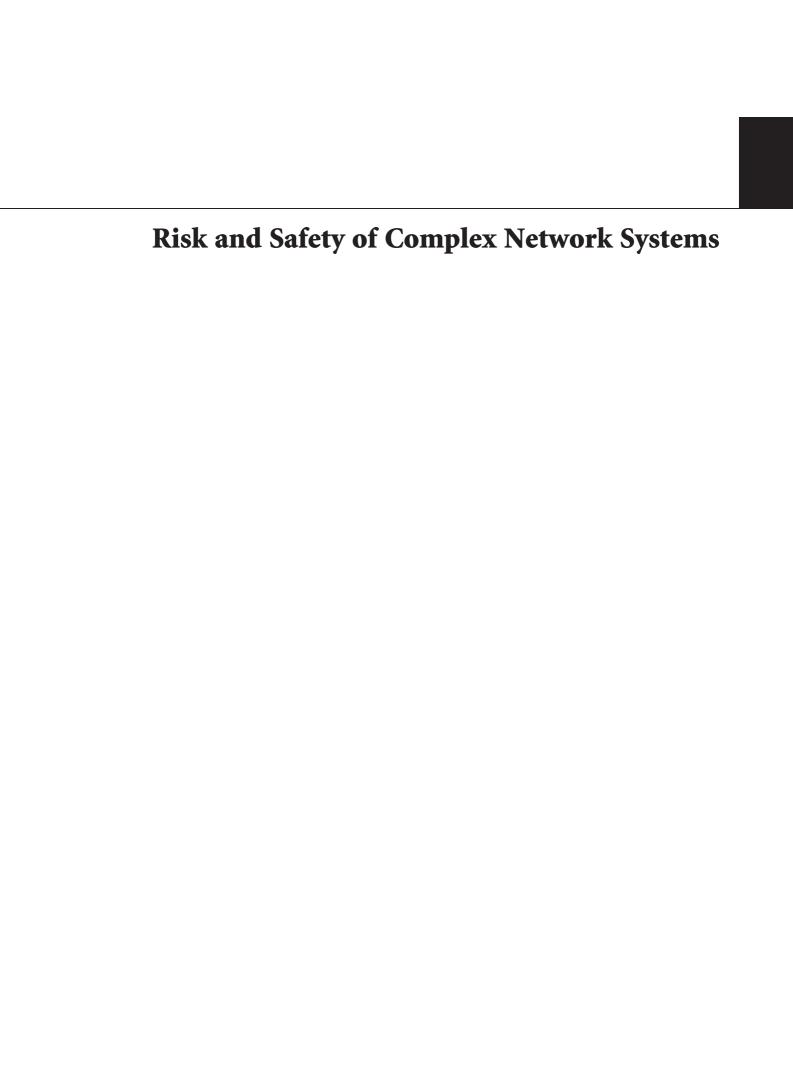
Risk and Safety of Complex Network Systems

Guest Editors: Xiao-Bing Hu, Adrian V. Gheorghe, Mark S. Leeson, Supeng Leng, Julien Bourgeois, and Xiaobo Qu





Risk and Safety of Complex Network Systems

Guest Editors: Xiao-Bing Hu, Adrian V. Gheorghe, Mark S. Leeson, Supeng Leng, Julien Bourgeois, and Xiaobo Qu



Editorial Board

Mohamed Abd El Aziz, Egypt Farid Abed-Meraim, France Silvia Abrahão, Spain Paolo Addesso, 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 Lionel Amodeo, France Sebastian Anita, Romania Renata Archetti, Italy Felice Arena, Italy Sabri Arik, Turkey Fumihiro 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 Maurizio Brocchini, Italy Julien Bruchon, France Javier Buldu', Spain Tito Busani, USA Pierfrancesco Cacciola, UK Salvatore Caddemi, Italy Jose E. Capilla, Spain Ana Carpio, Spain Miguel E. Cerrolaza, Spain M. Chadli, France Gregory Chagnon, France Ching-Ter Chang, Taiwan Michael J. Chappell, UK Kacem Chehdi, France Chunlin Chen, China Xinkai Chen, Japan 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'Acierno, Italy Weizhong Dai, USA Purushothaman 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 Ahmed El Hajjaji, France Fouad Erchiqui, Canada

Anders Eriksson, Sweden

Hua Fan, China Yann Favennec, France Giuseppe Fedele, Italy Roberto Fedele, Italy Jacques Ferland, Canada Jose R. Fernandez, Spain Simme D. Flapper, Netherlands Thierry Floquet, France Eric Florentin, France Francesco Franco, Italy Tomonari Furukawa, USA Mohamed Gadala, Canada Matteo Gaeta, Italy Zoran Gajic, USA Ugo Galvanetto, Italy Akemi Gálvez, Spain Rita Gamberini, Italy Maria Gandarias, Spain Arman Ganji, Canada Xin-Lin Gao, USA Zhong-Ke Gao, China Giovanni Garcea, Italy Fernando García, Spain Laura Gardini, Italy Alessandro Gasparetto, Italy Vincenzo Gattulli, Italy 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 M.I. Herreros, Spain Vincent Hilaire, France

Giovanni Falsone, Italy

Jaromir 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 Bin Jiang, China Zhongping Jiang, USA Ningde Jin, China Grand R. Joldes, Australia Tadeusz Kaczorek, Poland Tamas Kalmar-Nagy, Hungary Tomasz Kapitaniak, Poland Haranath Kar, India Konstantinos Karamanos, Belgium Chaudry Khalique, South Africa Do Wan Kim, Republic of Korea Nam-Il Kim, Republic of 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, UAE Peide Liu, China Peter Liu, Taiwan Wanquan Liu, Australia Yan-Jun Liu, China Jean J. Loiseau, France

Eckhard Hitzer, Japan

Paolo Lonetti, Italy Luis M. López-Ochoa, Spain Vassilios C. Loukopoulos, Greece Valentin Lychagin, Norway Fazal M. Mahomed, South Africa Yassir T. Makkawi, UK Noureddine Manamanni, France Didier Maquin, France Paolo Maria Mariano, Italy Benoit Marx, France Ge'rard 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 Javier Ortega-Garcia, Spain Alejandro Ortega-Moñux, Spain Naohisa Otsuka, Japan Erika Ottaviano, Italy Alkis S. Paipetis, Greece

Alessandro Palmeri, UK

Anna Pandolfi, Italy Elena Panteley, France Manuel Pastor, Spain Pubudu N. 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 Rosário 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 Kumbakonam Ramamani Rajagopal, USA Gianluca Ranzi, Australia Sivaguru Ravindran, USA Alessandro Reali, 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 Ferreira Santos, Denmark José A. Sanz-Herrera, Spain Nickolas S. Sapidis, Greece Evangelos J. Sapountzakis, Greece Andrey V. Savkin, Australia Valery Sbitney, Russia Thomas Schuster, Germany Mohammed Seaid, UK Lotfi Senhadji, France Joan Serra-Sagrista, Spain Leonid Shaikhet, Ukraine Hassan M. Shanechi, USA Sanjay K. Sharma, India Bo Shen, Germany Babak Shotorban, USA Zhan Shu, UK Dan Simon, Greece 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 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 Tsourdos, 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 Pandian Vasant, Malaysia Miguel E. Vázquez-Méndez, Spain

Josep Vehi, Spain K. Veluvolu, Republic of Korea Fons J. Verbeek, Netherlands Franck J. Vernerey, USA Georgios Veronis, USA Anna Vila, Spain Rafael J. Villanueva, Spain Uchechukwu E. Vincent, UK Mirko Viroli, Italy Michael Vynnycky, Sweden Junwu Wang, China Shuming Wang, Singapore Yan-Wu Wang, China Yongqi Wang, Germany Desheng D. Wu, Canada Yuqiang Wu, China Guangming Xie, China Xuejun Xie, China Gen Qi Xu, China Hang Xu, China Xinggang Yan, UK Luis J. Yebra, Spain Peng-Yeng Yin, Taiwan Ibrahim Zeid, USA Huaguang Zhang, China Qingling Zhang, China Jian Guo Zhou, UK Ouanxin Zhu, China Mustapha Zidi, France

Contents

Risk and Safety of Complex Network Systems

Xiao-Bing Hu, Adrian V. Gheorghe, Mark S. Leeson, Supeng Leng, Julien Bourgeois, and Xiaobo Qu Volume 2016, Article ID 8983915, 3 pages

Threshold for the Outbreak of Cascading Failures in Degree-Degree Uncorrelated Networks

Junbiao Liu, Xinyu Jin, Lurong Jiang, Yongxiang Xia, Bo Ouyang, Fang Dong, Yicong Lang, and Wenping Zhang

Volume 2015, Article ID 752893, 7 pages

Stability Analysis of Process Route Based on Weighted Network

Ying Xiang, Rong Mo, Zhiyong Chang, Hu Qiao, and Chunlei Li Volume 2015, Article ID 946281, 9 pages

Robust Satellite Scheduling Approach for Dynamic Emergency Tasks

Xuejun Zhai, Xiaonan Niu, Hong Tang, Lixin Wu, and Yonglin Shen Volume 2015, Article ID 482923, 20 pages

A CDT-Based Heuristic Zone Design Approach for Economic Census Investigators

Changixu Cheng, Xiaomei Song, Jing Yang, Xiatian Hu, Shi Shen, and Lijun Wang Volume 2015, Article ID 828374, 14 pages

Epidemic Spreading Characteristics and Immunity Measures Based on Complex Network with Contact Strength and Community Structure

Xueting Zhang, Bingfeng Ge, Qi Wang, Jiang Jiang, Hanlin You, and Yingwu Chen Volume 2015, Article ID 316092, 12 pages

An Efficient Top-k Query Processing with Result Integrity Verification in Two-Tiered Wireless Sensor Networks

Ruiliang He, Hua Dai, Geng Yang, Taochun Wang, and Jingjing Bao Volume 2015, Article ID 538482, 8 pages

A Comprehensive Sensitivity Analysis of a Data Center Network with Server Virtualization for Business Continuity

Tuan Anh Nguyen, Dugki Min, and Jong Sou Park Volume 2015, Article ID 521289, 20 pages

In-Plane Vibration Response of the Periodic Viaduct on Saturated Soil under Rayleigh Surface Wave

Hai-yan Ju and Ming-fu Fu

Volume 2015, Article ID 835460, 10 pages

Quantitative Risk Analysis of Offshore Fire and Explosion Based on the Analysis of Human and Organizational Factors

Yan Fu Wang, Yu Lian Li, Biao Zhang, Pei Na Yan, and Li Zhang Volume 2015, Article ID 537362, 10 pages

Improvement on Load-Induced Cascading Failure in Asymmetrical Interdependent Networks: Modeling and Analysis

Haiyan Han and Rennong Yang Volume 2015, Article ID 194568, 10 pages

A Secure Ciphertext Self-Destruction Scheme with Attribute-Based Encryption

Tonghao Yang, Junquan Li, and Bin Yu Volume 2015, Article ID 329626, 8 pages

Predicting Free Flow Speed and Crash Risk of Bicycle Traffic Flow Using Artificial Neural Network Models

Cheng Xu, Qiangwei Li, Zhaowei Qu, and Sheng Jin Volume 2015, Article ID 212050, 11 pages

Risk and Resilience Analysis of Complex Network Systems Considering Cascading Failure and Recovery Strategy Based on Coupled Map Lattices

Fuchun Ren, Tingdi Zhao, and Hongli Wang Volume 2015, Article ID 761818, 8 pages Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2016, Article ID 8983915, 3 pages http://dx.doi.org/10.1155/2016/8983915

Editorial

Risk and Safety of Complex Network Systems

Xiao-Bing Hu,¹ Adrian V. Gheorghe,² Mark S. Leeson,³ Supeng Leng,⁴ Julien Bourgeois,⁵ and Xiaobo Qu⁶

¹Beijing Normal University, Beijing 100875, China

Correspondence should be addressed to Xiao-Bing Hu; huxiaobing@bnu.edu.cn

Received 30 December 2015; Accepted 30 December 2015

Copyright © 2016 Xiao-Bing Hu 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.

Complex network systems cover most aspects of our daily life and are concerned with a wide range of communities, such as transportation, communication and information, energy systems, disaster and risk reduction and mitigation, finance, social networks and perception, and biological and medical systems, as well as academic researchers on theories of reliability, safety, risk, complexity, and networks [1-4]. Despite their enormous benefits to daily life, complex network systems also exhibit disadvantages, and one of the most challenging issues is the risk and safety of complex network systems. In a network system, the impact of a local hazard/fault/disturbance can easily spread out to the whole system due to domino effect, cascading effect, and/or ripple effect and eventually evolves into a large-scale disaster [5, 6]. The finding of "six degrees of separation" may partially illustrate how efficiently the impact of a local event can spread in a network [7]. The speedup of globalization process nowadays just makes the situation even worse [8].

The past two decades has witnessed too many examples of globally networked catastrophes. For instance, in the 2003 Northeast blackout in USA, a software bug in the alarm system at a control room of the FirstEnergy Corporation, located in Ohio, eventually caused a widespread power outage throughout parts of the Northeastern and Midwestern United States and the Canadian province of Ontario, affecting an estimated 10 million people in Ontario and 45 million people in eight U.S. states [9]. In the 2003 SARS plague, the first case of infection was speculated to be a farmer in Foshan County,

Guangzhou Province, China, then modern social network and advanced transportation system boosted the outbreak in 37 countries having claimed 774 lives worldwide [10]. In the 2008 global financial crisis, the bursting of the U.S. housing bubble ended up with the European sovereign-debt crisis and a global recession [11]. In the 2010 Eyjafjallajokull Volcano eruption in Iceland, despite the relatively small intensity and remote location, 20 countries across Western and Northern Europe closed their airspace to commercial jet traffic and it affected about 10 million travelers [12].

Dealing with the risk and safety of complex network systems has long been a common challenging task of concern to researchers and engineers from a wide range of communities. Although substantial relevant results have been reported in the literature, the means to assess the risk and to improve the safety of complex network systems still remains largely unresolved. For example, a most important finding on the structural robustness of networks reveals that a scale-free topology is more vulnerable to intended attacks to hub nodes (i.e., high-degree nodes), which suggests that such nodes are more important than low-degree nodes [1]. However, recent research indicates that, once a node oscillating function is introduced, low-degree nodes are actually more important than hub nodes, no matter what the intended attacks or random failures [13]. These conflicting results clearly remind us of the fact that we are far away from fairly good understanding of complex network systems. What makes things worse is that a real-world complex network system usually keeps changing

²Old Dominion University, Norfolk, VA, USA

³University of Warwick, Coventry CV4 7AL, UK

⁴University of Electronic Science and Technology of China, Chengdu 611731, China

⁵University of Franche-Comte, FEMTO-ST Institute, UMR CNRS 6174, Montbéliard, France

⁶Griffith University, Gold Coast, QLD 4222, Australia

and evolving, and therefore studying the past of the system might not necessarily generate a good/feasible solution for the present, but it could even become a source of risk in future [3]. As is pointed out in a contemporary Nature perspective paper [14], the world is currently faced with a diverse range of globally networked risks at a time when we are ill-equipped to understand or handle them.

Researchers and engineers in various communities are eager to see new progress to address the risk and safety of complex network systems. Besides the nature of hazard/fault/disturbance and the topology of system, many other factors such as heterogeneity in components, interaction functions and rules, amplifying effects, self-healing capability, and multiple feedback loops altogether play a crucial role in determining the system performance against hazard/fault/disturbance. The complexity roots in the combination of all these factors, resulting in nonlinear, dynamical, self-adaptive, and self-organizing system behavior, which is far beyond the capability of the existing methodologies. To study the risk and safety of complex network systems, there is an urgent demand for new interdisciplinary theories, models, methods, and experiments. This particularly needs us to fertilize a new thinking of globally networked risks and complex systems science [8, 14].

This special issue aims to serve as a platform for a wide range of communities to share their specialized knowledge and experiences for dealing with the risk and safety of various complex network systems. By bringing together both relevant theoretical works and case studies, the research community can be a source of inspiration and encouragement to promote and enrich research on the risk and safety of complex network systems. In this way, a little more light will hopefully be shed on the challenging task set-up by the perspective paper [14].

The accepted papers in this special issue cover many important topics on the risk and safety of complex network systems, including both theoretical studies and application-oriented researches.

Regarding mathematical theories of complex networks, F. Ren et al. reported a coupled map lattices based approach to analyze the risk of cascading process and recovering strategy in Watts-Strogatz (WS) small-world network and Barabási and Albert (BA) scale-free network, respectively; J. Liu et al. conducted some theoretical analyses on the threshold for the outbreak of the cascading failures in degree-degree uncorrelated networks, which are useful to improve network robustness given the limited capacity resource; H. Han and R. Yang investigated load-induced cascading failures in asymmetrical interdependent networks, discovering that network robustness is positively related to capacity, but negatively related to load.

In the area of transportation systems and civil engineering, C. Xu et al. developed some artificial neural network models to predict the free flow speed of bicycle traffic and then to analyze the associated crash risk, which are important for the evaluation, planning, and management of bicycle traffic safety; H. Ju and M. Fu focused on the vibration control of structures and conducted a deep investigation on the inplane vibration response of periodic viaduct on saturated soil under Rayleigh surface wave.

In the area of communication networks and information systems, R. He et al. proposed an efficient top-k query processing scheme with result integrity verification, named as ETQ-RIV, in two-tiered sensor networks, and the proposed scheme exhibits better performance than the relevant existing works in terms of both communication cost and query result redundancy rate; T. A. Nguyen et al. targeted at sensitivity assessment of availability for data center networks (DCNs), which plays a crucial role in design and management of cloud computing based businesses, and they presented a comprehensive availability modeling and sensitivity analysis of a DCell-based DCN with server virtualization for business continuity using stochastic reward nets.

In the area of cyber world security, T. Yang et al., based on the attribute-based encryption (ABE) and the distributed hash table (DHT) technologies, reported a secure ciphertext self-destruction scheme in order to improve the cloud storage security for handling expired data.

In the area of disaster reduction and emergence management, X. Zhai et al. proposed a robust satellite scheduling model to address a sequence of emergency tasks, in which both the benefit and robustness of schedule are simultaneously maximized in each stage; Y. F. Wang et al., based on hybrid causal logic model techniques, developed a quantitative risk analysis framework to assess the risk of offshore fire and explosion.

In the area of biological and medical systems, X. Zhang et al. studied epidemic spreading characteristics and immunity measures based on complex network with contact strength and community structure.

In the area of financial systems, C. Cheng et al. presented a heuristic multikernel growth approach based on constrained Delaunay triangulation in order to address a special zone design problem for economic census investigators; Y. Xiang et al. analyzed the production stability of process routes for complex parts based on weighted network, where a brittleness risk entropy, as an indicator of the vulnerability of nodes in a weighted network, was introduced to predict those easily failed subsystems in the entire network.

Acknowledgments

Being the guest editors, we would like to express our gratitude to all the authors who have contributed their work to this special issue.

> Xiao-Bing Hu Adrian V. Gheorghe Mark S. Leeson Supeng Leng Julien Bourgeois Xiaobo Qu

References

[1] R. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002

- [2] 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.
- [3] A. V. Gheorghe, System of Systems, InTech, 2012.
- [4] P. Ball, Why Society is a Complex Matter, Springer, Berlin, Germany, 2012.
- [5] X. Q. Huang, I. Vodenska, S. Havlin, and H. E. Stanley, "Cascading failures in bi-partite graphs: model for systemic risk propagation," *Scientific Reports*, vol. 3, article 1219, 2013.
- [6] X.-B. Hu, M. Wang, M. S. Leeson, E. L. Hines, and E. Di Paolo, "Deterministic ripple-spreading model for complex networks," *Physical Review E*, vol. 83, no. 4, Article ID 046123, 2011.
- [7] S. Milgram, "The small-world problem," *Psychology Today*, vol. 1, no. 1, pp. 61–67, 1967.
- [8] OECD, "Future global shocks—improving risk governance," OECD Reviews of Risk Management Policies, OECD, 2011.
- [9] E. J. Lerner, "What's wrong with the electric grid?" *The Industrial Physicist*, vol. 9, no. 5, pp. 8–13, 2003.
- [10] World Health Organization (WHO), Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003.
- [11] J. Crotty, "Structural causes of the global financial crisis: a critical assessment of the 'new financial architecture," *Cambridge Journal of Economics*, vol. 33, no. 4, pp. 563–580, 2009.
- [12] S. M. Wilkinson, S. Dunn, and S. Ma, "The vulnerability of the European air traffic network to spatial hazards," *Natural Hazards*, vol. 60, no. 3, pp. 1027–1036, 2012.
- [13] K. Morino, G. Tanaka, and K. Aihara, "Dynamical robustness in complex networks: the crucial role of low-degree nodes," *Scientific Reports*, vol. 2, article 232, pp. 1–6, 2012.
- [14] D. Helbing, "Globally networked risks and how to respond," *Nature*, vol. 497, no. 7447, pp. 51–59, 2013.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 752893, 7 pages http://dx.doi.org/10.1155/2015/752893

Research Article

Threshold for the Outbreak of Cascading Failures in Degree-Degree Uncorrelated Networks

Junbiao Liu, ¹ Xinyu Jin, ¹ Lurong Jiang, ² Yongxiang Xia, ¹ Bo Ouyang, ³ Fang Dong, ⁴ Yicong Lang, ¹ and Wenping Zhang ¹

¹College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

Correspondence should be addressed to Xinyu Jin; jinxy@zju.edu.cn

Received 26 June 2015; Accepted 26 November 2015

Academic Editor: Xiaobo Qu

Copyright © 2015 Junbiao Liu 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 complex networks, the size of the giant component formed by unfailed nodes is critically important for estimating the robustness of networks against cascading failures. In order to explore the critical moment of cascading failures break-out, we provide a cascade of overload failure model with local load sharing mechanism and then deduce the threshold of node capacity when the large-scale cascading failures happen and unfailed nodes in steady state cannot connect to each other to form a large connected subnetwork. We get the theoretical derivation of this threshold in degree-degree uncorrelated networks and validate the effectiveness of this method in simulation. This threshold provides us with a guidance to improve the network robustness under the premise of limited capacity resource when creating a network and assigning load. Therefore, this threshold is useful and important to analyze the robustness of networks. We believe that our research provides us with a guidance to improve the network robustness under the premise of limited capacity resource.

1. Introduction

Cascading failures are a sort of phenomena in which a random failure or intentional attack on one or a few nodes leads to serve chain reaction in the networks, which can cause collapse of a large fraction of nodes in the network. It is widely found in many real-world networks, such as power transmission [1, 2], communication [3], economic [4], and biological [5] networks. A real example of cascading failures is the well-known Northeast Blackout in 2003 [6]. In this case, the outage of a generator led to a serve chain reaction of power blackout, which affected approximately 50 million people in North America and caused financial losses of about \$6 billion.

In order to understand the essential mechanism of cascading failures, a number of cascading models have been proposed, such as betweenness-based model [7–10], sand-pile model [11, 12], and fiber-bundle model [13, 14]. In these models, a node fails when its load exceeds its capacity. The

failure of this node leads to the redistribution of load in network and can cause collapse of a large fraction of the network. Therefore, the cascading failure process of these models is highly dependent on the relation between load and capacity.

When the dynamical process of cascading failures terminates in steady state, the network breaks into several connected subnetworks formed by unfailed nodes. The size of the largest connected subnetwork (i.e., the giant component) can be used to measure the severity of cascading failures, which is critically important for estimating the robustness of networks [15, 16]. It is intuitive that when the capacity increases (or the load decreases), the size of cascading failures in network reduces, which is confirmed by all of those cascading failure models above. In fiber-bundle and sand-pile model, with the capacity under a critical value (or load above a critical value), the giant component disappears, or above this critical value the size of the giant component dramatically rises up [13, 17].

²School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China

³College of Electrical and Information Engineering, Hunan University, Changsha 410015, China

⁴School of Information and Electric Engineering, Zhejiang University City College, Hangzhou 310015, China

This critical value is a very important feature to measure the robustness of networks. However, to our knowledge, there is little research on quantitative analysis of the relation between node capacity and this threshold for the break-out of cascading failures at present.

In this paper, we provide a cascade of overload failure model with local load sharing mechanism and then explore the threshold of node capacity when the large-scale cascading failures happen and there does not exist a large connected subnetwork formed by unfailed nodes.

2. Model

Here we provide a cascade of overload failure model with local load sharing mechanism. In this model, the statuses of nodes are divided into two categories: the *unfailed* and *failed*. We assume that all nodes in networks are unfailed at the beginning. A node fails if its load exceeds its capacity (i.e., overload). When a node fails to work, this node is considered as transferring a fixed positive load Δ to each of its unfailed neighbors and being separated from the giant component [18–20]. When the cascading failures terminate, only those nodes in the giant component are supposed to work. It is natural to assume that the capacity C_{ν} of a node ν is proportional to its initial load L_{ν} [21, 22] as

$$C_{\nu} = (1 + \alpha) L_{\nu}, \tag{1}$$

where the constant α is the tolerance parameter. The initial load of each node is randomly distributed following a uniform distribution on the interval $[L_{\min}, L_{\max}]$. For simplicity, we set the lower bound of initial load $L_{\min}=0$ and the lower bound of initial load $L_{\max}=1$ [18, 19]. Then, for arbitrary node ν , the cumulative distribution function $\varphi(l)$ of L_{ν} is defined as

$$P\{L_{v} < l\} \triangleq \varphi(l) = \begin{cases} 0 & l \le 0 \\ l & 0 < l \le 1 \\ 1 & l > 1. \end{cases}$$
 (2)

The numerical process of cascading failures with local load sharing mechanism is summarized as follows.

- (1) Initialization. Generate a degree-degree uncorrelated network with N nodes. Assume all nodes are unfailed. The load of each node is uniformly distributed in $[L_{\min}, L_{\max}]$, where we let $L_{\min} = 0$ and $L_{\max} = 1$. The capacities of nodes are determined by (1).
- (2) Beginning. Choose very few nodes randomly in the network, and set them as failed.
- (3) Load Redistribution. In each round, each node which is failed in the last round transfers a fixed positive load Δ to each of its unfailed neighbors. An unfailed node turns out to be failed if it overloads in this round.
- (4) Halt. Repeat step (3) if there exist overloaded nodes in the network; otherwise, the process halts. Finally, only the unfailed nodes in the giant component are supposed to work.

3. Analysis

- 3.1. Description of Critical Conditions. When the cascading failure process ends, the unfailed nodes in the network form several connected subnetworks. The fraction of giant component, which is the relative size of the largest connected subnetwork, can be used as a measure of the network performance against cascading failures. Figure 1 shows the fraction G of the giant component as a function of tolerance parameter α in Erdős-Rényi (ER) random networks [23] and Barabási-Albert (BA) scale-free networks [24]. For each G with different average degree $\langle k \rangle$, there exists a critical tolerance parameter α_c , with α under which the giant component disappears and over which *G* increases dramatically and approaches 1 finally. When $\alpha < \alpha_c$, G approximately equals 0, which indicates that large-scale cascading failures occur and network breaks into extremely small clusters. There are two conditions to ensure that *G* approximately equals 0 when $\alpha < \alpha_c$:
 - (I) Large-scale cascading failures occur in the network; that is, the failed nodes connect to each other to form a large connected subnetwork.
 - (II) When the dynamical process of cascading failures terminates in steady state, there does not exist a large connected subnetwork formed by unfailed nodes.

In the rest of this section, we give a theoretical derivation of α_c with these two conditions above.

3.2. Large-Scale Cascading Failures Occur in the Network. First, let us consider condition (I) that large-scale cascading failures occur in the network. This condition equals the situation that the failed nodes connect to each other to form a large connected subnetwork. On average, each node fails and causes more than one of its neighbors to fail when large-scale cascading failures happen. Consider a node v of degree k_v and any of its neighbors w. The probability that node v fails and causes node w to fail is

$$P\{L_w + \Delta > (1 + \alpha)L_w\} = P\{L_w < \frac{\Delta}{\alpha}\} = \varphi(\frac{\Delta}{\alpha}),$$
 (3)

where L_w is the initial load of node w.

The cascading failures of nodes are a sort of site percolation process. Because of the locally tree-like approximation in percolation of degree-degree uncorrelated networks, the probability that node ν fails and causes m of its $k_{\nu}-1$ neighbors (subtract the node which causes node ν to fail) to fail is

$$\binom{k_{\nu}-1}{m}\varphi\left(\frac{\Delta}{\alpha}\right)^{m}\left(1-\varphi\left(\frac{\Delta}{\alpha}\right)\right)^{k_{\nu}-1-m}.$$
 (4)

Thus, on average, the node ν fails and causes its

$$\begin{split} &\sum_{m=1}^{k_{v}-1} m \binom{k_{v}-1}{m} \varphi \left(\frac{\Delta}{\alpha}\right)^{m} \left(1-\varphi \left(\frac{\Delta}{\alpha}\right)\right)^{k_{v}-1-m} \\ &= \sum_{m=1}^{k_{v}-1} m \binom{k_{v}-1}{m} \varphi \left(\frac{\Delta}{\alpha}\right)^{m} \left(1-\varphi \left(\frac{\Delta}{\alpha}\right)\right)^{k_{v}-1-m} \end{split}$$

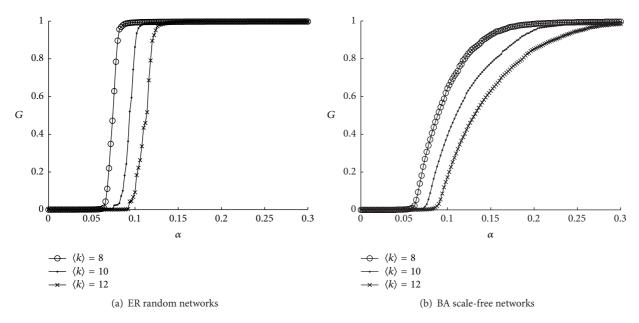


FIGURE 1: The fraction G of giant component changes with increase of tolerance parameter α in ER random and BA scale-free networks.

$$= \sum_{m=1}^{k_{\nu}-1} (k_{\nu} - 1) {k_{\nu} - 2 \choose m - 1} \varphi \left(\frac{\Delta}{\alpha}\right)^{m} \left(1 - \varphi\left(\frac{\Delta}{\alpha}\right)\right)^{k_{\nu}-1 - m}$$

$$= (k_{\nu} - 1) \varphi\left(\frac{\Delta}{\alpha}\right)$$
(5)

neighbors to fail.

Denote p_k as the probability that a randomly picked node has degree k. According to (5), the failure of arbitrary node in the network causes its $\sum_{k=0}^{\infty} p_k(k-1)\varphi(\Delta/\alpha) = (\langle k \rangle - 1)\varphi(\Delta/\alpha)$ neighbors to fail. When large-scale cascading failures occur in the network, the failure of arbitrary node causes more than one of its neighbors to fail; that is,

$$(\langle k \rangle - 1) \varphi\left(\frac{\Delta}{\alpha}\right) \ge 1.$$
 (6)

Thus, for condition (I) that large-scale cascading failures occur in the network, we have

$$\alpha \le (\langle k \rangle - 1) \Delta. \tag{7}$$

It is worth noting that there is not any specific assumption on degree distribution for the result above.

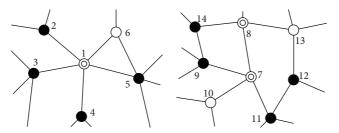
3.3. There Does Not Exist a Large Connected Subnetwork Formed by Unfailed Nodes. According to the local load sharing mechanism in our model, there may exist some nodes that do not propagate cascading failures. These nodes never fail with large tolerance parameter α , or their failure will not affect other nodes as all of their neighbors are already failed nodes. We call these nodes absorbing nodes. Before handling condition (II), we pay close attention to these absorbing nodes. Let us consider the situation that a node ν of degree k_{ν} does not fail after $k_{\nu}-1$ of its neighbors fail and share loads $(k_{\nu}-1)\Delta$ with it. Consequently, no matter whether node ν fails

or not, it can not affect other nodes any more. In this situation, node ν has the ability to absorb the loads of all its neighbors by itself; thus we call this node the *independent absorbing node*. The capacity and initial load of an independent absorbing node ν meet the following relation:

$$L_{v} + (k_{v} - 1) \Delta < (1 + \alpha) L_{v}.$$
 (8)

But this is not the only reason that a node happens to be an absorbing node. There is another situation for a node to be an absorbing node whose load does not satisfy (8). In this situation, considering a node v, m of its neighbors are absorbing nodes; thus we only need $n_v > k_v - m - 1$ to make node v an absorbing node. We call this kind of nodes dependent absorbing nodes because their ability to absorb loads depends on the other absorbing nodes of their neighbors. Dependent absorbing nodes can also prevent cascading failures locally.

Figure 2 gives two examples of absorbing nodes. In Figure 2(a), unfailed node 1 has five neighbors and four of them (nodes 2, 3, 4, and 5) are failed nodes. Assume that if node 6 does not fail in the following procedure of cascading failure, then node I does not fail either. Or in another case assume that node 6 fails and transfers a fixed value of load Δ to node 1. In the latter case, no matter whether node 1 fails or not, it can not affect any other node. Thus, node 1 is an independent absorbing node. In Figure 2(b), we assume that node 7 is not an independent absorbing node. It happens to fail if more than two of its four neighbors fail and transfer loads to it. Assume node 8 is an absorbing node; thus it never fails. Consequently, no matter whether node 7 fails or not, it can not share load with any unfailed neighbor. Therefore, node 7 in Figure 2(b) comes to be a dependent absorbing node because its ability to absorb loads depends on the absorbing node



(a) The independent absorbing (b) The dependent absorbing nodes nodes

FIGURE 2: Two examples of the absorbing nodes. In the procedure of cascading failures, an unfailed node (o) may turn to be a failed node (o) when it overloads or does not fail as it is an absorbing node (o).

Then we consider condition (II) that there does not exist a large connected subnetwork formed by unfailed nodes. We have explained that absorbing nodes survive and keep to be unfailed after cascading failures. With increase of α , each node has higher probability to be assigned with more capacity and be an absorbing node on average. If there exist a large number of absorbing nodes, they may connect to each other to form a large connected subnetwork which makes G greater than 0. We have divided absorbing nodes into independent absorbing nodes and dependent absorbing nodes. Furthermore, we call an absorbing node an *m*-absorbing node (m = 0, 1, 2, ..., k), if and only if m of its neighbors are absorbing nodes. We assume that independent absorbing nodes are not equivalent to 0-absorbing node. For an interdependent node, if it has m absorbing nodes of its neighbors, we still treat this node as an m-absorbing node.

Then, we let the probability that a node of degree k happens to be an absorbing node be a_k . The probability that node v happens to be an m-absorbing node is $a_k^{(m)}$. Then we have

$$a_k = \sum_{m=0}^k a_k^{(m)}. (9)$$

In a degree-degree uncorrelated network, consider a node v of degree k and any of its neighbors w. The probability of node w with degree j is $jp_j/\sum_i ip_i = jp_j/\langle k \rangle$. Node w happens to be an absorbing node with the probability a_j .

Then, the probability of node *v*'s arbitrary neighbor being an absorbing node is

$$\sigma_a = \sum_j \frac{j p_j a_j}{\langle k \rangle}.$$
 (10)

According to (10), the probability that exactly m of node v's neighbors are absorbing nodes is

$$\binom{k}{m}\sigma_a^m \left(1-\sigma_a\right)^{k-m}.\tag{11}$$

Node ν is an m-absorbing node if it stays unfailed with k-m-1 neighbors fail and transfer loads to it. Therefore, the probability is that node ν comes to be an m-absorbing node with probability

$$a_k^{(m)} = {k \choose m} \sigma_a^m \left(1 - \sigma_a\right)^{k-m}$$

$$\cdot P\left\{L_v + (k - m - 1) \Delta < (1 + \alpha) L_v\right\}$$

$$= {k \choose m} \sigma_a^m \left(1 - \sigma_a\right)^{k-m} \left(1 - \varphi\left(\frac{(k - m - 1) \Delta}{\alpha}\right)\right).$$
(12)

We substitute (12) into (9) and get the probability that node ν (i.e., arbitrary node of degree k) happens to be an absorbing node as

$$a_{k} = \sigma_{a}^{k} + \sum_{m=0}^{k-1} {k \choose m} \sigma_{a}^{m} \left(1 - \sigma_{a}\right)^{k-m} \left(1 - \varphi\left(\frac{(k-m-1)\Delta}{\alpha}\right)\right)$$

$$= \begin{cases} \sigma_{a}^{k} + \sum_{m=0}^{k-1} {k \choose m} \sigma_{a}^{m} \left(1 - \sigma_{a}\right)^{k-m} \left(1 - \frac{(k-m-1)\Delta}{\alpha}\right) & \text{if } k \leq \left[\frac{\alpha}{\Delta}\right] + 1\\ \sigma_{a}^{k} + \sum_{m=k-1-\left[\alpha/\Delta\right]}^{k-1} {k \choose m} \sigma_{a}^{m} \left(1 - \sigma_{a}\right)^{k-m} \left(1 - \frac{(k-m-1)\Delta}{\alpha}\right) & \text{elsewhere} \end{cases}$$

$$= \begin{cases} 1 - \frac{(k-1)\Delta}{\alpha} + \frac{k\sigma_{a}\Delta}{\alpha} - \frac{\sigma_{a}^{k}\Delta}{\alpha} & \text{if } k \leq \left[\frac{\alpha}{\Delta}\right] + 1\\ \sigma_{a}^{k} + \sum_{m=k-1-\left[\alpha/\Delta\right]}^{k-1} {k \choose m} \sigma_{a}^{m} \left(1 - \sigma_{a}\right)^{k-m} \left(1 - \frac{(k-m-1)\Delta}{\alpha}\right) & \text{elsewhere.} \end{cases}$$

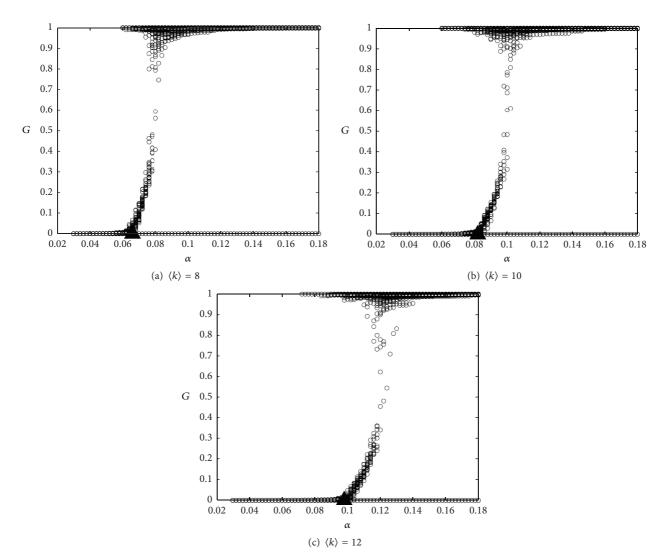


FIGURE 3: Validation of the analytical predictions of threshold α_c where the giant component G exists in ER random networks. With increase of tolerance parameter α , we draw each simulated G of 50 realizations, respectively. The threshold α_c is marked by solid triangle (\triangle).

According to the *cascade condition* [25, 26], the absorbing nodes can not connect to each other to form a large connected subnetwork with the following condition:

$$\frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k (k-1) p_k a_k < 1, \tag{14}$$

where p_k is probability that a randomly picked node has degree k and a_k is the probability that a node with degree k happens to be an absorbing node.

Now we complete the derivation by (7), (10), (13), and (14). These four equations depend on each other. We can get the threshold of tolerance parameter α_c with the following iteration process:

- (1) Assign $(\langle k \rangle 1)\Delta$ as an initial value to tolerance parameter α according to (7).
- (2) Get a_k via (10) and (13) with the current value of tolerance parameter α .

(3) Substitute a_k into (14). If the inequality of (14) is invalid, discount a very small value from α , and then repeat step (2); otherwise, let $\alpha_c = \alpha$ and finish the iteration process.

4. Simulation

We validate the analytical prediction of threshold α_c in ER random and BA scale-free networks with N=5000 nodes and varied average degree $\langle k \rangle = 8, 10, 12$. The load of each node is uniformly distributed on the interval [0,1]. We randomly set very few nodes to fail as the beginning of the dynamical procedure. When a node fails, it transfers a fixed load $\Delta=0.01$ to each of its unfailed neighbors. Figures 3 and 4 present the giant component G with different G0, and the threshold G1 for G2 rises from zero to nonzero. Note that different values of G2 only change the scale of abscissa. For each subfigure in Figures 3 and 4, each of 50 independent simulation results is drawn, respectively. The theoretical

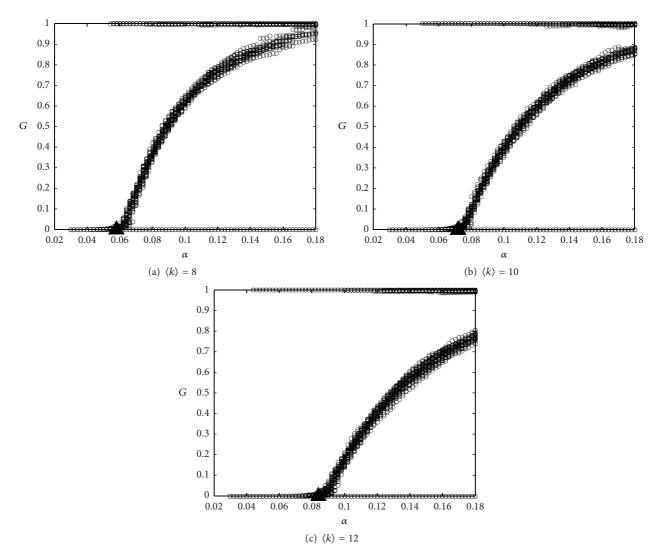


FIGURE 4: Validation of the analytical predictions of threshold α_c where the giant component G exists in BA scale-free networks. With increase of tolerance parameter α , we draw each simulated G of 50 realizations, respectively. The threshold α_c is marked by solid triangle (\triangle).

threshold α_c is marked by the solid triangle. α_c is exactly at the moment when G rises from zero to nonzero. Therefore, our theoretical predictions are in good agreement with the simulation results. It also can be seen in both ER and BA networks that, with higher value of average degree $\langle k \rangle$, the value of threshold α_c increases and the networks turn out to be robuster against cascading failure. This appearance is consistent with previous studies on cascading failure.

5. Conclusion

In this paper, we provide a cascade of overload failure model with local load sharing mechanism and then explore the threshold α_c of tolerance parameter for capacity, when the giant component G formed by unfailed nodes rises from zero to nonzero. We provide two conditions to ensure that G approximately equals zero when $\alpha < \alpha_c$, which are as follows: (I) the large-scale cascading failures occur; (II) unfailed nodes in steady state cannot connect to each other to form

a large connected subnetwork. With these two conditions, we get the theoretical derivation of α_c in degree-degree uncorrelated networks and validate the effectiveness of this theoretical derivation in simulations. We believe that when creating a network and assigning load with local load sharing mechanism, threshold α_c provides us with a guidance to improve the network robustness under the premise of limited capacity resource, especially for telecommunication networks and power grid networks.

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 Natural Science Foundations of China (NSFC61403347), National Science and

Technology Major Projects (2013ZX03005013), and opening foundation of the State Key Laboratory for Diagnosis and Treatment of Infectious Disease, The First Affiliated Hospital of Medical College, Zhejiang University (2014KF06).

References

- [1] Y. Koç, M. Warnier, P. Van Mieghem, R. E. Kooij, and F. M. Brazier, "The impact of the topology on cascading failures in a power grid model," *Physica A: Statistical Mechanics and Its Applications*, vol. 402, pp. 169–179, 2014.
- [2] P. Henneaux, "Probability of failure of overloaded lines in cascading failures," *International Journal of Electrical Power & Energy Systems*, vol. 73, pp. 141–148, 2015.
- [3] X.-L. Xu, W.-B. Du, and C. Hong, "Cascade defense via routing in complex networks," *International Journal of Modern Physics C*, vol. 26, no. 12, Article ID 1550141, 2015.
- [4] X. Huang, I. Vodenska, S. Havlin, and H. Eugene Stanley, "Cascading failures in bipartite graphs: Model for systemic risk propagation," *Scientific Reports*, vol. 3, article 1219, 2013.
- [5] C. Borrvall, B. Ebenman, and T. Jonsson, "Biodiversity lessens the risk of cascading extinction in model food webs," *Ecology Letters*, vol. 3, no. 2, pp. 131–136, 2000.
- [6] G. Andersson, P. Donalek, R. Farmer et al., "Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1922– 1928, 2005.
- [7] G. Barach, M. Tuchman, G. Cwilich, and S. Buldyrev, "Distributions of betweenness in cascades of overload failure in random regular networks," in *Proceedings of the APS March Meeting*, vol. 1 of *Abstract 16012*, Denver, Colo, USA, March 2014.
- [8] J. Wang, L. Rong, L. Zhang, and Z. Zhang, "Attack vulnerability of scale-free networks due to cascading failures," *Physica A*, vol. 387, no. 26, pp. 6671–6678, 2008.
- [9] F. Tan, Y. Xia, W. Zhang, and X. Jin, "Cascading failures of loads in interconnected networks under intentional attack," *EPL*, vol. 102, no. 2, Article ID 28009, 2013.
- [10] Y. Xia, J. Fan, and D. Hill, "Cascading failure in Watts-Strogatz small-world networks," *Physica A*, vol. 389, no. 6, pp. 1281–1285, 2010.
- [11] K.-I. Goh, D.-S. Lee, B. Kahng, and D. Kim, "Sandpile on scale-free networks," *Physical Review Letters*, vol. 91, no. 14, Article ID 148701, 2003.
- [12] M. Hoore and S. Moghimi-Araghi, "Critical behavior of a small-world sandpile model," *Journal of Physics. A. Mathematical and Theoretical*, vol. 46, no. 19, pp. 195001–195008, 2013.
- [13] Y. Moreno, J. B. Gómez, and A. F. Pacheco, "Instability of scale-free networks under node-breaking avalanches," *Europhysics Letters*, vol. 58, no. 4, pp. 630–636, 2002.
- [14] D.-H. Kim, B. J. Kim, and H. Jeong, "Universality class of the fiber bundle model on complex networks," *Physical Review Letters*, vol. 94, no. 2, Article ID 025501, 2005.
- [15] M. Á. Serrano, D. Krioukov, and M. Boguñá, "Percolation in self-similar networks," *Physical Review Letters*, vol. 106, no. 4, Article ID 048701, 2011.
- [16] G. Ganesan and S. J. S. Sansanwal Marg, "Size of the giant component in a random geometric graph," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 49, no. 4, pp. 1130–1140, 2013.

- [17] D.-S. Lee, K.-I. Goh, B. Kahng, and D. Kim, "Sandpile avalanche dynamics on scale-free networks," *Physica A*, vol. 338, no. 1-2, pp. 84–91, 2004.
- [18] I. Dobson, B. A. Carreras, and D. E. Newman, "A loading-dependent model of probabilistic cascading failure," *Probability in the Engineering and Informational Sciences*, vol. 19, no. 1, pp. 15–32, 2005.
- [19] G. Sansavini, M. Hajj, I. Puri, and E. Zio, "A deterministic representation of cascade spreading in complex networks," *Europhysics Letters*, vol. 87, no. 4, 2009.
- [20] L. Jiang, X. Jin, Y. Xia, B. O. Ouyang, and D. Wu, "Dynamic behavior of the interaction between epidemics and cascades on heterogeneous networks," *EPL*, vol. 108, no. 5, Article ID 58009, 2014
- [21] A. E. Motter and Y.-C. Lai, "Cascade-based attacks on complex networks," *Physical Review E*, vol. 66, no. 6, Article ID 065102, 2002
- [22] J. Lehmann and J. Bernasconi, "Stochastic load-redistribution model for cascading failure propagation," *Physical Review E*, vol. 81, no. 3, Article ID 031129, 2010.
- [23] P. Erdős and A. Rényi, "On the evolution of random graphs," in *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.
- [24] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [25] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, no. 2, Article ID 026118, 2001.
- [26] D. J. Watts, "A simple model of global cascades on random networks," Proceedings of the National Academy of Sciences of the United States of America, vol. 99, no. 9, pp. 5766–5771, 2002.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 946281, 9 pages http://dx.doi.org/10.1155/2015/946281

Research Article

Stability Analysis of Process Route Based on Weighted Network

Ying Xiang, Rong Mo, Zhiyong Chang, Hu Qiao, and Chunlei Li

The Key Laboratory of Contemporary Design and Integrated Manufacturing Technology, Ministry of Education, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Ying Xiang; yingcara@hotmail.com

Received 26 June 2015; Accepted 2 November 2015

Academic Editor: Xiaobo Qu

Copyright © 2015 Ying Xiang 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.

Aiming at the production stability of complex parts, a method based on weighted network was proposed to analyze the stability of the process routes of complex parts in order to improve the production stability. The weighted network of the process routes of complex parts was constructed by using the concept of machining cell which can transform the production cost and manufacturing time to the weights of network to do the decision-making of the process routes. Based on the production stability, the brittleness risk entropy of subsystem of weighted network was constructed by analyzing the probability of the brittle events that may lead to the collapse of the weighted network in machining cells. As the indicator of analyzing the vulnerability of weighted network node, the brittleness risk entropy can predict the easily failed subsystem in the entire network. Meanwhile, the brittle event, which may lead to the machining cell failure, was retrospect for the greater stability of the process routes. Finally, the correctness and effectiveness of this method were verified by using the manufacturing process of an aero-engine blade.

1. Introduction

The diversity of process route in business production is derived from two sides. First, the order and methods of manufacturing features of parts are not unique. Second, the manufacturing resource in enterprise is various and similar. Next, there are many factors in production which may lead to the instability. For example, the factors like the irrational process routes which may lead to the imbalance of manufacturing resource and then trigger production faults and the brittle events (toll wear) which may delay the processing may cause catastrophic collapse of production. Therefore, as the hinge between the products design and manufacture, process route must be chosen with reason that can raise production efficiency and reduce costs and meanwhile avoid the occurrence of brittle events to increase the stability of business production.

Recently, the research shows some methods describing the process of parts manufacturing, which are the object-oriented method [1, 2], the petri net-based modeling method [3–5], discrete event dynamic simulation method [6, 7], and so forth. These methods put the emphasis on describing and definition of the elements in manufacturing process, while

lacking description of the correlation among the elements. Based on this, researchers proposed graph theory and complicated network modeling methods [8–12]. They imported these methods into practice and described the correlation of different elements preliminarily. This graph theory-based process route is subject to the theory of graph itself, so it is limited in long process route. In addition, this method emphasizes that process route is optimal in some ways, while ignoring the stability of process route.

In the research about the system stability, Fouad et al. first propose the concept of system vulnerability and build an analytic technique on vulnerability by using transient energy function (TEF) method and neural networks [13]. In 2000, Albert et al. research vulnerability source based on complicated theory [14]. Therefrom, they bring the system vulnerability into a new age. Xu and Wang put up a scale-free coupled map lattices (CML) model that is a good math method for cascading failures of complicated network [15]. Jin et al. research the system vulnerability by using the theory of entropy and systematic mutation [16]. Some results emphasizing theories are achieved, which aim at universality researches about complicated system. For manufacture system, due to its complicacy and specificity, there is discrepancy

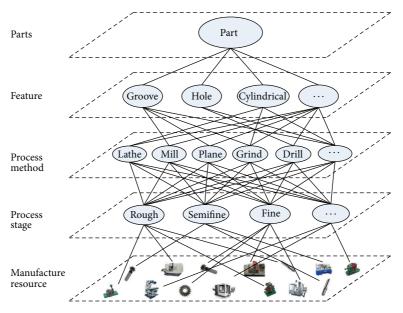


FIGURE 1: The hierarchical graph of parts processing.

between it and the common complicated system in system vulnerability field. Therefore, further efforts should be made to research it.

In this paper, a weighting network-supped process route method about complicated parts is put up based on production stability and system vulnerability. The weighted network of the process routes of complex parts was constructed by using machining cell. And then, the brittleness risk entropy of subsystem of weighted network was constructed by analyzing the probability of the brittle events. The brittle event, which may lead to the machining cell failure, was retrospect in order to improve the production stability. Also, the correctness and validity are verified by instance in this paper.

2. The Weighted Network of the Process Routes of Complex Parts

2.1. Machining Cell. Routings consist of some object-oriented manufacturing features which are arranged by the processing route and the connection among them. Beginning with the manufacturing features of elements, the route of process should be designed by considering different factors, such as the parts materials, the relationship among features, manufacture resource and the design experience of technologist, and so forth. It may be divided into these five steps: partsfeature-process method-process stage-manufacture resource (Figure 1).

Definition 1 (machining cell). The processing of characteristics of mechanical parts is sorted according to the principle "coarse to fine" to form a machining operation sequence, also known as the processing chain; each element in the processing chain is called "machining cell" [17]. It can be expressed as

$$U_{ab} = \{ F_a, L_l, G_b, D_d \}, \tag{1}$$

where F_a is the ath manufacturing feature which belongs to the machining cell, such as hole, groove, and plane; L_l is the processing method which belongs to F_a , for example, lathe, mill, plane, grind, and drill; G_b is processing stage which belongs to F_a . Usually, there are rough machining, semifine machining, and fine machining. D_d is manufacturing resources used in the processing stage, mainly referring to the machine, cutting, and fixture.

Therefore, a machining cell has only one manufacturing feature that may be a process and may also be a step. At the same time, a machining cell belongs to only one processing stage and one processing method and uses only one cutting and one machine. In the production line, all machining cell formed the high cohesion and low coupling station of having manufacturing capacity in a special way. And then, the roughcast can be processed between station and another station in the driving of the manufacturing technology. The product of meeting the design requirements was obtained ultimately.

2.2. The Construction of Weighted Network. Multistage machining processes have many machining processes and many coupling relationships among machining processes; thus complex network theory can be introduced in multistage machining processes.

In the design of process scheme for multistage machining processes, there are several process schemes that exist in interrelation with other machining cells in this scheme and the machining cell in the other process schemes. So the machining cells are mapped as weighted network nodes. Because the coupling relationships among machining cell are machining feature and datum relation, coupling relationships are mapped as edges among weighted network nodes. In machining process of product, as the production cost and manufacturing time are important characteristic and they are

one of the important factors of the process scheme design that reflects the cost and efficiency, the production cost and manufacturing time are mapped as the weights of weighted network.

Depending on complex network theory, a weighted network of multistage machining scheme for complex parts manufacturing process is constructed by the weighted network nodes, the edges among weighted network nodes, and the weights of weighted network. Figure 2 shows the weighted network construction of multistage machining scheme for complex parts manufacturing process.

In Figure 2, U_j denotes the jth machining cell; P_k denotes the kth stage machining process in complex parts manufacturing process; and V_j denotes the jth note in the weighted network of complex parts manufacturing process.

W is defined as weight matrix among weighted network nodes, so $W = [w_{1,1}, w_{1,2}, \ldots, w_{h,j}]_{n \times n}; w_{h,j} \ (0 \le h \le n, \ 0 \le j \le n)$ is the weight of weighted network from hth to jth and denotes the production cost and manufacturing time of machining cell.

Definition 2. The number of adjacent nodes with V_j is the degree of V_i , denoted as k_i .

Corollary 3. The number of adjacent nodes pointing to V_j is the out-degree of V_j , denoted as k_j^{out} . The out-degree is the influence scope of processing capacity of machining cell.

Corollary 4. The number of adjacent nodes pointing to V_j is the in-degree of V_j , denoted as k_j^{in} . The in-degree is the scope of processing objects of machining cell.

Definition 5. The strength of nodes s_j is the connection ability of V_i and other nodes:

$$s_j = \sum_{h \in \alpha} w_{jh},\tag{2}$$

where α is the number set of adjacent nodes with V_i .

Corollary 6. The strength of nodes s_j^{out} is the connection ability of V_j and adjacent nodes pointing to V_j , and s_j^{out} is the influence extent of processing capacity of machining cell:

$$s_j^{\text{out}} = \sum_{h \in \mathcal{B}} w_{jh},\tag{3}$$

where β is the number set of adjacent nodes pointing to V_i .

Corollary 7. The strength of nodes s_j^{in} is the connection ability of V_j and adjacent nodes pointing to V_j , and s_j^{in} is the assignment amount of machining cell:

$$s_j^{\text{in}} = \sum_{h \in \mathcal{V}} w_{jh},\tag{4}$$

where γ is the number set of adjacent nodes pointing to V_i .

Definition 8. The clustering coefficient is the influence extent of the clustering degree of adjacent nodes for this node in weighted network. The bigger the clustering degree of nodes is, the higher the key degree of machining cell is. In weighted network, c_i is expressed as

$$c_{j} = \frac{1}{s_{j}\left(k_{j}-1\right)} \cdot \sum_{h,l \in \alpha} \frac{\left(w_{j,h} + w_{k,l}\right)}{2} \left(a_{jh} a_{hl} a_{lj}\right). \tag{5}$$

3. The Decision-Making Model of Multistage Process Routes

In weighted network, machining cells are regarded as network nods, which represent different manufacturing phases, manufacturing methods, and manufacturing resources. Because there are various manufacturing methods and resources in every phase, the whole process may produce various process routes in weighted network. That is to say, we must do decision-making in different process routes to select the most reasonable one that can guide the manufacturing process.

The designer usually designs multiple process routes which contrapose one part. Different routes mean different cost and time in practice. The paper aims at the manufacturing costs and time and does decision-making in various routes to pursue the most reasonable weight in weighting network.

3.1. The Production Cost. The total cost of parts manufacturing includes two parts: the production cost $(M_{\rm in})$ and the changing cost (M_q) . The machining cell is used as the node in this paper. The machine cost and tool cost are totally called the inner production cost $(M_{\rm in})$ of machining cell. In this network, the cost changing between adjacent nodes is called the changing cost (M_q) . For example, if different production processes need different machines in manufacturing, M_q is cost of changing one machine to another which includes the cost of labor and transportation. Consequently, in the process route in practice, if there are m kinds of routes, n machining cells in one process route, the mathematical models of M_i in ith route may be described as

$$M_{i} = \left(M_{\text{in}} + M_{q}\right) = \left(\sum_{j=1}^{n} M_{i,j} + \sum_{j=1}^{n-1} M_{i,(j,j+1)}\right)$$

$$i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n,$$
(6)

where $M_{i,j}$ is the inner production cost of node j in the ith process route; $M_{i,(j,j+1)}$ is the changing cost from node j to node j+1 in the ith process route.

3.2. Manufacturing Time. The manufacturing time includes machining time and transportation time. Consequently, in the process route in practice, if there are m kinds of routes, n machining cells in one process route, the

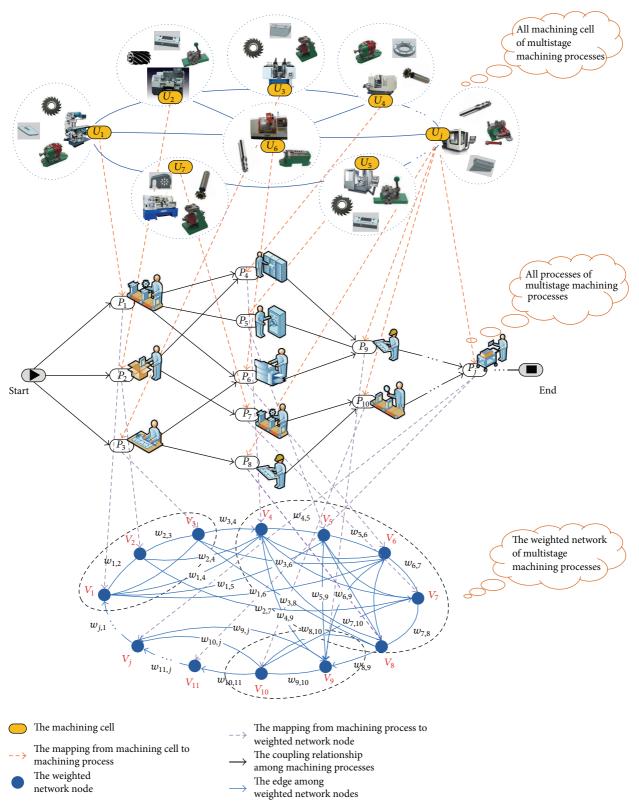


FIGURE 2: The weighted network construction of multistage machining process.

mathematical models of T_i in ith route may be described as

$$T_{i} = (T_{0} + T_{tr}) = \left(\sum_{j=1}^{n} T_{i,j} + \sum_{j=1}^{n-1} T_{i,(j,j+1)}\right)$$

$$i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n,$$
(7)

where $T_{i,j}$ is the machining time of the node j in the ith process route; $T_{i,(j,j+1)}$ is the transportation time between the node j and node j+1 in the ith process route.

By the means of the multiprocess route decision model mentioned above, we can get the ultimate machining cell by selecting the useful one, meanwhile judging whether the cell is in practicable status or not. The relationship among the machining cells shapes the multiprocess route decision network, and the useful machining cells form the nodes in this network. The weight of the network is the production cost and manufacturing time. It can be analyzed to get the most suitable decision to make the best balance between the production cost and manufacturing time. Due to considering the production cost and manufacturing time overall, the multiprocess route decision model is a kind of multiple target. In general, the math model of the multiprocess route decision mentioned above can be described as

$$f(M_i, T_i) = \frac{k_1}{M_i} + \frac{k_2}{T_i}.$$
 (8)

So

$$f(M_i, T_i) = \frac{k_1}{\left(\sum_{j=1}^n M_{i,j} + \sum_{j=1}^{n-1} M_{i,(j,j+1)}\right)} + \frac{k_2}{\left(\sum_{j=1}^n T_{i,j} + \sum_{j=1}^{n-1} T_{i,(j,j+1)}\right)}$$

$$i = 1, 2, \dots, m, \ j = 1, 2, \dots, n$$

$$(9)$$

$$k_1 + k_2 = 1, \quad k_1, k_2 \in [0, 1],$$

where adjusting k_1 and k_2 will decide the focus of process scheme decisions.

Currently, for the multifunction of the process route decision, there are abundant algorithms provided by the correlative papers [17–20]. This paper solves the multiprocess route decision model by the Adaptive Ant Colony Algorithm proposed by Chang et al. [17]; there is no longer repetition.

4. The Vulnerability Analysis of Machining Cell

In the process decision method, there may be a variety of reasonable process routes, but the machining cell may often be due to abnormal processing equipment which hindered the process. So we need to analyze the abnormal probability of machining cell in order to obtain the most reasonable and high stability process route.

Therefore, it is necessary to analyze the brittle risk of machining cell in order to grasp the abnormal probability of machining cell. That may improve the stability of the process route.

In weighting network the process routes of complex parts, taking the machining cell as the node of the network, the network S is configured by n subnetwork if taking the machining cell U_j as subsystem of the network S, may be expressed as S_i , so $S = \{S_1, S_2, \ldots, S_n\}$.

According to formula (8), the reasonable process route is obtained in weighted network of the process routes of complex parts. In the machining process of part, the brittle events (tool wear, equipment load, etc.) may cause the system malfunction in subsystem (i.e., there is brittle risk in subsystem). Along with the transfer and expansion of the collapse behavior, thereby the entire manufacturing process may be paralyzed. Brittleness risk of network system is the risk of subsystem failure because the brittleness of subsystem is stimulated. It is considered that the fundamental source of brittleness risk of the entire system is the uncertainty of brittle events of subsystem. So the subsystem failure triggered by brittle event is needed to be predicted and controlled in order to ensure the stability of the weighted network of the process route of complex parts.

If the subsystem S_j has brittle events $E = \{E_1, E_2, \dots, E_t\}$, the probability of brittle event E_r ($r = 1, 2, \dots, t$) is p_r ; there is a basic relationship:

$$\sum_{r=1}^{t} p_r = 1, \quad 0 \le p_r \le 1. \tag{10}$$

The probability of subsystem occurrence failure is q_r under the action of the brittle event E_r , and $0 \le q_r \le 1$, so the probability P_{Sj} of subsystem occurrence failure due to brittle events is shown as

$$P_{S_j} = 1 - \prod_{r=1}^{t} (1 - p_r q_r) \quad j = 1, 2, ..., n.$$
 (11)

The influence coefficient, the failure of machining cell effect collapse of the weighted network, is defined as ξ . That is, in the network system $S = \{S_1, S_2, \dots, S_n\}$, the influence coefficient, the failure of subsystem S_j effect collapse of the system S_j , is defined as ξ_j . The influence coefficient and weighting network clustering coefficient have a similar function, so the influence coefficient ξ_j may be mapped as the normalized values of the clustering coefficient:

$$\xi_{j} = \frac{c_{j}}{\sum_{j=1}^{n} c_{j}}.$$
 (12)

The effective coefficient of subsystem S_j collapse is the normalized values of subsystem collapse probability and influence coefficient:

$$g_{j} = \frac{P_{s_{j}} \xi_{j}}{\sum_{j=1}^{n} P_{s_{j}} \xi_{j}},$$

$$\sum_{j=1}^{n} g_{j} = 1, \quad 0 \le g_{j} \le 1.$$
(13)

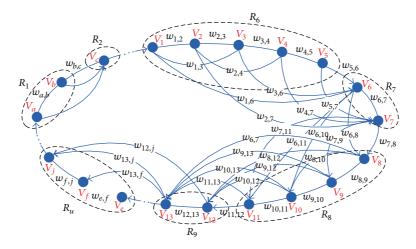


FIGURE 3: The weighted network of blade machining processes.

According to Shannon theory [21], the brittleness risk entropy H(S) of subsystem S_j is defined as the average value of the risk probability of brittle event in utility coefficient space, expressed as

$$H\left(S_{j}\right) = -g_{j}\log P_{s_{j}}.\tag{14}$$

So

$$H(S_{j}) = -\frac{\left(1 - \prod_{r=1}^{t} (1 - p_{r}q_{r})\right) \left(c_{j} / \sum_{j=1}^{n} c_{j}\right)}{\sum_{j=1}^{n} \left(1 - \prod_{r=1}^{t} (1 - p_{r}q_{r})\right) \left(c_{j} / \sum_{j=1}^{n} c_{j}\right)} \cdot \log\left(1 - \prod_{r=1}^{t} (1 - p_{r}q_{r})\right).$$

$$(15)$$

Brittleness risk entropy of subsystem can reflect not only the fundamental source of subsystem brittle which is uncertain of brittle events, but also influence of collapse coefficient. Brittleness risk entropy of subsystem combines these both effects for system brittleness risk effectively [16]. Brittleness risk entropy is used to measure brittleness risk probability of subsystem S_j in a moment, which is the uncertainty measurement of subsystem possible collapse, and may predict the brittleness risk of subsystem.

Formula (14) shows that there is a close relationship between the subsystem brittleness risk entropy and the node clustering coefficient that is topological properties of the weighted network, so weighting network of the complex parts processing, having different topological properties, has different subsystem brittleness risk entropy. Through the analysis of the subsystem of brittleness risk entropy, the brittleness risk entropy of each machining cell and the brittleness risk entropy of the process route which have the machining cell can be obtained. The minimum brittleness risk entropy of the process route, namely, the highest stability process route, can be obtained by analyzing the brittleness risk entropy of process route. In the production, the higher

stability and reasonable process route can guarantee the stability of production.

5. Case Study

An experiment case of the manufacturing process of an aeroengine blade [21] is performed to verify the effectiveness of this method. The engine manufacturers have advanced blade machining centers that introduce a large number of advanced types of equipment, such as axis CNC machine tools and five-axis CNC machine tools. According to the actual situation of the enterprise, the aero-engine blade design requirements, manufacturing costs, and production time basis, the proposed method is verified.

The weighted network of blade machining process is formed by the intrinsic link between each technology program, in the enterprise. In this network, one of the aeroengine blade machining tasks may have a variety of processing methods, and each processing method can be achieved through multiple processing element. In the preparation process, the weighted network of the machining process, regarding machining cell as node, is formed by considering variety of resources for each processing element and different conditions, shown in Figure 3. Due to the complexity of machining aerospace engine blades, limited space of this paper, an experiment case of a part of the blade manufacturing features is used to explain the building process of the weighted network.

The portion process of the blade and the information of the required machining cell are shown in Table 1, and the various processing tasks shown in Table 1 may be completed by the corresponding machining cell shown in Table 2.

The blade processing has 36 different technology programs from craft station of the blade top to positioning holes of craft station from Table 2, because of the multiple choice of the processing methods from 6 to 9, and the machining cells matched the processing methods, as shown in Figure 3. In processing blade, the local factory focuses on manufacturing

Task	Process task	Process method	Accuracy class	Surface roughness Ra/ μ m	Note
R_6	Blade top craft station	Milling	IT10	3.2	Poor milling, CNC machining, special fixtures
R_7	Blade dorsal positioning surface	Milling	IT10	3.2	Four-axis CNC machining, special fixtures
R_8	Tenon root	Wire cutting	IT10	3.2	CNC wire-cutting machine, special fixtures
R_9	Positioning holes of craft station	Drilling	IT9	3.2	Four-axis CNC machining, special fixtures
:	:	:	:	:	:

TABLE 1: Process tasks of a blade.

TABLE 2: The machining cell information matched with process tasks.

Task	The number of optional machining cells	The number of corresponding nodes	The processing method	Accuracy class	Surface roughness Ra/µm	The location of machining cell	Production cost (thousand)	Manufacturing time (day)
	U_{61}	V_1		11~7	6.4~3.2	Other factories	0.6	1
	U_{62}	V_2		10~7	6.4~1.6	Other factories	0.5	2
R_6	U_{63}	V_3	Milling	10~7	6.4~1.6	Other factories	0.55	2
	U_{64}	V_4		11~7	6.4~3.2	Other factories	0.45	2
	U_{65}	V_5		10~7	6.4~1.6	Local factory	0.8	1
R_7	U_{71}	V_6	Milling	10~6	6.4~3.2	Local factory	1.2	1
	U_{72}	V_7	Milling	9~5	6.4~0.8	Local factory	1.8	1
R_8	U_{81}	V_8		12~8	6.4~3.2	Other factories	0.2	2
	U_{82}	V_9	Wire	10~7	6.4~1.6	Other factories	0.2	3
	U_{83}	V_{10}	cutting	12~8	6.4~3.2	Other factories	0.15	2
	U_{84}	V_{11}		9~5	6.4~0.8	Other factories	0.3	2
D	U_{91}	V_{12}	Drilling	10~6	6.4~3.2	Local factory	1.8	1
R_9	U_{92}	V_{13}	Drilling	9~5	6.4~0.8	Local factory	2.2	1
:	:	:	:	:	:	:	:	:

time; therefore, assuming that $k_1 = 0.3$, $k_2 = 0.7$, the decision function of preferred technology program is as follows:

$$f(M,T) = \frac{0.3}{\left(\sum_{j=1}^{n} M_{i,j} + \sum_{j=1}^{n-1} M_{i,(j,j+1)}\right)} + \frac{0.7}{\left(\sum_{j=1}^{n} T_{i,j} + \sum_{j=1}^{n-1} T_{i,(j,j+1)}\right)}.$$
(16)

In the weighted network of blade manufacturing, the decision indexes of every technology program are shown in Table 3. There is a part of program index because of the limited space, in route 3 $(V_1-V_6-V_{10}-V_{12})$, route 30 $(V_5-V_6-V_{10}-V_{12})$, route 31 $(V_5-V_6-V_{10}-V_{13})$, $f_3=0.3/3.9+0.7/15=0.12359$, $f_30=0.3/4.3+0.7/12=0.128101$, and $f_31=0.3/4.35+0.7/13=0.122812$.

The production cost and manufacturing time are the most reasonable in these programs.

Nodes' vulnerability is related to clustering coefficient closely in the weighted network of this blade. According to formula (4), nodes' clustering coefficients are derived and the distribution is shown in Figure 4.

Formula (14) derives that node V_6 's Brittleness risk entropy is the highest, as shown in Figure 5. In the reasonable process routes, the node V_1 corresponding to the machining cell U_{61} and the node V_5 corresponding to the machining cell U_{65} can be selected; the node V_{12} corresponding to the machining cell U_{91} and the node V_{13} corresponding to the machining cell U_{92} can be selected. Thus, the route 30 is better than route 3 because that brittleness risk entropy of V_1 is greater than brittleness risk entropy of V_5 . The route 30 is better than route 31 because that brittleness risk entropy of V_{12} is greater than brittleness risk entropy of V_{13} . Therefore, the route 30 (V_5 - V_6 - V_{10} - V_{12}) is the highest stability process route.

During the manufacturing process of aero-engine blade, node V_6 corresponds to machine cell U_{71} and by Figure 5 we derive that V_6 is prone to collapse. Machine cell U_{71} is a part of most reasonable process route. By analyzing the event that triggered machine cell U_{71} collapses, we find that the most important reason is this machine cell easily overloaded. So enterprise should arrange U_{71} corresponding four-axis CNC machine's usage reasonable in the manufacturing

	Optional technology program																				
Program	Program 1			Program 2			Program 3			Program 4			Program 5			• • •					
	1	6	8	12	1	6	8	13	1	6	10	12	1	6	10	13	1	6	11	12	• • •
Production cost M		4.05			1.4			3.9			3.95			4.15							
Manufacturing time T_0		17			16			15			16				17						
Decision function value		0.115251				0.111932				0.12359			0.119699				0.113466				

TABLE 3: The evaluation of a part of craft program.

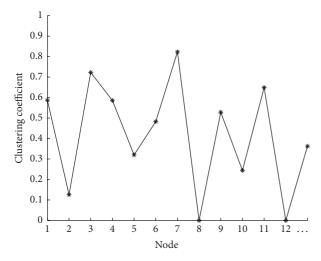


FIGURE 4: The clustering coefficients of a part of nodes.

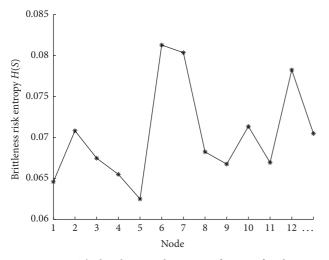


Figure 5: The brittleness risk entropy of a part of nodes.

process to avoid production disruption caused by four-axis CNC machine overload. This is consistent with the actual production.

6. Conclusions

(1) Based on part's manufacturing feature, weighted network of the process routes of complex parts is established by using machine cell, and the topological characteristics of weighted network are analyzed.

- (2) Through transforming production cost and manufacturing time to the weight of network, the multiprocess routes decision-making of weighted network mathematical model is established to make process route decision.
- (3) The factors lead to weighted network collapse which is analyzed to establish brittleness risk entropy of the subsystem of weighted network to obtain the highest stability process route.
- (4) Proposed method is used to establish the weighted network of one aero-engine blade manufacturing process to choose the reasonable manufacturing process route. Thereby the effectiveness of the proposed method is verified by analyzing the vulnerability of weighted network nodes.
- (5) The further research area will focus on deriving the substitute of brittle machine cell quickly according to brittle distance theory of complex network and analyzing the pattern of system brittleness propagation.

Conflict of Interests

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

Acknowledgment

The project was supported by the National Natural Science Foundation of China (51375359).

References

- [1] Y. N. Sotskov, A. Dolgui, and M.-C. Portmann, "Stability analysis of an optimal balance for an assembly line with fixed cycle time," *European Journal of Operational Research*, vol. 168, no. 3, pp. 783–797, 2006.
- [2] X. Shao and Y. Rao, Manufacturing System Operation Optimization Theory and Methods, Science Press, Beijing, China, 2010.
- [3] M. Uzam, "The use of the Petri net reduction approach for an optimal deadlock prevention policy for flexible manufacturing systems," *International Journal of Advanced Manufacturing Technology*, vol. 23, no. 3-4, pp. 204–219, 2004.
- [4] G. Meja, C. Montoya, J. Cardona, and A. L. Castro, "Petri nets and genetic algorithms for complex manufacturing systems scheduling," *International Journal of Production Research*, vol. 50, no. 3, pp. 791–803, 2012.

- [5] Z. Cai, "Timed reconfigurable petri net-based modeling of reconf igurable manufacturing system," Acta Aeronautica et Astronautica Sinica, vol. 25, no. 6, pp. 615–618, 2004.
- [6] B. Bouslah, A. Gharbi, R. Pellerin, and A. Hajji, "Optimal production control policy in unreliable batch processing manufacturing systems with transportation delay," *International Journal of Production Research*, vol. 51, no. 1, pp. 264–280, 2013.
- [7] Q. Wang and C. R. Chatwin, "Key issues and developments in modelling and simulation-based methodologies for manufacturing systems analysis, design and performance evaluation," *International Journal of Advanced Manufacturing Technology*, vol. 25, no. 11-12, pp. 1254–1265, 2005.
- [8] X.-H. Liu, X.-T. Zhang, and W.-J. Liu, "Multi-process routes decision-making methodology based on improved max-min ant system," *Computer Integrated Manufacturing Systems*, vol. 14, no. 12, pp. 2414–2420, 2008.
- [9] Y. Qin, L. Zhao, Y. Yao, and D. Xu, "Multistage machining processes variation propagation analysis based on machining processes weighted network performance," *The International Journal of Advanced Manufacturing Technology*, vol. 55, no. 5–8, pp. 487–499, 2011.
- [10] T. Yang, D. Zhang, B. Chen, and S. Li, "Capacity planning for mixed production line based on directed weighted networks," *China Mechanical Engineering*, vol. 22, no. 7, pp. 824–829, 2011.
- [11] F. Jian, P. Jiang, D. Liu et al., "Error propagation control method for multistage batches machining process of blades," *Computer Integrated Manufacturing Systems*, vol. 18, no. 1, pp. 76–86, 2012.
- [12] W.-J. Huang, L.-G. Cai, Y.-J. Hu, X.-L. Wang, and L. Ling, "Process planning optimization based on genetic algorithm and topological sort algorithm for digraph," *Computer Integrated Manufacturing Systems*, vol. 15, no. 9, pp. 1770–1778, 2009.
- [13] A. A. Fouad, Q. Zhou, and V. Vittal, "System vulnerability as a concept to assess power system dynamic security," *IEEE Transactions on Power Systems*, vol. 9, no. 2, pp. 1009–1015, 1994.
- [14] R. Albert, H. Jeong, and A.-L. Barabási, "The internet's achilles heel: error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, 2000.
- [15] J. Xu and X. F. Wang, "Cascading failures in scale-free coupled map lattices," *Physica A: Statistical Mechanics and its Applications*, vol. 349, no. 3-4, pp. 685–692, 2005.
- [16] H.-Z. Jin, Q. Wei, J. Guo et al., Brittle Theory and Application of Complex Systems, Northwestern Polytechnical University Press, Xi'an, China, 2010.
- [17] Z. Chang, J. Yang, J. Zhao, and H. Wei, "Optimization of process based on adaptive ant colony algorithm," *Chinese Journal of Mechanical Engineering*, vol. 48, no. 9, pp. 163–169, 2012.
- [18] M. K. Tiwari, Y. Dashora, S. Kumar, and R. Shankar, "Ant colony optimization to select the best process plan in an automated manufacturing environment," *Proceedings of the Institution of Mechanical Engineers, Part B. Journal of Engineering Manufacture*, vol. 220, no. 9, pp. 1457–1472, 2006.
- [19] A. G. Krishna and K. M. Rao, "Optimisation of operations sequence in CAPP using an ant colony algorithm," *The International Journal of Advanced Manufacturing Technology*, vol. 29, no. 1-2, pp. 159–164, 2006.
- [20] P. Wang, D.-H. Zhang, B. Chen, S. Li, and G.-H. Ma, "Evaluation of multi-process plans based on fuzzy comprehensive evaluation and grey relational analysis," *Journal of Aerospace Power*, vol. 27, no. 9, pp. 2075–2085, 2012.
- [21] H. Jin, Q. Wei, J. Guo et al., Brittle Theory and Application of Complex Systems, Northwestern Polytechnical University Press, Shanghai, China, 2010.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 482923, 20 pages http://dx.doi.org/10.1155/2015/482923

Research Article

Robust Satellite Scheduling Approach for Dynamic Emergency Tasks

Xuejun Zhai, 1,2 Xiaonan Niu, 1,2 Hong Tang, 1,2 Lixin Wu, 3 and Yonglin Shen 4

¹ State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
² Key Laboratory of Equipment Change and Natural Disaster, Ministry of Education, Beijing Normal University

²Key Laboratory of Environment Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing 100875, China

Correspondence should be addressed to Hong Tang; hongtang@bnu.edu.cn

Received 30 June 2015; Accepted 15 October 2015

Academic Editor: Xiaobo Qu

Copyright © 2015 Xuejun Zhai 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.

Earth observation satellites play a significant role in rapid responses to emergent events on the Earth's surface, for example, earthquakes. In this paper, we propose a robust satellite scheduling model to address a sequence of emergency tasks, in which both the profit and robustness of the schedule are simultaneously maximized in each stage. Both the multiobjective genetic algorithm NSGA2 and rule-based heuristic algorithm are employed to obtain solutions of the model. NSGA2 is used to obtain a flexible and highly robust initial schedule. When every set of emergency tasks arrives, a combined algorithm called HA-NSGA2 is used to adjust the initial schedule. The heuristic algorithm (HA) is designed to insert these tasks dynamically to the waiting queue of the initial schedule. Then the multiobjective genetic algorithm NSGA2 is employed to find the optimal solution that has maximum revenue and robustness. Meanwhile, to improve the revenue and resource utilization, we adopt a compact task merging strategy considering the duration of task execution in the heuristic algorithm. Several experiments are used to evaluate the performance of HA-NSGA2. All simulation experiments show that the performance of HA-NSGA2 is significantly improved.

1. Introduction

Earth observing satellites (EOSs), orbiting the Earth, are able to collect images of specified areas of the Earth's surface at the request of customers by using observation sensors [1]. EOSs not only play key roles in application for environment surveillance, reconnaissance, resource investigation, and other fields, but also support many important services, such as remote sensing, navigation, geodesy, and monitoring [2]. As a result, many countries, especially China, tend to increase their investments to develop associated techniques and EOSs. For example, China plans to launch four small optical satellites and four small SAR satellites to form a natural disaster-monitoring constellation [3]. Although the number of EOSs is continuously increasing, they are still unable to satisfy the requirements of various customers [4]. Therefore, it is very important to develop effective methods of satellites observation scheduling to make full use of scarce satellite resources to better satisfy the demands of customers. The satellite observation scheduling problem (SOSP) is to reasonably assign satellite resources and time windows to Earth observation tasks on the precondition of satisfying complex constraints [5]. Additionally, SOSP can be seen as a kind of multidimensional knapsack problem that is NP-hard [6].

From the previous survey, the satellite observation tasks scheduling problem can be divided into two classes, that is, static tasks scheduling and dynamic tasks scheduling. Over the past decades, the satellite scheduling imaging problem has been intensively investigated [7–10]. Regarding static single satellite scheduling, Lin et al. employed mathematical programming methods such as Tabu search, Lagrange relaxation, and liner search to solve the scheduling problem and acquired a near-optimal schedule [11–13]. Wolfe and Sorensen proposed three corresponding algorithms including a dispatch algorithm, a look-ahead algorithm, and a genetic algorithm to

³School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China ⁴Faculty of Information Engineering, China University of Geosciences, Wuhan 430074, China

model the imaging satellite scheduling problem [14]. Gabrel and Murat presented two upper bound procedures based on the graph theory and a column generation technique [15]. Vasquez and Hao formulated the satellite observation scheduling problem into a generalized version of the well-known knapsack model and used the Tabu search method to solve the problem [16]. Furthermore, they proposed a partition-based approach UPPB to obtain tight upper bounds to evaluate the TS search algorithm [17]. Several exact methods, such as depth first branch and bound or Russian dolls search and some approximate methods, such as greedy search and Tabu search, were used by Bensana et al. to solve the scheduling problem of the SPOT5 satellite. They viewed the scheduling problem as a value constraint satisfaction problem or an integer linear programming problem [6].

Although the single satellite scheduling problem can be solved perfectly by the above methods, with the development of the EOS system, these methods are unable to satisfy the needs of multisatellite system scheduling. In later studies, many researchers adopted a heuristic algorithm to solve the multisatellite scheduling problem such as Zweben et al. [18], Frank et al. [19], Bianchessi and Righini [20], Wang and Reinelt [21], and Marinelli et al. [22]. Bianchessi and Righini proposed a FIFO heuristic and a look-ahead insert heuristic algorithm to solve the scheduling problem [20]. Frank et al. used a constraint based interval planning framework to model the problem and proposed a heuristic for guiding this search procedure. Wang and Reinelt considered image downloads and proposed a priority-based conflict-avoided heuristic algorithm [21]. With the development of the intelligent optimization algorithm, many researchers proposed using intelligent algorithms to solve the complex scheduling problem. The intelligent algorithms mainly include genetic algorithm and ant colony optimization algorithm. Barbulescu et al. [23], Wang et al. [24], Baek et al. [25], Chen et al. [26], and Mansour and Dessouky [7] used the genetic algorithm (GA) to address the satellite scheduling problem. Barbulescu et al. proposed Genitor, a genetic algorithm, and proved that it performs well for a broad range of problem instances [23]. A multiobjective earth observation satellite scheduling method called SPEA2, based on the strength of the Pareto evolutionary algorithm, was proposed by Wang et al. [24]. Additionally, Mansour and Dessouky solved the satellite scheduling problem by developing a genetic algorithm considering two objectives, that is, the profits and the number of acquired photographs [7]. Ant colony optimization (ACO) was also used to address this problem. Based on ant colony optimization, Zhang et al. [27] and Liu et al. [28] considered multisatellite resources and task merging, respectively. Wu et al. presented a novel two-phase based scheduling method with the consideration of task clustering [29]. Moreover, Qiu et al. proposed the first finish first schedule with discard task moving back (FFFS-DTMB) and accommodate discard task predicting coexistence with task moving back (ADTPC-DTMB) algorithms to solve the problem [30]. Globus et al. developed an evolutionary algorithm to solve this problem and compared it with existing algorithms [31, 32]. From the survey of static tasks scheduling, we find that the heuristic algorithm is shown to perform well for simple problems and

the genetic algorithm is proven to be suitable for large scale, more complex problems [23].

Static scheduling assumes that all imaging tasks have been submitted before scheduling and once the scheduling scheme is produced, it is immutable until all tasks have been finished. In practice, because emergent events usually occur unexpectedly, with uncertainties of occurrence time and number, traditional static scheduling cannot attend to these emergent tasks in time. It is suggested that dynamic scheduling copes with these unexpected factors.

Until now, there are a few research efforts towards the dynamic scheduling on Earth observation satellites. Pemberton and Greenwald described the problem of dynamic scheduling and discussed contingency conditions [33]. Kramer and Smith proposed a repair-based search method for the oversubscribed scheduling problem [34]. Verfaillie and Schiex developed an approach employing local adjustment to solve the dynamic constraint satisfaction problem (DCSP) [35]. Considering the tradeoff between the performance and degree of adjustment, Wang et al. proposed a rulebased heuristic algorithm to solve the dynamic scheduling problem [36]. Also, Wang et al. established a multiobjective mathematic programming model for the dynamic real-time scheduling of EOSs and proposed a dynamic real-time scheduling algorithm called DMTRH to solve the problem [37]. Through analyzing the main constraints, Dishan et al. constructed an integer programming model and proposed a Rolling Horizon (RH) strategy with a mixed triggering mode to schedule the dynamic tasks and common tasks together [2].

To the best of our knowledge, EOSs play an important role in the process of disaster relief. For example, when an earthquake occurs, new emergency tasks will be submitted to the satellite observation system to acquire images that contain disaster information. According to the needs of disaster relief, more and more different emergency tasks will arrive. So far, just a few dynamic scheduling methods consider dealing with the emergency tasks. After analyzing the dynamic properties of satellite scheduling, Wang et al. proposed an optimization model and two heuristic algorithms to solve the problem [38]. However, they did not consider the task merging method in the heuristic algorithms. Wu et al. adopted the ant colony optimization plus iteration local search (ACO-ILS) approach to resolve the multisatellite emergency tasks and common tasks scheduling problem [4]. However, they did not account for the uncertainty of emergency tasks. Moreover, Wang et al. presented a dynamic scheduling algorithm called TMBSR-DES, which comprehensively considers task merging, backward shift, and rehabilitation [39]. The task merging strategy in TMBSR-DES only combines new tasks. The robustness of schedule has not been considered in all of the above algorithms.

To sum up, the studies on dynamic emergency scheduling mentioned above have the following shortages.

- (1) Some dynamic scheduling methods ignored the timeliness of emergency tasks.
- (2) The real-time scheduling method [2, 37, 38] took into account the timeliness of dynamic emergency tasks,

- but it did not consider the influence of the common tasks on satellites.
- (3) Most of studies overlooked the robustness of dynamic satellite scheduling.

In this paper, we propose a robust scheduling approach oriented to both dynamic tasks and common tasks. We establish a robust satellite scheduling model to maximize the profit and robustness of the schedule. To solve the model, we design a combined algorithm HA-NSGA2 that is made up of multiobjective genetic algorithm NSGA2 and rule-based heuristic algorithm. NSGA2 is to acquire a flexible and highly robustness initial schedule in the scheduling of common tasks. Then a heuristic algorithm is designed to insert more emergency tasks dynamically. Meanwhile, to improve the revenue and resource utilization, we adopt a compact task merging strategy considering the duration of task execution in the heuristic algorithm. Based on the heuristic algorithm, NSGA2 is employed to optimize the solution so as to keep high continuous robustness for new emergency tasks.

The major contributions of this paper are summarized as follows. (1) The robustness is considered in the whole process of common tasks and dynamic emergency tasks scheduling to improve the scheduling ability of resisting disturbances. (2) To leave more opportunities for emergency tasks and improve the revenue and resource utilization, a compact task merging strategy is used. (3) Through considering the advantages of the heuristic algorithm and NSGA2 comprehensively, we propose an efficient solution to deal with the emergency tasks dynamically.

The remainder of this paper is organized as follows. The characteristics of emergency tasks and the solution framework are described in Section 2. In Section 3, we present the robust scheduling model. We introduce the corresponding solution to the model including the multiobjective genetic algorithm, rule-based heuristic algorithm, and neighborhood operator in detail, in Section 4. The simulation results and performance analysis, through comparison with other methods, are given in Section 5. Section 6 concludes the paper with some future research directions.

2. Problem Description

Earth observing satellites (EOSs), orbiting the Earth, collect images of the Earth's surface. As shown in Figure 1(a), a strip of EOSs can be formed on the ground by the subsatellite point of satellite and the field of view of the sensor, slewing angle of the sensor, and observation duration. Satellite mission scheduling allocates the limited satellite resources to observing tasks efficiently and reasonably. Figure 1(b) shows that each schedule is a sequence of tasks ordered in time for an EOS.

The EOSs work in a complex environment in which there exist many dynamic factors, as follows:

(i) The change in task properties: the change in user requirements or the reasonable task attributes will lead to a change in task priority. When the priority of the task changes, we cancel the task before it is

- performed. Then the task, which is viewed as an emergency task, is added to the emergency task set.
- (ii) The change in the state of satellites: sometimes, certain satellites may be out of use because of malfunction, memory, or energy shortage, and so on. As a result, the initial schedule cannot be executed continuously. These tasks which are affected by the satellites can be seen as emergency tasks to be inserted.
- (iii) The arrival of emergency tasks: according to the actual requirement, users may insert some new tasks when a scheduling scheme is executed. Specifically, there are some new incoming emergency tasks caused by emergent events, for example, earthquakes, fire, and landslides. Take an earthquake as an example. There is an urgent need to reasonably utilize the existing satellites to rapidly image the affected area during a short time period. Once the emergency tasks are submitted and scheduled, EOSs can acquire remote sensing images of disaster area quickly, which can provide rapid and effective information for quick investigation and assessment of earthquake damage.
- (iv) Uncertainty of weather conditions: some imaging tasks may not be completed because of a change in weather conditions. To satisfy user requirements, these tasks must be rearranged.

The four cases mentioned above can be seen as emergency tasks to be inserted into a scheduling scheme. Thus, the dynamic scheduling problem can be described as a unified form of inserting new tasks. In this paper, we research dynamic scheduling focus on new incoming emergency tasks. The general method is to produce a temporary schedule and then to adjust the schedule as quickly as possible. When dynamically adjusting the initial schedule, the solution stability is an important problem. In fact, the satellite application is a complex process and the orders can only be uploaded with the special equipment within limited visible time windows, so excessive changes may cause great operational trouble [36]. Hence, we should not only consider arranging more emergency tasks dynamically but also decrease the disturbances to the initial schedule as much as possible. Moreover, the uncertainty of arrival time is an important property of new tasks. Thus, we should consider the timeliness of emergency tasks in dynamic scheduling. Therefore, the goal of our research is to arrange emergency tasks as more as possible in smaller disturbances, considering the timeliness of the emergency tasks. As a result, the optimization objectives in this paper are revenue and robustness.

In this paper, we propose a dynamic scheduling approach that can deal with the emergency tasks in real time dynamically. It is actually a two-stage method to produce a temporary schedule and then to adjust the schedule as quickly as possible as is shown in Figure 2. The robust scheduling approach proposed in the paper is oriented to both dynamic emergency tasks and common tasks. We describe the flowchart of the solution in Figure 3. Before scheduling, pretreatment is conducted to compute the available opportunities for each task. Then we establish a robust satellite scheduling model

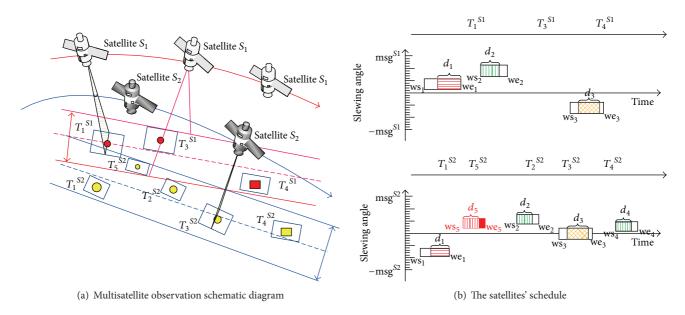


FIGURE 1: Illustration of satellite observing activity.

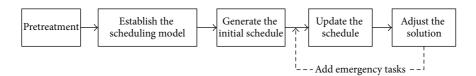


FIGURE 2: The framework of solution to the dynamic scheduling problem.

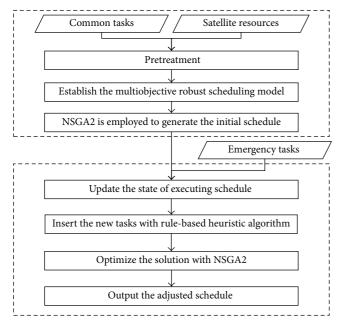


FIGURE 3: The flowchart of the dynamic scheduling approach.

with two objectives considering both revenue and robustness. To optimize objectives of the model, the multiobjective genetic algorithm NSGA2 is used to obtain robust solutions.

We can acquire an initial schedule with high revenue and stronger ability to accept emergency tasks. With the arrival of emergency tasks, we update the state of the tasks in the current executing schedule. To arrange the emergency tasks dynamically, the rule-based heuristic algorithm with compact task merging strategy is designed. Thus, we can get some temporary adjusted schedule. Lastly, NSGA2 is adopted to further improve the solution by searching the global space. It is worth mentioning that we design a neighborhood operator to initialize the population in NSGA2.

3. Model

The satellite scheduling amounts to a reasonable arrangement of satellites, sensors, time windows, and sensor slewing angle for observation tasks to maximize one or more objectives, for example, the overall observation profit, while being subject to related constraints. As a result, the satellite scheduling problem usually consists of five parts: tasks, satellite resources, opportunities, objectives, and constraints. A detailed introduction of the model is as follows.

3.1. Tasks. In this paper, we concentrate on two types of tasks: the common task and emergency task. Common imaging tasks are usually submitted to the planning system in advance. The characteristics of emergency tasks are different from those of common tasks. Emergency events usually occur

unexpectedly, with uncertainties of occurrence time and number. Thus, the emergency task number and arrival times are not known a priori. Moreover, dynamic tasks need to be completed in time, because users want to obtain observation results within certain time limits.

- (i) Common task: let $T = \{t_1, t_2, \dots, t_N\}$ be the common task set, where N is the number of common tasks. We assume that all tasks are independent, aperiodic, and nonpreemptive. Each task is associated with the weight, p_i , and the indispensable duration of task execution d_i .
- (ii) Emergency task: let $T^E = \{t_{N+1}, t_{N+2}, \dots, t_{N+N^E}\}$ be the emergency task set. Each emergency task is presented by $t_n = (p_n, d_n, \operatorname{at}_n, \operatorname{dt}_n)$, where p_n is the weight, d_n is the duration, at_n is the arrival time, and dt_n is the deadline.

In addition, to make the denotation of common tasks consistent with that of emergency tasks, we assume that each common task has an arrival time at_n = 0 and a deadline $dt_n = t_s^f$.

Considering that many tasks are commonly submitted together, thus we assume in this paper that the new tasks arrive in batch style. Let $T_S = \{t_S^0, t_S^1, \ldots, t_S^e, t_S^f\}$ be the scheduling time set, where t_S^i ($i \geq 1$) is the ith dynamic scheduling time, t_S^0 is the initial scheduling time, e is the total batch of emergency tasks, and t_S^f is the end time of scheduling.

Imaging tasks can be divided into two types: point target and area target. The type depends on the size of the target area and the attributes of the satellite, such as the field of view and height of satellite. The point target is small so that it can be observed by single observations trip. The area target, which covers a wide geographical area, cannot be photographed in a scene. Usually, the area target can be segmented into several point targets to be imaged. In this paper, we focus on dealing with point targets.

3.2. Satellite Resources. As for the resources, we consider a set $S = \{s^1, s^2, \dots, s^M\}$, where M is the satellite number. We assume that the sensors of the satellites considered in our study are able to slew laterally. Each resource $s^j \in S$ is denoted by $s^j = (\Delta \theta^j, \Delta d^j, \operatorname{sl}^j, \operatorname{st}^j, \operatorname{msg}^j, \operatorname{orb}^j, \operatorname{su}^j, \operatorname{sd}^j, \operatorname{duty}^j)$, where $\Delta \theta^j, \Delta d^j, \operatorname{sl}^j, \operatorname{msg}^j, \operatorname{orb}^j, \operatorname{su}^j, \operatorname{sd}^j, \operatorname{and} \operatorname{duty}^j$ are the field of view, the longest opening time, slewing rate, attitude stability time, maximum slewing angle, the flight time in each orbit, the start-up time of sensor, the retention time of shutdown, and the longest imaging time in each orbit.

3.3. Available Opportunities. For available opportunities, let $AO = \{ao_1^1, ao_1^2, \dots, ao_j^j, ao_2^1, \dots, ao_j^j, \dots, ao_N^M\}$ be the opportunity set of every task on each satellite resource. For $ao_i^j \in AO$, $ao_i^j = \{ao_{i1}^j, ao_{i2}^j, ao_{i3}^j, \dots, ao_{ik_{ij}}^j\}$ represents the available opportunities of task t_i on resource s^j , where k_{ij} represents the number of available opportunities. Giving an available opportunity $ao_{ik}^j \in ao_i^j$, it is represented by $ao_{ik}^j = \{[ws_{ik}^j, we_{ik}^j], \theta_{ik}^j\}$, where ws_{ik}^j and ws_{ik}^j are denoted by the start time and end time of the window ws_{ik}^j , and ws_{ik}^j is the swing angle. The set of time windows of task i on satellite i is i is i is i in satellite i is i in i in

3.4. Optimization Objectives. We assume that t_S^k is the current scheduling time, and then the model can be described as follows.

The primary objective is to maximize the revenue of all scheduled tasks:

$$\max: \sum_{i=1}^{N+N^E} \sum_{j=1}^{M} \sum_{k=1}^{K_{ij}} x_{i,k}^j p_i, \tag{1}$$

where $x_{i,k}^{j}$ is the decision variable whether the task i is scheduled:

$$x_{i,k}^{j} = \begin{cases} 1, & \text{if task } i \text{ is executed by satellite } j \text{ in the } k \text{th time window} \\ 0, & \text{otherwise.} \end{cases}$$
 (2)

The second objective is called the neighborhood-based robust indicator [40]. The robustness based on neighborhood is proposed from the advances in robust optimization of continuous functions [41, 42]. The idea is that robust optima are located on broad peaks (plateaus) in the fitness landscape, whereas brittle optima are located on narrow peaks, as shown in Figure 4. A tradeoff is made between the broadness and height of the peaks, such that a low and broad peak may be preferable to a narrow and high peak. The problem of satellite scheduling is a discrete problem. There are some schedules which are different sequence of the scheduled tasks ordered in time but have the same revenue. If only

considering the revenue, any one of these schedules is a feasible solution. However, dynamic task scheduling not only requires arranging more emergency tasks, but also requires reducing the frequency and difficulty of adjustment as much as possible. Thus, the property of resisting disturbances and self-adjusting is very important for dynamic task scheduling.

Hence, we define the neighborhood-based robustness as the total revenue of scheduled tasks that can be reassigned to other timeslots in the scheme directly. It is used to measure the ability of a scheduling scheme to rearrange scheduled tasks. The higher value of the neighborhood-based robust indicator is, the more possible it is to rearrange scheduled

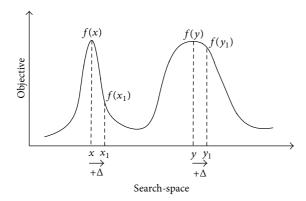


FIGURE 4: The idea in the robust optimization technique for continuous problem [47].

tasks. Therefore, we maximize the value of the neighborhood-based robust indicator:

$$\max: \sum_{i=1}^{N+N^{E}} \sum_{j=1}^{M} \sum_{k=1}^{K_{ij}} p_{i} \times x_{i,k}^{j} \times y_{i}, \tag{3}$$

where y_i is the parameter that indicates whether the task $t_i \in SS$ can be rearranged in another timeslot:

 y_i

$$= \begin{cases} 1, & \text{if task } t_i \text{ can be rearranged in another timeslot} \\ 0, & \text{otherwise.} \end{cases}$$
 (4)

- 3.5. *Constraints*. There are some constraints that are used to analyze the complex restricted condition in the model.
- (1) Each task t_i must be executed in an available opportunity ao_{ik}^j , $ao_{ik}^j \in ao_i^j$.

Hence, we have the available opportunity constraint:

$$C_{1}: \begin{cases} x_{i,k}^{j} \left(\operatorname{ts}_{i}^{j} - \operatorname{ws}_{ik}^{j} \right) \geq 0 \\ x_{i,k}^{j} \left(\operatorname{ts}_{i}^{j} + d_{i} - \operatorname{we}_{ik}^{j} \right) \leq 0 \\ a_{i} \leq t_{S}^{k} \leq \operatorname{ts}_{i}^{j} \leq \operatorname{dt}_{i}, \end{cases}$$
 (5)

where ts_i^j denotes the start time of task t_i . Actually, for a given scheduled task t_i on s^j , its starting observation time ts_i^j and observation angle θ_i^j are determined. All scheduled tasks form the current schedule $\mathsf{SS} = \bigcup_{j \in [1,M]} T^j$, where T^j is a sequence of scheduled tasks on satellite j.

(2) Each task only needs to be executed once, and the execution process does not involve preemptive service. So we have the following uniqueness constraint:

$$C_2: \sum_{j=1}^{M} \sum_{k=1}^{K_{ij}} x_{i,k}^j \le 1.$$
 (6)

(3) There should be an adequate transmission time for sensor in any two adjacent tasks t_u , t_v assigned to the same satellite s^j to finish the actions such as shutting down,

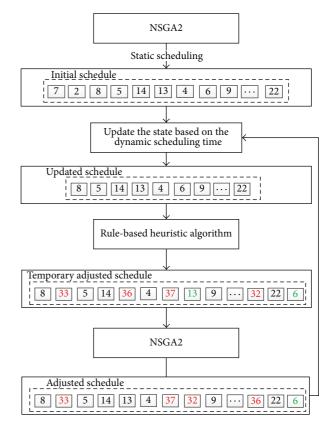


FIGURE 5: The flowchart of the algorithm.

swinging the observation angle, stabilizing gesture, and startup. So we have the switch time constraint:

$$C_3: \mathsf{te}_u^j + \mathsf{sd}^j + \mathsf{sl}^j \times \left| \theta_v^j - \theta_u^j \right| + \mathsf{st}^j + \mathsf{su}^j \le \mathsf{ts}_v^j. \tag{7}$$

(4) Another constraint is the total imaging time of any satellite, which should be less than the allowable longest imaging time of its sensor during every time period. Let the scheduling period be $[t_S^k, f]$. Hereby, we have the imaging time constraint:

$$C_4: \sum_{i \in T^j_{i_b}} d_i \le \operatorname{duty}^j, \tag{8}$$

where $T_{t_b}^J$ denotes a sequence of scheduled tasks on satellite j which flies during the time span $[t_b, t_b + \operatorname{orb}^j]$, where $t_b \in [t_s^k, f - \operatorname{orb}^j]$.

4. Algorithms

In this section, we introduce the dynamic emergency tasks scheduling algorithm HA-NSGA2 proposed in this paper. The algorithm, which is actually a combined algorithm, aims at improving the efficiency of dynamic emergency tasks scheduling. The flowchart of the algorithm is depicted in Figure 5. Firstly, the multiobjective genetic algorithm NSGA2 with two objectives, that is, revenue and robustness, is used for common tasks scheduling to obtain the optimal solution.

With the arrival of emergency tasks, the rule-based heuristic algorithm is applied to insert these emergency tasks into the initial schedule as much as possible. To improve the imaging efficiency as much as possible, we propose a compact task merging method that considers task execution duration in the rule-based heuristic algorithm. If two or more targets are geographically adjacent, they may be combined into a compact composite task by the compact task merging method. The details of compact task merging strategy are described in Appendix B. Based on the rule-based heuristic algorithm, the multiobjective genetic algorithm NSGA2 is employed to find the optimal solution. The neighborhood operator, which is embedded in NSGA2, aims at generating the initial population.

4.1. Multiobjective Genetic Algorithm. Satellites observation scheduling problem has been proven to be NP-complete. Thus, it is impossible to find polynomial time algorithms for finding optimal solutions. Genetic algorithms are the most successful algorithms for efficiently dealing with the complexity of computationally hard problems from the networking domain [23, 43]. Before introducing the multiobjective genetic algorithm NSGA2, we firstly make clear some terms, such as Pareto Dominance, Pareto optimality, Pareto Optimal Set, Pareto Front, Nondominated Sorting, and crowding distance, which will be referred to in the algorithm. We give the associative definitions in Appendix A.

As shown in Figure 6, we present the main process of NSGA2. First, we initialize the parameters and generate the initial population P_{father} . After computing the fitness function of P_{father} , we make use of the binary tournament selection mechanism, single point crossover operator, and single point mutation operator for the current individuals until the number of the next generation population P_{child} reaches the size M. Then, we set the total population as $Q = P_{\text{father}} \cup P_{\text{child}}$ by combining P_{father} and P_{child} . Next, the nondominated sorting algorithm is used to divide the population Q into different Pareto Front F_{rank} with the rank of each individual. We choose the individuals from $F_{\rm rank}$ until the number of the next population $|P_{\text{next}}|$ satisfies the request, ranking from small to large. If $|P_{\text{next}}|$ is greater than M, then we sort the current F_{rank} in order according to the crowding distance and choose the better individuals as P_{next} . Finally, if the result satisfies the termination conditions, then we output the nondominated solution set. Otherwise, we let P_{father} = P_{next} and repeat the process.

The main parts of the genetic algorithm will be described in detail including initialization, selection operator, crossover operator, mutation operator, nondominated sorting, and the computation of crowding distance.

4.1.1. Initialization. Unlike local search techniques, the genetic algorithm uses a population of individuals giving thus the search a larger scope and chances to find a better solution. In this paper, the satellite scheduling problem is encoded into a set of chromosomes. Each chromosome composed of an imaging task sequence of different satellites represents a solution and is assigned a fitness value to determine how good each chromosome is. The structure of a chromosome

is shown in Figure 7. In the algorithm, some parameters such as the size of population M, the maximum iterative number T, the probability of crossover $p_{\rm crossover}$, and mutation probability $p_{\rm mutation}$ must be set. The initial population is generated by random strategy.

4.1.2. Selection Operator. Several selection strategies have been proposed in the previous literatures such as roulette wheel selection and binary tournament selection. We adopt binary tournament selection in this paper. The main steps are to choose two individuals randomly from a population and to select the individual that dominates the others.

4.1.3. Crossover Operator. Crossover operator plays an important role in the genetic algorithm. The best genetic information of parents can be transmitted to offspring during generations of the evolution process by the crossover operator. After parent chromosomes have been selected from the population, we set the crossover probability $p_{\rm crossover}$ to allow the majority of chromosomes to mate and pass to their offspring. Figure 8 shows an example of the crossover. According to the characteristics of the satellite scheduling problem, we cross the parent chromosomes to generate child chromosomes and then delete the same genes in common chromosome. Finally we attempt to arrange the tasks differently for both parent chromosomes in child chromosomes.

4.1.4. Mutation Operator. Mutation operator intends to keep the variety of the population and improve the individuals by small local perturbations. Additionally, the mutation operator provides a component of randomness in the neighborhood of the individuals of the population. Therefore, genetic algorithm has the ability to avoid falling prematurely into local optima and eventually escapes from them during the search process. For example, if we set the probability $p_{\text{mutation}} = 0.2$, then the overall procedure of the random mutation operator on the chromosome is shown in Figure 9. We choose a scheduled task that has another window randomly such as t_8 . Then the task is rearranged in another available window.

4.1.5. *Nondominated Sorting*. The main procedure of non-dominated sorting is described in Algorithm 1 in detail.

4.1.6. Crowding Distance Computation. The main procedure of computing the crowding distance is described in Algorithm 2 in detail.

4.2. Rule-Based Heuristic Algorithm. The solution with maximum revenue and robustness, which is produced by NSGA2, is selected as the initial schedule. When the schedule is executed, a number of emergency tasks arrive. In this case, the current executing schedule must be modified to arrange these emergency tasks.

When a batch of emergency tasks arrives, based on the current scheduling time, tasks may have different states: finished task, running task, waiting task, and new task. In the example shown in Figure 10, t_S^k is the current scheduling time. If task i has been executed before the current time, then task

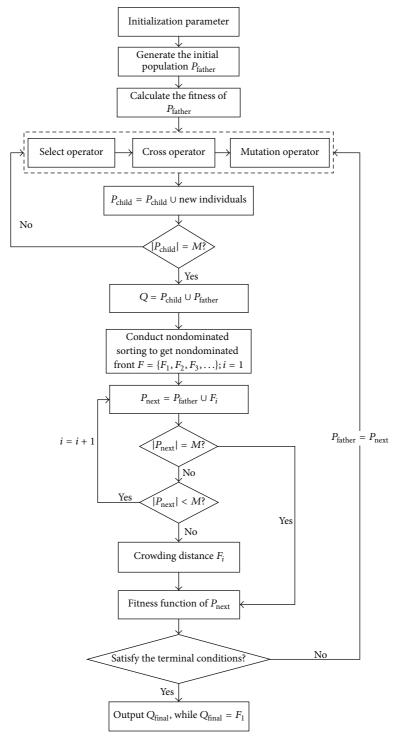


FIGURE 6: The main process of NSGA-2.

i is a finished task while $te_i < t_S^k$. If $ts_i < t_S^k < te_i$, then task i is a running task. Task i is a waiting task if $ts_i > t_S^k$. If task i is an emergent task that has not been planned, then task i is a new task.

According to the analysis on the state of the tasks, the waiting tasks in the current schedule can be adjusted. Focusing on the new emergency tasks, we assign these

tasks to the current schedule. In this paper, we employ a rule-based heuristic algorithm considering compact task merging (CTM-DAHA) to insert the new tasks dynamically [44]. The main idea of the heuristic algorithm is to assign the most conflict-free time window to the task most in need. With CTM-DAHA, we obtain the temporary adjusted schedules.

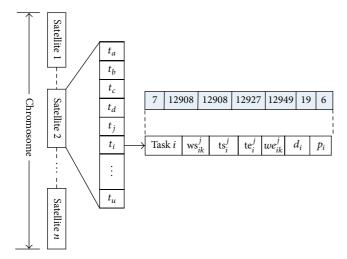


FIGURE 7: The structure of chromosome.

To describe the dynamic adjusting rule-based heuristic algorithm considering compact task merging clearly, we first give some definitions.

Definition 1. The need of the task indicates how badly the task needs to be performed [30, 44]:

$$N(t_i) = \frac{p_i}{|AO_i|}. (9)$$

Definition 2. The contention of the task window is measured by counting the number of tasks that need the time window, weighted by the weights of the tasks:

$$C(w) = \sum_{t_i \in T(w)} p(t_i), \qquad (10)$$

where T(w) is the set of tasks that could be executed at any moment within time window w and $p(t_i)$ is the weight of the tasks t_i .

The dynamic adjusting rule-based heuristic algorithm CTM-DAHA is presented as shown in Figure 11 [44]. To accommodate emergency tasks in the initial scheme, firstly we update the state of tasks in the schedule. Then, all new tasks are inserted into a waiting queue according to the needs of the tasks $N(t_i)$, from high to low. The time windows of each task are ranked from low to high according to the contention C(w). For a task, we first judge whether it can merge with any other existing task in the current schedule. If all time windows of the task fail to be merged, we attempt to insert the task directly if it does not conflict with any other task. If direct insertion fails, we judge whether the task can be inserted by rearranging the conflict tasks. If direct insertion and insertion by rearranging do not succeed, some tasks may be deleted for inserting an emergency task. Then the entire process will repeat until the waiting queue is empty. Take, for example, a current executing schedule as shown in Figure 12(a). With the arrival of emergency tasks, that is, $T^E = \{t_8, t_9, t_{10}, t_{11}\},\$ they should be arranged in time. Task t_8 can be combined

with t_2 by compact task merging. Task t_9 can be inserted directly. Task t_{10} is inserted into the schedule by rearranging the conflict task t_5 . Task t_7 is deleted from the schedule so that task t_{11} with higher weight can be inserted. Figure 12(b) shows the planning task list of the adjusted schedule.

4.3. Neighborhood Operator. The solution obtained by CTM-DAHA still needs to be optimized. Based on the temporary adjusted schedule, we use NSGA2 to find the optimal solution. Different from the multiobjective genetic algorithm described in Section 4.1, we design a neighborhood operator to generate the initial population.

The main idea of the neighborhood operator is to search all neighborhoods of the current schedule. As shown in Figure 13, the current schedule SS is represented by the red point and the yellow points denote the neighborhoods of SS. Specifically, we change the locus of a task to move it to another position, as shown in Figure 14.

The procedure of the neighborhood operator is described as shown in Algorithm 3.

After initializing the population using the neighborhood operator, the next serial of operation is selection, crossover, mutation, and nondominated sorting as expressed in Section 4.1. Finally, using the multiobjective genetic algorithm NSGA2, we obtain the optimal solution as the new adjusted schedule.

5. Experimental Simulation and Discussion

In this section, we evaluate the performance of the proposed algorithm HA-NSGA2. To reveal the advantages of the algorithm, we compare it with RBHA-DTSP [36] and CTM-DAHA [44]. The two methods both use multiobjective genetic algorithm to make the initial schedule for common tasks. To evaluate the advantage of the compact task merging method (CTM), we perform experiments to compare the scheduling results obtained by HA-NSGA2 with the two task merging strategies: traditional task merging strategy (TTM) and compact task merging method (CTM). Specifically, we embed the two task merging methods into HA-NSGA2, respectively, so as to prove the effectiveness of compact task merging strategy.

- (i) RBHA-DTSP: the basic idea of RBHA-DTSP is to carry out some amount of iterative repair search inside each emergency task's available opportunities. With the repair search for a given emergency task, some rules are adopted to decide which tasks should be retracted [36].
- (ii) CTM-DAHA: this is a dynamic adjusting rule-based heuristic algorithm considering compact task merging (CTM-DAHA) which is designed to insert the emergency tasks into the initial schedule.

To evaluate the result we use the following performance metrics.

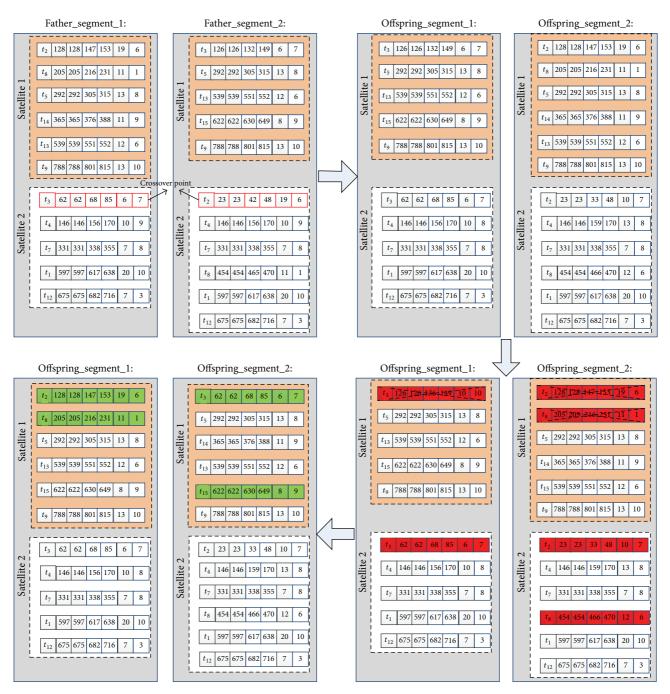


FIGURE 8: Crossover operator.

Emergency task revenue (ETR) can be described as

ETR =
$$\sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{K_{ij}} x_{i,k}^{j} p_{i}.$$
 (11)

Satisfaction ratio (SR) is defined as SR = total number of tasks that can be arranged in the scheme/total number of tasks.

Scheduled tasks robustness (STR) can be defined as

STR =
$$\sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{K_{ij}} p_i \times x_{i,k}^j \times y_i$$
. (12)

In fact, the satellite application is a complex process and the orders can only be uploaded with special equipment within limited visible time windows, so it is important to make the least adjustment to keep the stability of the schedule as well as to insert more tasks to keep the efficiency of the

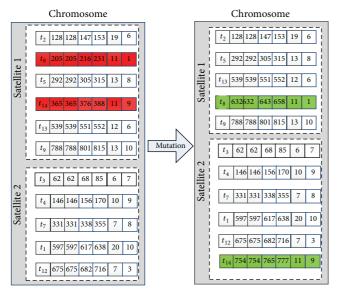


FIGURE 9: Mutation operator.

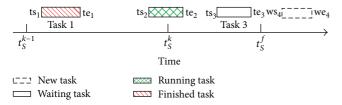


FIGURE 10: The states of tasks.

schedule. Therefore, the main measurement of perturbation needs to be considered.

Perturbation measurement (PM) is a metric used to describe the distance between the adjusted scheme and the initial scheme. PM can be defined with the total number of initial tasks that change in the adjusted scheme compared with the initial scheme.

5.1. Simulation Method and Parameters. Without loss of generality, we acquire the imaging targets randomly in the area: latitude $-180^{\circ} \sim 180^{\circ}$ and longitude $-65^{\circ} \sim 65^{\circ}$. The scheduling period is 24 h, which can be denoted by [0, 86400]. The number of common tasks is 200. In addition, there are five batches of new coming emergency tasks and each batch is given the same size: 50. The arrival time of a batch t_S^k is a random positive number distributed in $[t_S^0, t_S^T]$. In the experiment, we simulate 5 batches of emergency tasks and T_S is set as $T_S = \{t_S^0, t_S^1, \dots, t_S^e, t_S^f\} =$ {0, 59, 13300, 26100, 41800, 65000, 86400}. Moreover, the deadline of an emergency task is the end time of scheduling: 86400. Without loss of generality, the weights of all tasks are randomly distributed in [1, 10]. Two sensors on different satellites are simulated to accomplish the tasks. The parameters of sensors are shown in Table 1. The available opportunities, including time windows and slewing angles associated with tasks, are computed using STK (Satellite

TABLE 1: Parameters of satellites.

Parameters	Satellite s ¹	Satellite s ²
Average height (km)	780.099	505.984
Orbital inclination (degree)	98.5	97.421
Local descending node time (am)	10:30	10:30
Argument of perigee	90	90
Eccentricity ratio (degree)	0.0011	0
Nodical period (min)	100.38	97.716
Maximum slewing angle (degree)	±32	±32
Field of view (degree)	8.4	6

Tool Kit). The time window is removed if its span is shorter than the duration of the corresponding task or its finished time is later than the deadline. Moreover, we assume that the maximum time for each satellite to open its sensor once is 60 seconds and that the longest imaging time in any period time of orb^j is 150 seconds.

5.2. The Simulation Result. We carry out two groups of experiments in this section. We investigate the performance impact of different algorithms, that is, RBHA-DTSP, CTM-DAHA, and HA-NSGA2, in the first group. The experimental results are depicted in Figure 15. The second group of experiments is designed to evaluate the efficiency of the task merging method: CTM and TTM.

Figure 15(a) shows the result of the revenue of scheduled emergency tasks in each adjusted schedule. Among the algorithms, HA-NSGA2 shows the best scheduling revenue. The explanation is that HA-NSGA2 optimizes the robustness and uses the compact task merging strategy. In contrast, CTM-DAHA does not improve the robustness and RBHA-DTSP does not consider task merging. Based on Figure 15(b), we can see that, with more batches of emergency tasks, the robustness of each adjusted schedule decreases. This can be attributed to the fact that, with the time of scheduling increasing, the scheduling period becomes shorter. This means that the tasks' available opportunities in a schedule decrease and the number of tasks which can be rearranged decreases. As a result, the robustness becomes smaller. In addition, in each scheduling, HA-NSGA2 can generate a more robust schedule except the last scheduling. In the last scheduling, the schedule obtained by HA-NSGA2 shows the lowest robustness because HA-NSGA2 scheduled more tasks and less spare time is left for other tasks to be rearranged. As a result, the robustness is worse. Figure 15(c) presents the satisfaction ratio of every scheduling result. It is found that the satisfaction ratio of all algorithms decreases as the arrival of more emergency tasks. This can be attributed to the fact that the number of scheduled tasks decreases compared with the increasing emergency tasks. Moreover, HA-NSGA2 shows the highest satisfaction ratio among all of the algorithms. Table 2 describes the number of scheduled emergency tasks and the number of adjusted tasks in the adjusted schedules. It is found that there is little difference of disturbance among the algorithms. On the whole, the ability

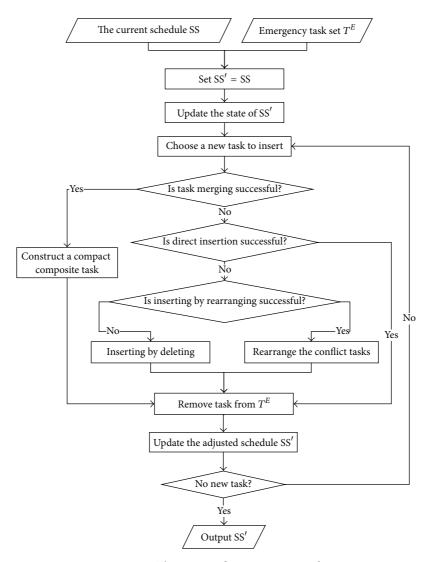


FIGURE 11: The process of inserting a new task.

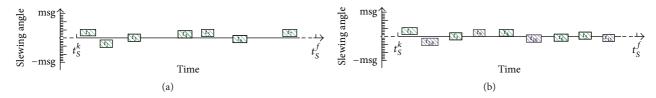


FIGURE 12: (a) The current executing schedule. (b) The adjusted schedule.

dealing with emergency tasks of HA-NSGA2 is superior to that of the other algorithms.

Figure 16 shows the performance impact of the two task merging methods. Figure 16(a) demonstrates that, with the increase in scheduling time, the compact task merging method (CTM) can get a higher total emergency task revenue. This result indicates that more emergency tasks can be arranged by the compact task merging method and more opportunities can be left for other emergency tasks. From Figure 16(b), it is found that robustness of the schedule of the compact task merging method is lower than that

of the traditional task merging strategy. This result can be attributed to two factors. First, the algorithm with the compact task merging strategy can arrange more emergency tasks compared with traditional task merging method. The other is the composite task, which is unable to participate in the robustness computing because there is only one synthetic available window, acquired by compact task merging. Moreover, the number of composite tasks by compact task merging is more than that by traditional task merging. So the compact task merging strategy is better than the traditional merging strategy in the available windows and the satellite

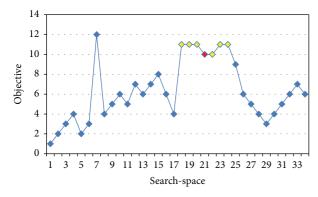


FIGURE 13: The neighborhods of the current schedule.

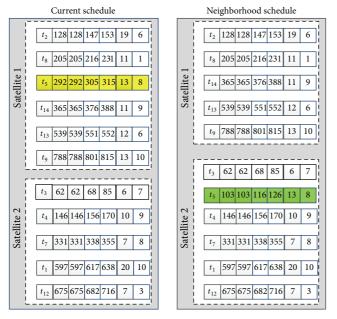


FIGURE 14: Illustration of neighborhood operator.

TABLE 2: The perturbation measurements of different algorithms.

Batch	RBHA	RBHA-DTSP		DAHA	HA-NSGA2		
	N	PM	N	PM	N	PM	
1	47	2	47	2	48	4	
2	46	5	46	5	47	5	
3	41	3	41	1	42	1	
4	43	3	46	4	47	4	
5	43	1	42	1	44	1	

Note: N: the number of scheduled emergency tasks.

resources left. Figure 16(c) depicts that the compact task merging method can arrange more emergencies and obtain a more satisfactory schedule. Table 3 describes the way of adjusting of emergency tasks and the distance of the adjusted scheme from the initial scheme. In the same perturbation measurement, HA-NSGA2 with the compact task merging method can arrange more emergency tasks. Also, CTM can merge more tasks than CTM. The result proves that

TABLE 3: The perturbation measurements of different task merging strategies.

D . 1	36.1.1		777. 6		> TTD		D) (
Batch	Method	N	TM	NI	NIR	NID	PM
1	CTM	48	4	42	1	1	4
1	TTM	47	1	44	1	1	4
2	CTM	47	4	38	3	2	5
2	TTM	46	1	41	2	2	6
3	CTM	42	3	38	1	0	1
	TTM	41	2	38	1	0	2
4	CTM	47	7	36	1	3	4
-	TTM	46	1	39	1	4	6
5	CTM	44	3	40	0	1	1
3	TTM	42	1	43	0	0	1

Note: N: the number of scheduled emergency tasks, PM: the number of adjusted tasks in the schedules, NITM: the number of emergency tasks inserted by task merging, NI: the number of emergency tasks inserted directly, NIR: the number of emergency tasks inserted by rearranging, and NID: the number of emergency tasks inserted by deleting.

the compact task merging strategy has an advantage over the existing traditional method. The compact task merging method can improve the chances of inserting emergency tasks.

5.3. Discussion. Based on the extensive experiments, we can evaluate the performance of HA-NSGA2 proposed in the paper. Using the algorithm HA-NSGA2 with two objectives, that is, the revenue and robustness, we can obtain the scheduling scheme that has higher revenue and stronger ability to resist disturbance in emergency tasks scheduling. HA-NSGA2 algorithm can insert more emergency tasks while keeping the high revenue as is shown in Figure 15. Moreover, HA-NSGA2 can significantly reduce the scheduling disturbance in the process of scheduling emergency tasks. The experiments prove that the HA-NSGA2 is capable of finding optimal solutions with the ability to resist disturbance and to address emergency tasks efficiently. Moreover, the compact task merging strategy can improve the chances of merging multiple tasks and save the satellite resources. So the method proposed in this paper is more practical and satisfies the requests of the users.

6. Conclusion and Future Work

To solve dynamic scheduling problem for emergency tasks in a multisatellite observation system, we propose a mixed algorithm named HA-NSGA2 to improve the ability of emergency tasks scheduling from two aspects. In our algorithm two optimization objectives, that is, revenue and robustness, are considered continuously. Firstly, we schedule common tasks by using the multiobjective genetic algorithm NSGA2 to obtain an initial solution with high revenue and robustness. Then, the heuristic algorithm which embeds a compact task merging strategy is used to handle emergency tasks. Finally, to optimize the solution, NAGA2, which embeds a novel neighborhood operator to generate the initial population,

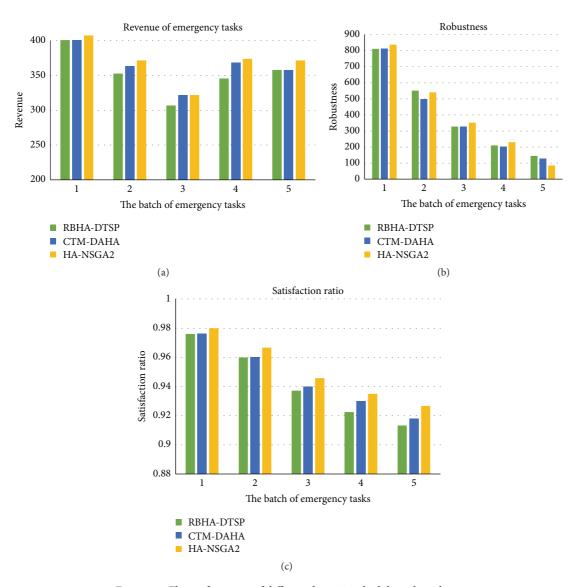


Figure 15: The performance of different dynamic scheduling algorithms.

is adopted to search the optimal result in the local scope. Extensive experimental simulation proves that the robustness is an important factor in dynamic scheduling. The algorithm HA-NSGA2 proposed in this paper optimizes the robustness and obtains the more robust scheduling results with the highest revenue compared with other algorithms. Moreover, the compact task merging strategy can improve the chances of merging multiple tasks. Therefore, we conclude that the robustness of the schedule has considerable significance in the dynamic scheduling problem.

In the future, our studies will address the following issues: first, we will conduct some research on area targets and consider more critical issues, such as resource availability, data dimension, and communication. In addition, we will improve our scheduling model by considering more constraints, such as memory capacity, energy consumption, and speed of upload. Finally, we will extend our experiments in real scene.

Appendices

To help the readers understand, we make some terms clear that will be referred to in NSGA2 and describe the compact task merging strategy in detail in these appendices. In Appendix A we introduce some definitions of the multiobjective genetic algorithm NSGA2. Compact task merging strategy is presented in Appendix B.

A. The Associative Definitions for NSGA2

Some fundamental definitions of the multiobjective genetic algorithm NSGA2 are given in the following [44, 45].

Definition for Pareto Dominance. For any $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^m$: \mathbf{u} dominates \mathbf{v} (in symbols $\mathbf{u} \prec \mathbf{v}$) if and only if: $\forall i = 1, ..., m : \mathbf{u} \leq \mathbf{v}$ and $\exists i \in \{1, ..., m\} : \mathbf{u} < \mathbf{v}$.

```
for each p \in Q // every individual in population p \in Q
  S_p = \emptyset;
                   // let the set of individuals dominated by p be null
n_p = 0;
                   // the number of individuals dominating p be 0
for each q \in Q
  if (p < q) then
                                // if p dominates q
        S_p = S_p \cup q; add q into S_p
                           // if q dominates p
  else if (p > q) then
        n_p = n_p + 1;
end for;
end for;
for each p \in Q
  if (n_p == 0) then
                           // the rank of p is 1st
     p_{\text{rank}} = 1;
     F_1=F_1\cup p;
                           // add p into the set of the first non-dominated front F_1
end for;
i = 1;
                            // initialize the rank of the non-dominated fronts
while (F_i \neq \emptyset)
  for each p \in F_i
     for each q \in S_p
        n_q = n_q - 1; // the number of individuals dominating p is reduced by one
           if (n_a == 0) then // if there is no individual dominating q
             q_{\text{rank}} = i + 1;
                                 // the rank of q plus one
             F_{i+1} = F_{i+1} \cup q; // add q into the non-dominated front F_{i+1}
     end for:
  end for;
i = i + 1;
end while.
```

Algorithm 1

```
I = |F_{\text{rank}}|;
                                                     // number of solutions in F_{\rm rank}
for each p \in F_{rank}
F_{\text{rank}}[p]_{\text{distance}} = 0;
                                                      // initialize the distance
end for;
for each objective m
    F_{\text{rank},m} = \text{sort}(F_{\text{rank}}, m);
                                                     // sort using each objective value
    for each i = 2 \rightarrow I - 1
        F_{\mathrm{rank},m}[i]_{\mathrm{distance}} = (F_{\mathrm{rank},m}[i+1] - F_{\mathrm{rank},m}[i-1])/(F_{\mathrm{rank},m}^{\mathrm{max}} - F_{\mathrm{rank},m}^{\mathrm{min}}); //\mathrm{crowding-distance} \ \ \mathrm{of} \ \mathrm{individual} \ \ i \ i \ m_{\mathrm{-}} \mathrm{th} \ \mathrm{objective}
        end for;
end for;
for each p \in F_{rank}
    for each objective m
        F_{\text{rank}}[p]_{\text{distance}} = F_{\text{rank}}[p]_{\text{distance}} + F_{\text{rank},m}[p]_{\text{distance}}; // crowding-distance of p
    end for;
end for.
```

```
(1) According to the arrival time of a batch of emergency tasks, select the waiting tasks in the current schedule SS as T^w
(2) for all i \in [1, 2, ..., |T^w|], do
(3) for all j \in [1, ..., M], do
(4) compute all available opportunities for task i on satellite j
(5) end for
(6) compute the number of available opportunities |W_i| for task i
(7) end for
(8) let the neighborhood set of the current schedule SS^N be SS^N = SS
(9) while T^w \neq \emptyset do
       take the serial number i of the first task in T^w
(10)
       if |W_i| \le 1 do
(11)
(12)
          remove t_i from T^w and then go to Step (9)
(13)
       end if
(14)
       else
         try to change the locus of task i in the chromosome
(15)
         if succeed, update the chromosome and add the chromosome to SS^N
(16)
(17)
         remove t_i from T^w and then go to Step (9)
(18)
(19)
         else
          remove t_i from T^w and then go to Step (9)
(20)
(21) end while
(22) output the neighborhood set of the current schedule SS^N
```

Algorithm 3

Definition for Pareto Optimality. The idea that a candidate solution $\mathbf{x} \in \Omega$ is the optimal solution of Pareto means that $\mathbf{x} \in \Omega$ makes $F(\mathbf{x}') \prec F(\mathbf{x})$. \mathbf{x} is a K-dimensional decision variable in the decision space, and $F(\mathbf{x})$ is an objective space.

Definition for Pareto Optimal Set. Consider

$$P = \left\{ x \in \Omega \mid \neg \exists x' \in \Omega \text{ st } F(x') \prec F(x) \right\}. \tag{A.1}$$

Definition for Pareto Front. Consider

$$PF = \{F(x) \mid x \in P\}.$$
 (A.2)

Definition for Nondominated Sorting. Nondominated sorting is used to confirm the rank of each individual in the whole population according to the dominated space. For example, $\forall x_i \in \Omega$, we can compute a dominated set of x_i , $P_{x_i} = \{x \in \Omega \mid \forall x' \in \Omega, \operatorname{st}: F(x') \prec F(x_i)\}$. The size of P_{x_i} is the rank of x_i , $\operatorname{Rank}(x_i) = |P_{x_i}|$. We divide the population into different Pareto Front with the rank of each individual, $\Omega = \{F_1 \cup F_2 \cup \cdots \cup F_{\max_rank}\}$.

Definition for Crowding Distance. The crowding distance approaches aim to obtain a uniform spread of solutions along a similar Pareto Front without using a fitness sharing parameter. In this paper, there are two objectives, that is, revenue and robustness. Firstly, we rank all individuals in order according to the revenue and then compute the distance of each individual in the sequence by formula $\operatorname{dis}_{-k}(x_i)$

 $(f(x_{i+1})-f(x_{i+1}))/(f_k^{\max}-f_k^{\min}).$ Thus, the crowding distance can be defined as

Distance
$$(x_i) = \frac{f_{\text{revenue}}(x_{i+1}) - f_{\text{revenue}}(x_{i-1})}{f_{\text{revenue}}^{\text{max}} - f_{\text{revenue}}^{\text{min}}} + \frac{f_{\text{robust}}(x_{i+1}) - f_{\text{robust}}(x_{i-1})}{f_{\text{robust}}^{\text{max}} - f_{\text{robust}}^{\text{min}}}.$$
(A.3)

Definition for Fitness Function. The determination of an appropriate fitness function is crucial to the performance of the genetic algorithm. Thus, we should construct optimization objectives with a weight sum to represent the fitness function [46]:

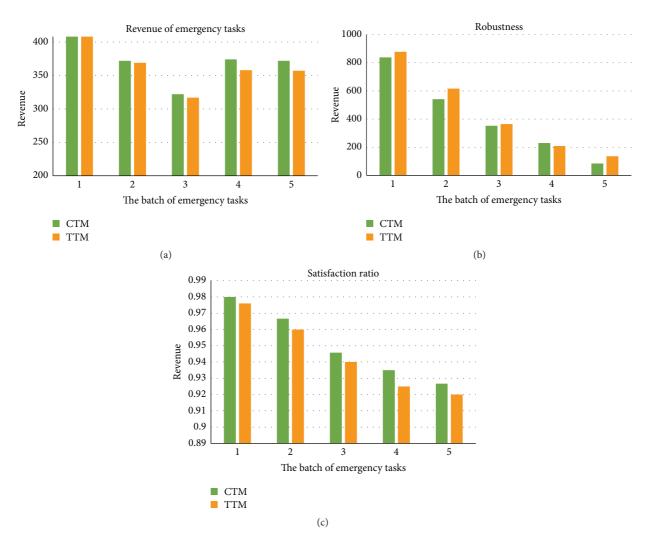
$$F = w_{\text{revenue}} \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{K_{ij}} x_{i,k}^{j} p_{i}$$

$$+ w_{\text{robustness}} \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{K_{ij}} p_{i} \times x_{i,k}^{j} \times \text{Rearrange}(t_{i}),$$
(A.4)

where F is the fitness and w_{revenue} and $w_{\text{robustness}}$ are weights.

B. The Compact Task Merging Strategy

The compact task merging strategy, which is embedded in the rule-based heuristic algorithm, is used to improve the imaging efficiency. If two or more targets are geographically adjacent, we can rationally tune the slewing angle and the



 $\label{figure 16} \mbox{Figure 16: The performances of different task merging methods.}$

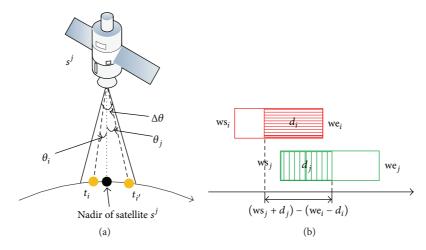


FIGURE 17: Task merging constraints.

observation duration of the sensor to enable an observation strip to cover them. In other words, tasks in the same swath of a sensor may be merged into one composite task. To improve the imaging opportunities of the new tasks, a task merging strategy is required. Suppose there are two tasks t_i and t_j that could be imaged by satellite j. And $ao_i = \{[ws_i, we_i], \hat{\theta}_i\} \in$ AO_i^J is an available opportunity of task t_i . Accordingly, ao t_i $\{[ws_j, we_j], \theta_j\} \in AO_j^j$ is an available opportunity of task t_i . Since the length of a visible time window must be larger than the observation duration of task, there often exists some unnecessary time to finish merged tasks according to the traditional task merging strategy. Therefore, the duration of task execution is an important factor in task merging. By considering the duration of task execution, we employ a new compact task merging method to construct the compact composite tasks in this paper.

When a task merging mechanism is embedded into the schedule scheme, we must judge when two tasks can be combined into a composite task and determine how to construct a composite task. Without loss of generality, between two tasks t_i and t_j , the window start time ws_i of task t_i is assumed to be no later than that of task t_i in the following.

Theorem B.1. Two feasible tasks t_i and t_j can be combined into a compact composite task $t_{i,j}$ if and only if they satisfy

$$(ws_i + d_i) - (we_i - d_i) \le \Delta d, \tag{B.1}$$

$$\left|\theta_i - \theta_i\right| \le \Delta\theta. \tag{B.2}$$

Equation (B.1) is the time window constraint. As shown in Figure 17(b), we illustrate the case where two time windows of tasks intersect.

Equation (B.2) is the slewing angle constraint. It is shown in Figure 17(a) that tasks t_i and t_j must be located in the same swath of the sensor.

Theorem B.2. If two feasible tasks t_i and t_j can be merged into a compact composite task $t_{i,j}$, then its time window $W_{i,j} = [ws_{i,j}, we_{i,j}]$ should range from

$$ws_{i,j} = \begin{cases} we_i - d_i, & \text{if } |W_i \cap W_j| \le \min(d_i, d_j) \\ \min\{ws_j, \max(ws_i, ws_j + d_j - d_i)\}, & \text{else} \end{cases}$$
(B.3)

to

$$we_{i,j} = \begin{cases} w_j + d_j, & \text{if } |W_i \cap W_j| \le \min(d_i, d_j) \\ \min(we_i - d_i, we_j - d_j) + \max(d_i, d_j), & \text{else.} \end{cases}$$
(B.4)

Its indispensable duration of task execution should be

$$\begin{aligned} &d_{i,j} \\ &= \begin{cases} ws_j + d_j - \left(we_i - d_i\right), & if \ \left|W_i \cap W_j\right| \leq \min\left(d_i, d_j\right) \\ \max\left(d_i, d_j\right), & else \end{cases} \end{aligned} \tag{B.5}$$

and the slewing angle is given by

$$\begin{split} \theta_{i,j} \\ &= \begin{cases} \max\left\{\theta_i - \frac{\Delta\theta_s}{2}, 0\right\}, & if \ \theta_i \geq 0, \ \left|\theta_i\right| \geq \left|\theta_j\right| \\ \min\left\{\theta_i + \frac{\Delta\theta_s}{2}, 0\right\}, & if \ \theta_i < 0, \ \left|\theta_i\right| \geq \left|\theta_j\right| \end{cases} & (B.6) \\ \max\left\{\theta_j - \frac{\Delta\theta_s}{2}, 0\right\}, & if \ \theta_j \geq 0, \ \left|\theta_j\right| > \left|\theta_i\right| \\ \min\left\{\theta_j + \frac{\Delta\theta_s}{2}, 0\right\}, & if \ \theta_j < 0, \ \left|\theta_j\right| > \left|\theta_i\right|. \end{cases} \end{split}$$

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 Basic Research Program of China (no. 2011CB707102), the National Natural Science Foundation of China (no. 41571334), and the Fundamental Research Funds for the Central Universities.

References

- [1] N. Bianchessi, J.-F. Cordeau, J. Desrosiers, G. Laporte, and V. Raymond, "A heuristic for the multi-satellite, multi-orbit and multi-user management of Earth observation satellites," *European Journal of Operational Research*, vol. 177, no. 2, pp. 750–762, 2007.
- [2] Q. Dishan, H. Chuan, L. Jin, and M. Manhao, "A dynamic scheduling method of earth-observing satellites by employing rolling horizon strategy," *The Scientific World Journal*, vol. 2013, Article ID 304047, 11 pages, 2013.

- [3] P. Wang, G. Reinelt, P. Gao, and Y. Tan, "A model, a heuristic and a decision support system to solve the scheduling problem of an earth observing satellite constellation," *Computers & Industrial Engineering*, vol. 61, no. 2, pp. 322–335, 2011.
- [4] G. Wu, M. Ma, J. Zhu, and D. Qiu, "Multi-satellite observation integrated scheduling method oriented to emergency tasks and common tasks," *Journal of Systems Engineering and Electronics*, vol. 23, no. 5, pp. 723–733, 2012.
- [5] N. G. Hall and M. J. Magazine, "Maximizing the value of a space mission," *European Journal of Operational Research*, vol. 78, no. 2, pp. 224–241, 1994.
- [6] E. Bensana, G. Verfaillie, J. Agnese, N. Bataille, and D. Blumstein, "Exact & INEXACT methods for daily management of earth observation satellit," in *Proceedings of the Space Mission Operations and Ground Data Systems (SpaceOps* '96), pp. 394–507, 1996.
- [7] M. A. A. Mansour and M. M. Dessouky, "A genetic algorithm approach for solving the daily photograph selection problem of the SPOT5 satellite," *Computers & Industrial Engineering*, vol. 58, no. 3, pp. 509–520, 2010.
- [8] D. Habet, M. Vasquez, and Y. Vimont, "Bounding the optimum for the problem of scheduling the photographs of an agile earth observing satellite," *Computational Optimization and Applications*, vol. 47, no. 2, pp. 307–333, 2010.
- [9] L. Barbulescu, A. E. Howe, J. P. Watson, and L. D. Whitley, "Satellite range scheduling: a comparison of genetic, heuristic and local search," in *Parallel Problem Solving from Nature— PPSN VII*, vol. 2439 of *Lecture Notes in Computer Science*, pp. 611–620, Springer, Berlin, Germany, 2002.
- [10] M. Lemaître, G. Verfaillie, F. Jouhaud, J.-M. Lachiver, and N. Bataille, "Selecting and scheduling observations of agile satellites," *Aerospace Science and Technology*, vol. 6, no. 5, pp. 367–381, 2002.
- [11] W.-C. Lin and S.-C. Chang, "Hybrid algorithms for satellite imaging scheduling," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2518–2523, IEEE, Waikoloa Village, Hawaii, USA, October 2005.
- [12] W.-C. Lin, D.-Y. Liao, C.-Y. Liu, and Y.-Y. Lee, "Daily imaging scheduling of an earth observation satellite," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, vol. 35, no. 2, pp. 213–223, 2005.
- [13] W.-C. Lin and D.-Y. Liao, "A tabu search algorithm for satellite imaging scheduling," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '04)*, pp. 1601–1606, October 2004.
- [14] W. J. Wolfe and S. E. Sorensen, "Three scheduling algorithms applied to the earth observing systems domain," *Management Science*, vol. 46, no. 1, pp. 148–168, 2000.
- [15] V. Gabrel and C. Murat, "Mathematical programming for earth observation satellite mission planning," in *Operations Research* in *Space and Air*, pp. 103–122, Springer, New York, NY, USA, 2003.
- [16] M. Vasquez and J.-K. Hao, "A 'logic-constrained' knapsack formulation and a tabu algorithm for the daily photograph scheduling of an earth observation satellite," *Computational Optimization and Applications*, vol. 20, no. 2, pp. 137–157, 2001.
- [17] M. Vasquez and J.-K. Hao, "Upper bounds for the SPOT 5 daily photograph scheduling problem," *Journal of Combinatorial Optimization*, vol. 7, no. 1, pp. 87–103, 2003.
- [18] M. Zweben, E. Davis, B. Daun, and M. J. Deale, "Scheduling and rescheduling with iterative repair," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 6, pp. 1588–1596, 1993.

- [19] J. Frank, A. Jonsson, R. Morris, and D. E. Smith, "Planning and scheduling for fleets of earth observing satellites," in *Proceedings* of the 6th International Symposium on Artificial Intelligence, Robotics, Automation and Space, Montreal, Canada, June 2001.
- [20] N. Bianchessi and G. Righini, "Planning and scheduling algorithms for the COSMO-SkyMed constellation," Aerospace Science and Technology, vol. 12, no. 7, pp. 535–544, 2008.
- [21] P. Wang and G. Reinelt, "A heuristic for an earth observing satellite constellation scheduling problem with download considerations," *Electronic Notes in Discrete Mathematics*, vol. 36, pp. 711–718, 2010.
- [22] F. Marinelli, S. Nocella, F. Rossi, and S. Smriglio, "A Lagrangian heuristic for satellite range scheduling with resource constraints," *Computers & Operations Research*, vol. 38, no. 11, pp. 1572–1583, 2011.
- [23] L. Barbulescu, J.-P. Watson, L. D. Whitley, and A. E. Howe, "Scheduling space-ground communications for the air force satellite control network," *Journal of Scheduling*, vol. 7, no. 1, pp. 7–34, 2004.
- [24] J. Wang, N. Jing, J. Li, and Z. H. Chen, "A multi-objective imaging scheduling approach for earth observing satellites," in Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (GECCO '07), pp. 2211–2218, London, UK, 2007.
- [25] S.-W. Baek, S.-M. Han, K.-R. Cho et al., "Development of a scheduling algorithm and GUI for autonomous satellite missions," *Acta Astronautica*, vol. 68, no. 7-8, pp. 1396–1402, 2011.
- [26] Y. Chen, D. Zhang, M. Zhou, and H. Zou, "Multi-satellite observation scheduling algorithm based on hybrid genetic particle swarm optimization," in *Advances in Information Technology and Industry Applications*, vol. 136 of *Lecture Notes in Electrical Engineering*, pp. 441–448, Springer, Berlin, Germany, 2012.
- [27] Z. Zhang, N. Zhang, and Z. Feng, "Multi-satellite control resource scheduling based on ant colony optimization," *Expert Systems with Applications*, vol. 41, no. 6, pp. 2816–2823, 2014.
- [28] X. Liu, B. Bai, Y. Chen, and Y. Feng, "Multi satellites scheduling algorithm based on task merging mechanism," *Applied Mathematics and Computation*, vol. 230, pp. 687–700, 2014.
- [29] G. Wu, J. Liu, M. Ma, and D. Qiu, "A two-phase scheduling method with the consideration of task clustering for earth observing satellites," *Computers & Operations Research*, vol. 40, no. 7, pp. 1884–1894, 2013.
- [30] D. Qiu, L. Zhang, J. Zhu, and H. Li, "FFFS-DTMB and ADTPC-DTMB algorithmin multi-satellites mission planning," Acta Aeronautica et Astronautica Sinica, vol. 30, pp. 2178–2184, 2009.
- [31] A. Globus, J. Crawford, J. Lohn, and A. Pryor, "A comparison of techniques for scheduling earth observing satellites," in Proceedings of the 16th Conference on Innovative Applications of Artificial Intelligence (IAAI '04), pp. 836–843, AAAI, San Jose, Calif, USA, July 2004.
- [32] A. Globus, J. Crawford, J. Lohn, and A. Pryor, "Scheduling earth observing satellites with evolutionary algorithms," in *Proceedings of the International Conference on Space Mission Challenges for Information Technology (SMC-IT '03)*, Pasadena, Calif, USA, July 2003.
- [33] J. C. Pemberton and L. G. Greenwald, "On the need for dynamic scheduling of imaging satellites," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, pp. 165–171, 2002.
- [34] L. A. Kramer and S. F. Smith, "Maximizing flexibility: a retraction heuristic for oversubscribed scheduling problems,"

- in Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI '03), pp. 1218–1223, August 2003.
- [35] G. Verfaillie and T. Schiex, "Solution reuse in dynamic constraint satisfaction problems," in *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI '94)*, pp. 307–312, Seattle, Wash, USA, July 1994.
- [36] J. Wang, J. Li, and Y. Tan, "Study on heuristic algorithm for dynamic scheduling problem of earth observing satellites," in Proceedings of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD '07), pp. 9–14, Qingdao, China, July 2007.
- [37] J. Wang, X. Zhu, L. T. Yang, J. Zhu, and M. Ma, "Towards dynamic real-time scheduling for multiple earth observation satellites," *Journal of Computer and System Sciences*, vol. 81, no. 1, pp. 110–124, 2015.
- [38] M. Wang, G. Dai, and M. Vasile, "Heuristic scheduling algorithm oriented dynamic tasks for imaging satellites," *Mathematical Problems in Engineering*, vol. 2014, Article ID 234928, 11 pages, 2014.
- [39] J. Wang, X. Zhu, D. Qiu, and L. T. Yang, "Dynamic scheduling for emergency tasks on distributed imaging satellites with task merging," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 9, pp. 2275–2285, 2014.
- [40] B.-C. Bai, Y.-Z. Ci, and Y.-W. Chen, "Dynamic task merging in multi-satellites observing scheduling," *Journal of System Simulation*, vol. 21, no. 9, pp. 2646–2649, 2009.
- [41] J. Branke, "Creating robust solutions by means of evolutionary algorithms," in *Parallel Problem Solving from Nature—PPSN V*, vol. 1498 of *Lecture Notes in Computer Science*, pp. 119–128, Springer, Berlin, Germany, 1998.
- [42] S. Tsutsui and A. Ghosh, "Genetic algorithms with a robust solution searching scheme," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 3, pp. 201–208, 1997.
- [43] F. Xhafa, J. Sun, A. Barolli, A. Biberaj, and L. Barolli, "Genetic algorithms for satellite scheduling problems," *Mobile Informa*tion Systems, vol. 8, no. 4, pp. 351–377, 2012.
- [44] X. Niu, H. Tang, L. Wu, R. Deng, and X. Zhai, "Imaging-duration embedded dynamic scheduling of Earth observation satellites for emergent events," *Mathematical Problems in Engineering*, vol. 2015, Article ID 731734, 31 pages, 2015.
- [45] T. Mao, Z. Xu, R. Hou, and M. Peng, "Efficient satellite scheduling based on improved vector evaluated genetic algorithm," *Journal of Networks*, vol. 7, no. 3, pp. 517–523, 2012.
- [46] B. Sun, W. Wang, X. Xie, and Q. Qin, "Satellite mission scheduling based on genetic algorithm," *Kybernetes*, vol. 39, 8, pp. 1255–1261, 2010.
- [47] M. T. Jensen, "Generating robust and flexible job shop schedules using genetic algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 3, pp. 275–288, 2003.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 828374, 14 pages http://dx.doi.org/10.1155/2015/828374

Research Article

A CDT-Based Heuristic Zone Design Approach for Economic Census Investigators

Changixu Cheng,^{1,2} Xiaomei Song,^{2,3} Jing Yang,¹ Xiatian Hu,² Shi Shen,¹ and Lijun Wang⁴

Correspondence should be addressed to Changixu Cheng; chengcx@bnu.edu.cn

Received 30 July 2015; Accepted 19 October 2015

Academic Editor: Xiaobo Qu

Copyright © 2015 Changixu Cheng 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

This paper addresses a special zone design problem for economic census investigators that is motivated by a real-world application. This paper presented a heuristic multikernel growth approach via Constrained Delaunay Triangulation (CDT). This approach not only solved the barriers problem but also dealt with the polygon data in zoning procedure. In addition, it uses a new heuristic method to speed up the zoning process greatly on the premise of the required quality of zoning. At last, two special instances for economic census were performed, highlighting the performance of this approach.

1. Introduction

Zone design is widely used as a method of spatial partitioning in geospatial sciences. It is similar to districting problem, which is a classical topic in optimization research. Both of the zone design and the districting problem use a certain of optimize operations divide a geographical region into some zones (or districts). The difference between them is to meet different criteria or constraints, such as balance, compactness, and contiguity. However, zone design pays more attention to how to build a spatial auxiliary graph. Sometimes, the zone design, also called district design or territory design, is usually considered a strategic activity [1]. The zone design has a broad range of applications, such as in political redistricting [2-7], sales territory alignment [8-10], school redistricting [11], logistics districting [12–15], and delimitating zones for land-use allocation and/or land acquisition and apportionment [16-18].

Traditional methods for zone design mainly focus on the complex geometries and spatial relationships between them. However, there are still some problems to be solved, such as nodes representation problem and geometrical barriers problem. Solving these problems from a GIScience perspective

may lead to a tractable solution, highlighting the importance of geographical insights [19, 20]. The use of ordinary and nonordinary Voronoi diagrams to solve districting problems has been reported in many literatures [21-23]. Novaes et al. [22] took advantage of the Voronoi diagram to detect the spatial relationship and solved geometrical barriers problem successfully. However, all above approaches are based on point data but are not suitable for polygon data. Murray and Tong [19] have raised the issue that only point-based representations are too simplistic to represent more complex vector objects like polygons in GIS. On the other hand, traditional methods make the shape of zones tend to be a circle or a square [21, 24]. That no longer applies with the presence of geometrical barriers problem. Another remarkable feature of traditional districting methods is that running procedures for optimizing the districting results usually cost too much time, which is in seconds or minutes [25–27].

So it is hard to use the previous studies or existing optimization software platform (such as CPLEX or MOSEK) to find an effective solution. To solve problems in this paper, we have to find a way different from the traditional ones. This paper presented a heuristic multikernel growth approach via Constrained Delaunay Triangulation (CDT) and introduced

¹ADREM, Beijing Normal University, Beijing 100875, China

²LREIS, Institute of Geographic Science and Resources Research, Beijing 100101, China

³School of Software, Tsinghua University, Beijing 100084, China

⁴China Internet Network Information Center, Beijing 100101, China

how to divide work zones for economic census investigators. Section 2 reviewed some relative work. Section 3 detailed how to construct an auxiliary graph for polygon data and barrier data and how to heuristically grow and generate spatial partitions. Section 4 gave some test to show functionality and performance of this method. Section 5 made some conclusions and some further work.

2. Problems Definition

This paper discussed a special zoning problem for economic census application. The objective of the problem is to partition a region into some census zones for investigators. Every investigator takes charge of surveying the financial situation of all companies in one zone. The partitioned zones should be contiguous, compact, and balanced. The main constraints presented are as follows: (1) contiguity makes sure that units in every zone are geographically adjacent; (2) compactness requires total travelling time in one zone as small as possible; (3) balance ensures that the workloads of every investigator are as equal as possible. The workload mainly includes investigating time and travelling time. In our application, data sources are polygon data and line obstacle data. Polygon data stands for the shape of buildings, where companies located. Some polygons are touching, but some of them are not touching, while line obstacle data stands for fences or busy roads which cannot be passed by. Additionally, it is difficult for ordinary users to give some upper and lower limits on a scalar attribute (e.g., investigating time and travelling time), because they hope to use it as simple as possible. To solve problems in this paper, we have to find a way different from the traditional ones.

First problem is about how to generate auxiliary graph for these touching or nontouching building polygons. Most of literatures often use a point to represent a polygon and then generate auxiliary graph based on these simplified points. However, in most of cases, the auxiliary graph mentioned above may not truly represent buildings' connection and touching relationships, because sometimes a polygon with complex shape may distort the connection and touching relationships of neighbor polygons. The partitions based on simple point-representation may result in some errors. Although Chou and Li [28, 29] and Ríos-Mercado and Fernández [27] discussed how to generate auxiliary graph directly based on polygons, they only considered touching polygons rather than nontouching polygons. Therefore, it is necessary to generate an auxiliary graph based on touching or nontouching polygons. In this paper, we used Constrained Delaunay Triangulation (CDT) to generate an Auxiliary Graph for nontouching Polygons (AGP). The shortest distance between discrete polygons in the auxiliary graph is related to the shapes of touching polygons, which would be described in Section 3.1.

The second problem is how to adjust AGP when line obstacles are involved. Novaes et al. [22] discussed some geometrical barriers problem, but they still focused on point data based on Voronoi diagram. Their method is not suitable for our polygon data based on CDT. In our application, geographical barriers may be polylines or polygons; they will affect the connection of building polygons. When a barrier

crosses some edges of an auxiliary graph, these edges (paths) will be cut. The adjusted graph is called *auxiliary graph for polygons and barriers* (AGPB).

The third problem is how to speed up the zoning process. In fact, spatial partition is a classical NP-hard problem. Most of literatures adopt the multikernel growth approach. The approach firstly selects a certain number of basic nodes as "seeds" (centers) for some zones and then successively adds every other node to its neighboring center until all have been assigned [23]. In the optimization phase, they need to constantly swap two adjacent nodes of zones in order to achieve optimized zones. When it comes to too many nodes, it is exhausted. It is feasible to consider at heuristics [6, 8, 30, 31], such as Tabu Search [32, 33] and simulated annealing [17, 34, 35]. The problem of traditional solutions is that bad seeds selection would generate bad partitions initiation. That means they often took more time to get optimization partitions. The procedure usually spends a lot of time after $20,000 \sim 80,000$ iterations [7]. In the paper, we use some heuristic strategies about spatial structure information to participate in the location of nodes and then avoided some unnecessary iterations.

3. A Heuristic Zone Design Approach Based on Constrained Delaunay Triangulation

There are two major phases in our method. Phase 1 is to construct an auxiliary graph for polygons and barriers and at last write them into XML files. Phase 2 is to use a spatial heuristic strategy to realize multikernel growth. The heuristic strategy continues to run as a simple optimization after all nodes are allocated. A flow chart of the general procedure is presented in Figure 1. Details of the steps are provided below.

3.1. Phase I: Auxiliary Graph Construction. The left part of Figure 1 shows the flow chart of Phase I. The AGPB construction is an optional operation. Users can input polygon data and barriers data based on actual situations. Finally, an auxiliary graph, which reflects the connectivity between nodes, is written into XML files.

3.1.1. Construction of AGP. When polygon data is input, topology checks and polygon-touching checks should be made first of all. This can avoid polygons being overlapped and can record nodes relation of touching. Then the AGP construction should be made and it can be depicted as follows: Firstly, DT is generated based on the vertexes of polygons and then is adjusted as Constrained Delaunay Triangulation (CDT) by counter lines of polygons. Secondly, all the edges contained in the polygons are deleted. Thirdly, shortest paths between the polygons are generated based on these triangulations. The construction of AGP is generated on CDT because a graph (DT) needs to be constructed by using the vertexes of the polygon; in most cases, the counter line of a polygon will intersect with the edges of DT. The special CDT is generated as follows: counter lines of polygons are seen as barrier lines and also as edges of triangulations that are forcibly inserted. Then, triangulations in polygons

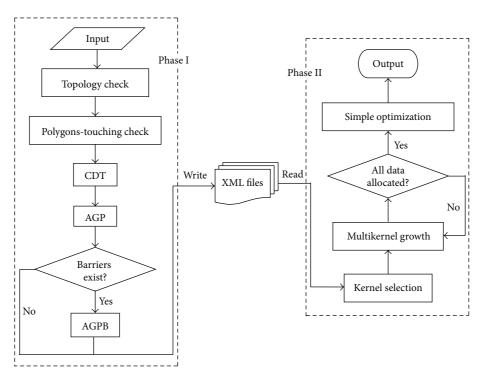


FIGURE 1: Diagram of general procedure flow.

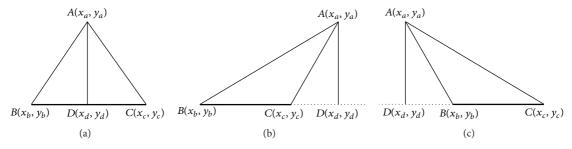


FIGURE 2: Search the short path.

need to be deleted. After the deletion of all the triangulations contained in the polygons, the polygons are attached in the graph as nodes. The relationship between adjacent nodes can be obtained by recursions.

In the third step above, the shortest paths between polygons are obtained by taking advantage of Delaunay Triangles. In fact, after the second step of CDT construction, triangles must connect two or three polygons. If three points of a triangle belong to three different polygons, its three edges are all likely to be the shortest paths between two polygons. Therefore, all edges in this type of triangles should be stored as candidates for the shortest path. If points of a triangulation belong to two polygons, it must be such a situation that a point belongs to a polygon and the other two points belong to another polygon. In those cases, the shortest path may not be the one with the edges existing in the triangle but rather the vertical distance from a point to a line. In Figure 2, we assume that there is a Delaunay Triangle connected with two polygons, and vertex $A(x_a, y_a)$ belongs to one triangle

and vertexes $B(x_b, y_b)$ and $C(x_c, y_c)$ belong to the other one. Assume that the edge AD stands for the shortest edge which is from vertex A to the straight line determined by vertexes B and C. The location of vertex D has two situations: either location inside the edge BC or location outside the edge BC. We assume that the slope of straight line BC is k. If point D moves from vertex B along x-axis by dx units, it will move along y-axis by $dy = x_b + k * dx$ units. Straight line AD is perpendicular to straight line BC and we obtain the formula as follows:

$$\frac{y_a - y_d}{x_a - x_d} = \frac{y_a - (y_b + k * dx)}{x_a - (x_b + dx)} = -\frac{1}{k}.$$
 (1)

Equation (1) can be further rearranged as follows:

$$dx = \frac{k * (y_a - y_b) + x_a - x_b}{1 + k^2}.$$
 (2)

If $0 \le dx \le x_c - x_b$, vertex *D* is located between vertex *B* and vertex *C* and the shortest path in the triangulation is the segment *AD*.

If $dx > x_c - x_b$, vertex *D* is located in the extension line of segment *BC* and the shortest path in the triangulation is the segment *AC*.

If dx < 0, vertex D is located in the extension line of segment CB and the shortest path in the triangulation is the segment AB.

3.1.2. Construction of AGPB. Further, the existing graph should be adjusted when geographical barriers exist. After the construction of AGP, we continue to put the vertexes of geometrical barriers into the graph and construct AGPB. In the process above we should find out all the edges intersecting with counter lines of geometrical barriers. Assume that the edge with the start point A and end point B needs to be adjusted. According to the graph, the geometrical barriers associated with points A and B can be found out and the set of their vertexes is assumed to be noted as $A_{\rm aset}$ and $B_{\rm aset}$, respectively.

Assume that the line with start point A and end point B is defined as Line $\langle A, X, B \rangle$ and X stands for a point or a line or the collection of both. For example, the Line $\langle C, D, E \rangle$ means points C, D, and E form a line; the line of Line $\langle A, \text{Line}\langle C, D, E \rangle$, $B \rangle$ is composed of points A, C, D, E, and B. Suppose O (defined as in (3)) is the set of barriers and is defined as follows where the integer c means the count of geometrical barriers. Consider

$$\mathbf{O} = \{O_1, O_2, \dots, O_c\} \quad (1 < c < \infty). \tag{3}$$

 O_i $(i=1,\ldots,c)$ represents a polygon or a line. Assume that the line with the minimum length in a set L of lines is noted as $\operatorname{Line}_{\min \operatorname{length}}\{L\}$. The shortest path function (noted as $\operatorname{Iter}_{\operatorname{Line}}(\cdot)$) of points A and B can be calculated as follows:

Iter_Line
$$(A, X, B) = \text{Line}_{\text{minlength}} \{ \text{Line}_{-\alpha} \langle A, X, B \rangle ,$$

$$\text{Line}_{-\beta} \langle A, X, B \rangle , \text{Line}_{-\gamma} \langle A, X, B \rangle , \qquad (4)$$

$$\text{Line}_{-\delta} \langle A, X, B \rangle \} .$$

As can be seen from the formula, the shortest path is calculated by considering four cases. Calculations of four cases are as follows:

(1) The shortest path is composed of three points, which are point A, point X from A_{aset} , and point B. Mathematical expression is as follows:

Line_
$$\alpha \langle A, X, B \rangle = \text{Line}_{\text{minlength}} \left\{ \text{Line} \langle A, X, B \rangle \right.$$

$$\in R^2 \mid \text{Line} \langle A, X, B \rangle \cap \left(\bigcup_{j=1}^n O_j \right) = \emptyset, \ X \in A_{\text{aset}} \right\}. \tag{5}$$

And we set a set α as follows:

$$\alpha = \left\{ X \in A_{\text{aset}} \mid \text{Line } \langle A, X, B \rangle \cap \left(\bigcup_{j=1}^{n} O_{j} \right) = \emptyset \right\}. \quad (6)$$

(2) The shortest path is composed of three points, which are point A, point X from B_{aset} , and point B. Mathematical expression is as follows:

Line_
$$\beta \langle A, X, B \rangle = \text{Line}_{\text{minlength}} \left\{ \text{Line } \langle A, X, B \rangle \right.$$

$$\in R^2 \mid \text{Line } \langle A, X, B \rangle \cap \left(\bigcup_{j=1}^n O_j \right) = \emptyset, \ X \in B_{\text{aset}} \right\}. \tag{7}$$

And we set a set β as follows:

$$\beta = \left\{ X \in B_{\text{aset}} \mid \text{Line } \langle A, X, B \rangle \cap \left(\bigcup_{j=1}^{n} O_{j} \right) = \emptyset \right\}. \tag{8}$$

(3) The shortest path is composed of four points, which are point A, point K from $(A_{aset}-\alpha)$, point L from $(B_{aset}-\beta)$, and point B. Mathematical expression is as follows:

$$\begin{aligned} &\operatorname{Line}_{-\gamma}\langle A,X,B\rangle = \operatorname{Line}_{\mathrm{minlength}} \, \left\{ \operatorname{Line} \, \langle A,K,L,B\rangle \right. \\ & \in R^2 \mid \operatorname{Line} \, \langle A,X_1,X_2,B\rangle \cap \left(\bigcup_{j=1}^n O_j \right) = \emptyset, \, \, K \end{aligned} \quad (9) \\ & \in \left(A_{\mathrm{aset}}\text{-}\alpha \right), \, \, L \in \left(B_{\mathrm{aset}}\text{-}\beta \right) \right\}. \end{aligned}$$

And we set a set γ as follows:

(4) An iterative process will be taken into account in this case. One point in the $(A_{aset}$ - α - $\gamma)$ and one point in the $(B_{aset}$ - β - $\gamma)$ are put into the function (I) and a set of lines (noted as *sublines*) can be obtained. Finally select the shortest one from *sublines*, put points A to B into the first and last position of the shortest subline, and the shortest path would be found out. The line on

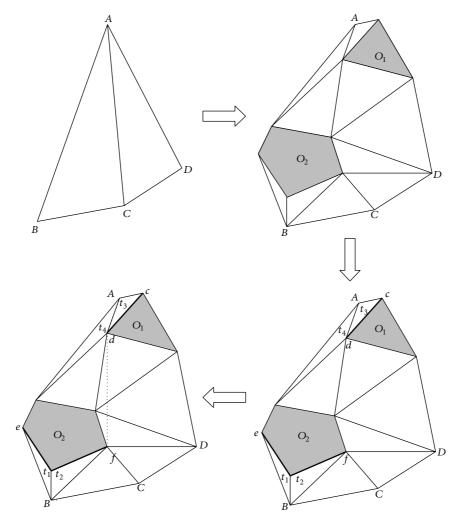


FIGURE 3: Search the short path around barriers.

behalf of the shortest path is composed of not less than four points. Mathematical expression is as follows:

Line_
$$\delta \langle A, X, B \rangle$$

$$= \operatorname{Line}_{\operatorname{minlength}} \left\{ \operatorname{Line} \left\langle A, \operatorname{Line} \left\langle M, Y, N \right\rangle, B \right\rangle$$

$$\in R^{2} \mid \operatorname{Line} \left\langle A, \operatorname{Iter_Line} \left(A, Y, B \right), B \right\rangle \cap \left(\bigcup_{j=1}^{n} O_{j} \right) \right. \tag{11}$$

$$= \emptyset, \ M \in \left(A_{\operatorname{aset}} - \alpha - \gamma \right), \ N \in \left(B_{\operatorname{aset}} - \beta - \gamma \right) \right\}.$$

The Antonio example is used here to illuminate the new method. Firstly the Delaunay Triangulation is created based on points A, B, C, and D. Add vertexes of barriers O_1 and O_2 to adjust the DT. The DT after adjustment is shown in Figure 3. According to DT, triangles t_1 , t_2 , t_3 , and t_4 are found out. So $A_{\rm aset} = \{c,d\}$ and $B_{\rm aset} = \{e,f\}$, where c, d,

e, and *f* represent vertexes of barriers shown in the figure. Line_ $\alpha\langle A, X, B \rangle$, Line_ $\beta\langle A, X, B \rangle$, and Line_ $\delta\langle A, X, B \rangle$ do not exist, while Line_ $\gamma\langle A, X, B \rangle = \text{Line}\langle A, d, f, B \rangle$. So the shortest path from point *A* to point *B* is $A \rightarrow d \rightarrow f \rightarrow B$.

Phase I mainly accomplishes operations for determining the spatial relations of data. It is relatively independent of Phase II. The graph obtained from Phase I can be stored and used at any time in Phase II with different parameters. Therefore, the design of two phases can save significant amount of time during graph construction.

3.2. Phase II: The Heuristic Multikernel Growth

3.2.1. Problem Formulation. Before describing the heuristic multikernel growth, it is necessary to formally specify the problem formulation of interest. The problem formulation can be specified from the follow three criteria: balance of workloads, compactness of zones, and contiguity of nodes.

The following parameters are defined:

N = node set.

E = edge set.

i = index of nodes.

k = index of nodes.

j = index of potential zones.

m = number of zones to plan.

 n_i = node i.

 w_i = workload of the node i.

 N_i = node set of the zone j.

 d_{ik} = Euclidean distance between nodes i and k.

 e_{ik} = edge connecting nodes i and k.

 MST_j = minimum spanning tree of the potential zone j.

 α = weight of compactness.

The following decision and auxiliary variables are defined:

$$a_{i \to j} = \begin{cases} 1, & \text{if node } i \text{ blongs to the district } j \\ 0, & \text{otherwise,} \end{cases}$$

$$b_{ik \to j} = \begin{cases} 1, & \text{if } e_{ik} \text{ blongs to the edge set of MST}_j \\ 0, & \text{otherwise,} \end{cases}$$
 (12)

 $e_{i \leftrightarrow k}$

$$= \begin{cases} 1, & \text{if node } i \text{ and } k \text{ are connected by an existing edge } e_{ik} \\ 0, & \text{otherwise.} \end{cases}$$

The mathematical formulation is as follows:

Balance

Minimize
$$\left(\max\left\{W_1,\ldots,W_j,\ldots,W_m\right\}\right)$$

- $\min\left\{W_1,\ldots,W_j,\ldots,W_m\right\}$, (13)

where

$$W_j = \sum_{n_i \in N_i} \left(w_j * a_{i \to j} \right) + \alpha * D_j \quad 1 \le j \le m.$$
 (14)

Compactness

$$Minimize (max \{D_1, \dots, D_j, \dots, D_m\}), \qquad (15)$$

where

$$D_{j} = \sum_{n_{i} \in N_{j}} \sum_{n_{k} \in N_{j}} \left(d_{ik} * a_{i \to j} * a_{k \to j} * b_{ik \to j} \right)$$

$$1 \le j \le m.$$
(16)

Contiguity. If a zone is an undirected graph whose nodes stand for buildings of the zone, an edge e_{ik} exists between nodes i and k if and only if $e_{i \leftrightarrow k} = 1$. A zone is contiguous if and only if the graph is connected (a path exists between every pair of adjacent nodes). A zone is feasible only if it is

contiguous. Constraint (17) means a node belongs to only one zone while constraint (18) means a zone contains more than one node:

$$\sum_{j} a_{i \to j} = 1, \quad n_i \in N \tag{17}$$

$$\sum_{n_i \in N_j} \sum_{n_k \in N_j} e_{i \leftrightarrow k} \ge 1. \tag{18}$$

The formulations above are drove by a specific zone problem in the real-world application. The problem is different from the traditional ones because we cannot give out the exact bounds for varieties or constraints. So we cannot completely use the methods of previous work or the existing optimization software platform (such as CPLEX or MOSEK) to solve the problem. In this work, multikernel growth is employed as a simple heuristic partition method. This method begins by selecting a certain number of basic nodes as "kernels" for the zones. The algorithm then successively adds, to each kernel, neighboring basic nodes by primarily considering the workload balance and compactness objectives.

3.2.2. A Heuristic Strategy. The specific multiobjective problem is solved in this work by the heuristic multikernel growth method (AMKG for short) based on an auxiliary graph. The contiguity objective is assumed to be achieved as long as nodes are connected by edges in an auxiliary graph. The balance and compactness objectives are achieved by a heuristic strategy: the zone with the smallest workload is selected and allocated with a specific basic node which contains a relative large workload and is relatively close to the zones. Furthermore, the specific basic node is found out firstly from unallocated nodes around the zone otherwise, if it failed to be got, from allocated nodes around the zone. The specific basic node is selected by calculating the "cost" which is defined by (19). To achieve the workload balance and compactness objective, "kernels" are greedily allocated basic nodes with the largest "cost" via an iterative procedure. The "cost" of node *i* can be defined as follows:

$$cost(i, j) = w_i + \beta * \frac{w_i}{D_{ij}} * \frac{\overline{D}}{\overline{W}},$$
 (19)

where β stands for a positive number of type double, \overline{D} means the average distance value of the edges in graph network, \overline{W} means the average value of workload, and $D_{ij} = \min_{n_i \in N_i} (d_{ik} * a_{i \mapsto j} * a_{k \mapsto j})$,

$$a_{i \mapsto j} = \begin{cases} 0, & \text{if node } i \text{ blongs to district } j \\ 1, & \text{otherwise.} \end{cases}$$
 (20)

We can see from (19) that the compactness objective can be generally achieved because D_{ij} would be as small as possible in the basic nodes allocation when the largest "cost" is selected.

It shoud be noted that in the heuristic multikernel growth the allocating basic node may be an unallocated one but may be an already allocated one. The proposed algorithm is with a certain degree of self-regulation and not sensitive to the initial

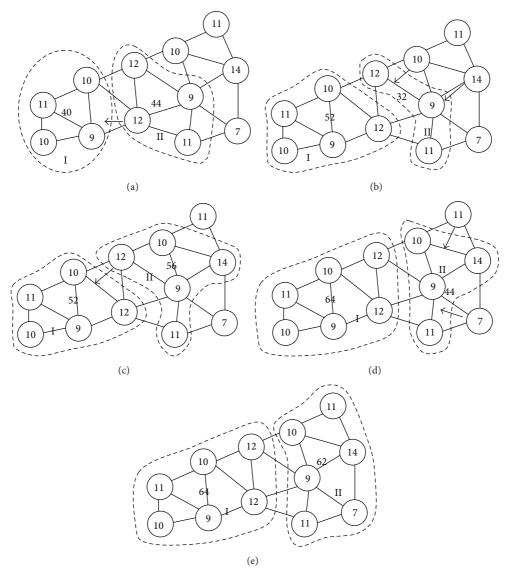


FIGURE 4: Kernels are growing.

choice of kernels. If allocating nodes must be unallocated nodes and the initial kernels are baddly selected, some zones would have no new unallocated node allocated in the process of zones growth. The final result may make a serious deviation from the balance object. For example, Figure 4(a) shows that zone I with the minimun workload (which is 40) needs to continue to be allocated nodes, but no unallocated nodes can be allocated. The proposed heuristic multikernel growth makes the zones growth "greedy" and "competitive." The zones with minimum workload may "eat" the adjacent node which belonged to other zones. In Figure 4(a), zone I gets the node with workload of 12 from zone II. Zone II needs to be allocated nodes because its workload becomes minimum. In accordance with the heuristic described previously the nodes with workload of 14 and 10 were allocated to zone II. According to the heuristic rules, the procedure would be continued as in Figures 4(c) and 4(d) and the final results are shown in Figure 4(e).

3.2.3. A Simple Optimization. The heuristic multikernel growth also has a simple and quick procedure of optimization for the result of nodes allocation after all nodes are allocated. The most classic optimization procedure for zoning problem may be the local search. It usually needs a huge time to run and it is hard to know when to stop the optimization procedure. The optimization procedure of the heuristic multikernel growth follows the same heurisic strategy as the nodes allocation. Its objective is to make the maximun workload of zones as small as possible (see (21)). So it is simple to be implemented and a little time comsuming:

Minimize
$$\left(\max\left\{W_1,\ldots,W_j,\ldots,W_m\right\}\right)$$
. (21)

3.2.4. The Implementation. The improved multikernel growth method includes three steps: (1) seeds selection; (2) nodes allocation; (3) simple optimization. There exist several approaches for determining a new configuration of

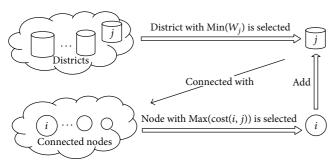


FIGURE 5: Allocating the basic nodes.

zone centers. Kalcsics et al. [26] thought a commonly used method was to solve in each territory resulting from the last allocation phase a 1-median problem. But it would be very complex, because much iteration should be taken to avoid bad cases. In our method, the initial seeds should be nodes which contain the largest number of companies. Of course it may not be optimization to take the initial seeds as kernels. If it is not optimization, this method will use a heuristic algorithm to adjust kernels. Nodes allocation takes into account three criteria: contiguity, balance, and compactness. The contiguity objective is assumed to be achieved as long as nodes are connected by edges in the graph. The balance and compactness objectives are achieved step by step in the allocation procedure. The general allocation procedure of improved multikernel growth method is illustrated in Figure 5. Zone j with the minimum workload would be selected to "grow." The connected basic node i with the maximum cost(i, j) will be selected and allocated to zone j. The simple optimization is similar to nodes allocation. Both of them follow the same heuristic strategy, but the main iteration of simple optimization is that if the first node with maximum "cost" in the nodes sorted list enlarges the minimum workload of zones or reduces the maximum workload of zones, it should be switched; otherwise the next node with the second maximum "cost" would be selected and repeat that logic. The iterative procedure would be continued until the maximum workload of zones could not be smaller any more by the heuristic strategy.

Assuming that the number of graphs which kernels generate is noted as g, the workload of the i-sub graph $(1 \le j \le g)$ is noted as W_j . The main procedure of improved multikernel growth is shown as follows:

(a) Take one graph, the workload of which is noted as W_{sg} , and identify the number of kernels which is c where

$$c = m * \frac{W_{sg}}{\sum_{i=1}^{g} W_i}.$$
 (22)

- (b) Obtain *c* nodes with the largest number of companies as centers for zones, and take the number of companies as the initial workload of zones.
- (c) Find out the zone (noted as *K*) with the minimum workload and find out its adjacent nodes via auxiliary subgraph firstly from unallocated nodes. If there are no unallocated nodes, the algorithm will find out its

- adjacent nodes from allocated ones; simultaneously calculate their cost by (19).
- (d) Allocate the node with the maximum cost to the zone K and recalculate its workload W_i .
- (e) If all the nodes in the subgraph are allocated, turn to step (c); otherwise continue.
- (f) Take the zone with the minimum workload, and try to "eat" its adjacent node with the biggest "cost." If the maximum workload is not smaller or the zone whose basic unit is "eaten" becomes not continuous, another adjacent node with the next biggest "cost" would be tried
- (g) Continue step (f) until the maximum workload could not be smaller any more.
- (h) Exit the procedure if the entire graph is handled; otherwise turn to step (a).

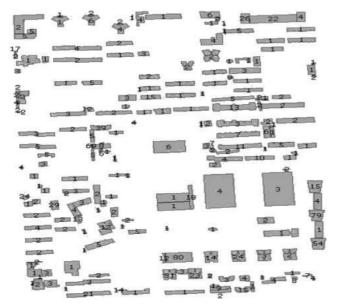
4. Computational Experiments and Results

To test the performance of the proposed solution procedure, a series of experiments were generated. All problem experiments were solved on a 2.53 GHz Pentium processor with 1.93 AGPB of RAM running with Windows XP as the operating system. Our procedure model is solved based on ArcEngine 9.3. A real-world application from an economic census whose operations consist of surveying the financial state of companies within a service region was performed in this work. Test data were all obtained from the real data of an economic census in Beijing, China. We solved two different sizes of instances, which are classified by number of nodes and zones:

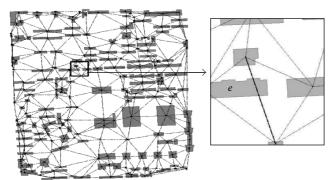
Test one: 213 nodes and 3 zones.
Test two: 741 nodes and 3\4\5 zones.

Each instance has two phases. Phase I checked the topology of polygons and got touching polygons recording zero distance. Then AGP would be generated. After AGP was created, we continued to turn the graph into AGPB wherever geometrical barriers existed. In Phase II, the heuristic multikernel growth approach was performed. The nodes with a relatively largest workload were selected as initial kernels. Nodes allocation and simple optimization would be applied following proposed heuristic strategy.

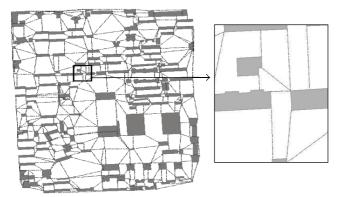
In test one, polygons were input and AGP was built in Phase I. In Phase II, several experiments are repeated on different values of parameters α and β . In Test two, geographical barriers are considered. Therefore, AGPB was built in Phase I. Several distances are solved based on different numbers of zones with the same values of α and β . Further, a comparison test between considering barriers and nonconsidering barriers is given in Section 4.2. This paper uses the Standard Deviation of Workload Values of zones to evaluate the balance of workload (STDEV_w) and uses the AVERAGE sum length of all MST edges in zones to evaluate the compactness of zones (AVERAGE_c). A smaller value of STDEV_w means a better balance result while a smaller value of AVERAGE_c means a better compactness result. Finally, time cost of instances was discussed.



(a) Polygon data and workload



(b) Construct the AGP based on points



(c) Construct the AGP based on polygons

Figure 6: Generation of the auxiliary graph.

4.1. Test One. The test data of test one is selected by a random rectangle from the economic census data. 213 buildings are selected into the test data and 1189 companies are involved. This indicates that there are 213 nodes whose total workload is 1189, and the number of companies located in every building is shown in Figure 6(a). Two spatial auxiliary graphs based on polygon center points and polygons were, respectively, built

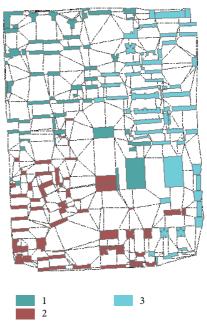


Figure 7: Result when $\alpha = 0$ and $\beta = 10$ according to polygons.

as shown in Figures 6(b) and 6(c). In Figure 6(b), edge e is a spatial relational edge crossing a building polygon. Edge e should not exist as a path in real world. What is more, length of edges in Figure 6(b) is generally longer than that in Figure 6(c) which could more accurately represent the true path length.

In generation of the graph, triangles between polygons were initially created and then AGP was created consecutively. Figure 6(b) shows details of AGPB. Finally, we decided to generate three zones from the test data.

In the first step of Phase II, we selected three nodes with the largest number of companies as kernels. In the second step, we set different values to α and different results are obtained. In Figure 7, we thought travel time on road can be negligible relative to the investigation time in companies, so we set $\alpha=0$ and $\beta=10$. The average value of workload is 396.33 (the number of companies actually) and that of three zones is 397, 396, and 396, respectively. This means that the result has got a good balance of workload for zoning. We also can find out that in Figure 7 continuity of zones has been held. Compactness of zones is used to control the total travel time in our application so it would be acceptable due to negligible travel time.

When $\alpha > 0$, the workload content included the travel cost and the compactness objective would take into account the distance of edges between polygons by setting $\beta = 10$. We set different values for α and the numerical results are presented in Figure 8. The parameter n stands for the number of companies, d for the average length of MST, and w for the workload in zones.

In Figure 8, we note that, in general, values of STDEV $_w$ and AVERAGE $_c$ almost have no large change with the change of the value of α . Two charts (see Figure 8) are generated from test results and support the assumption stated above. This indicates that the balance of workload and

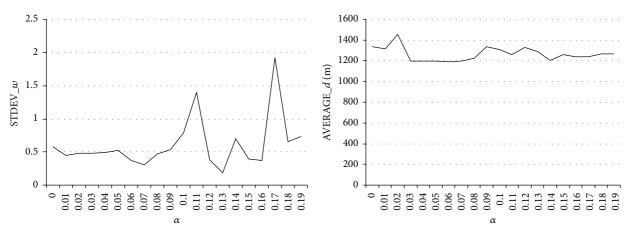


FIGURE 8: Standard deviation of workload and average length of MST.

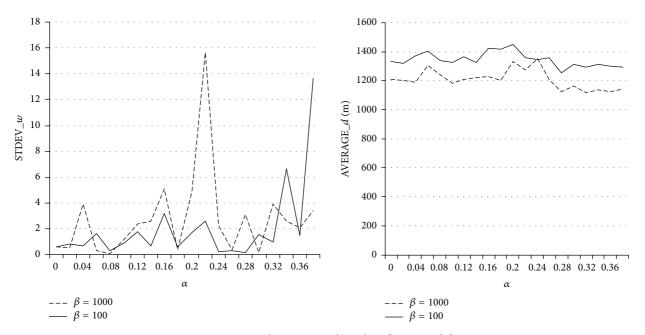


FIGURE 9: STDEV_w and AVERAGE_d/m when $\beta = 100$ and $\beta = 1000$.

the compactness objectives are affected not too much by the composition of workload content. In practical application, the appropriate value of α needs to be considered when considering workload content and we can set its value as a ratio of investigation time to travel time. With changes in the value of α , standard deviation of workload is stable and is always approximately 0.6 in most cases. The average value of distances changes around 1250. This indicates that the proposed heuristic works and holds the balance of workload and compactness to a certain extent when β is specified and whatever value α is.

To test the effect of the parameter β in the heuristic multikernel growth, another twenty experiments were performed with twenty different values of α and two specific values of β . The test result of experiments is shown as in Figure 9. It is noted that when the value of β is relatively small, the balance of workload is in general well held but the compactness is badly achieved. This means that β is a control to coordinate compactness and balance objective. From Figure 9 we can find out that a specific β corresponds to a relatively specific value of AVERAGE_c. So in the real application we usually specify β by considering the travel time limits or means of transportation.

4.2. Test Two. Test data of large instances are selected by a random rectangle from the economic census data. There are 741 polygon nodes whose total workload is 7107 and the number of companies located in every building is shown in Figure 10(a). In addition, we considered some busy roads or fences through the test data as geographical barriers, which can be seen in Figure 10(b). Finally, we aimed at generating 3\4\5 zones from the test data.

To illustrate the impact of geometrical barriers to the zoning results, we have made comparison experiments without

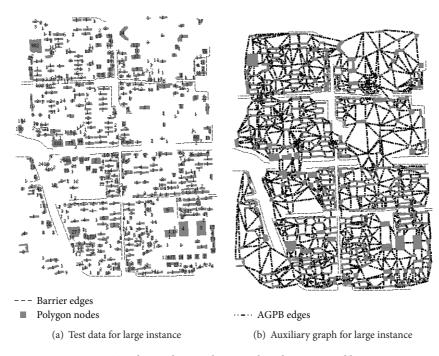


FIGURE 10: Test data and its auxiliary graph with geometrical barriers.

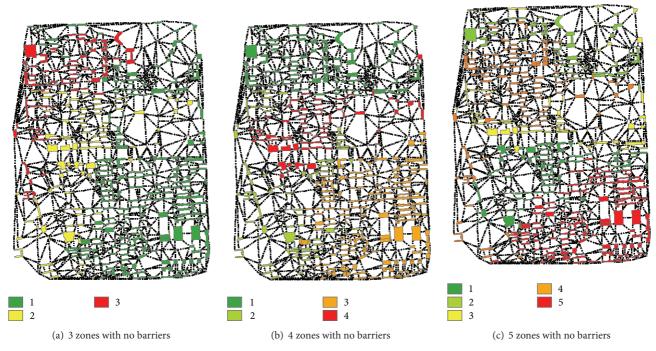


FIGURE 11: Results of 3\4\5 zones with no barriers.

barriers but with the same parameters α and β . In Figures 11 and 12, we could find out that geometrical barriers had affected the zoning results.

The comparison between effects of zoning results also has been made. Set $\alpha = 0.08$ and $\beta = 100$, and effects of balance

and compactness between the 3\4\5 zones with barriers or no barriers are shown in Table 1. We can see that geometrical barriers have affected the zoning results mainly affecting the distribution of members in zones but having relatively little effect to zoning results.

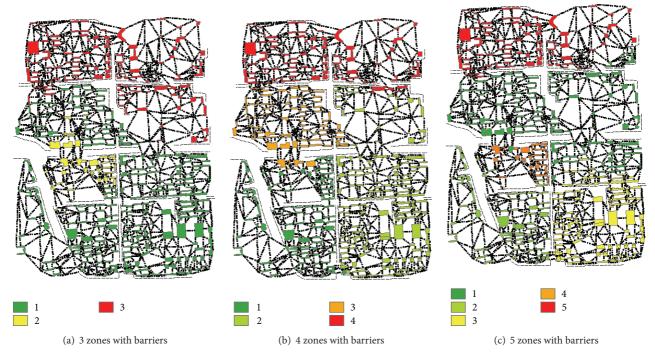


FIGURE 12: Results of 3\4\5 zones with barriers.

TABLE 1: Results of large instances.

α β	Rarriere	Three zones		Four zones		Five zones		
	Р	Barriers	STDEV_ w	AVERAGE_c	$STDEV_w$	AVERAGE_c	$STDEV_w$	AVERAGE_c
0.08	100	Do not exist	0.215094	12598.68	0.555236	9234.909	1.406661	7297.194
0.08	100	Exist	0.598867	11977.43	0.429749	9460.601	0.505428	7682.173

TABLE 2: Time cost in different instances.

Instances	Parameters	Barriers	Phase I	Phase II	Total time
	$\alpha, \beta = 10$	Do not exist	15.406 s	6.25 ms	15.412 s
213_3	$\alpha, \beta = 100$	Do not exist	15.406 s	6.25 ms	15.412 s
	$\alpha, \beta = 1000$	Do not exist	15.406 s	6.25 ms	15.412 s
741.2	$\alpha = 0.08 \beta = 100$	Exist	53.875 s	46.875 ms	53.922 s
741_3	$\alpha = 0.08, \ \beta = 100$	Do not exist	52.921 s	53.125 ms	52.974 s
741 4	$\alpha = 0.08 \beta = 100$	Exist	53.875 s	48.437 ms	53.923 s
741_4	$\alpha = 0.08, \ \beta = 100$	Do not exist	52.921 s	51.562 ms	52.973 s
741 5	$\alpha = 0.08 \beta = 100$	Exist	53.875 s	48.438 ms	53.923 s
741_5	$\alpha = 0.08, \ \beta = 100$	Do not exist	52.921 s	51.563 ms	52.973 s

Constrained Delaunay Triangulation can successfully deal with our zoning problem with barriers (see Figures 11 and 12) while holding an acceptable quality of zoning.

4.3. Time Cost. For different situations, we may construct different graphs by identifying the data types and existence of barriers in Phase I and adopt the heuristic multikernel growth in Phase II. In Sections 4.1 and 4.2, we observe that acceptable results can be obtained. What is more, time cost of the procedure is very low, which is one of notable features discussed in this paper. In Table 2, time cost for instances discussed above is presented.

Instances of 213_3 have been repeated for 20 times with different values of α and computational time of Phase II is the average value of results from 20 repeated tests. We can observe that the time cost in Phase II of the heuristic multikernel growth is so little that it can be almost negligible comparing with the time cost in the whole process of zoning. In this zoning procedure, running times in our results are in milliseconds while traditional methods usually consider a few seconds or minutes to optimize zoning results [25–27]. The construction of the graph in Phase I consumed too much computational time. Fortunately, however, the procedure of Phase I is separate and does not interact with that of Phase

II. Therefore, once the graph is built up and stored in a file, there is no necessity to rebuild the auxiliary graph and the file only needs to be read prior to the procedure of the proposed heuristic strategy with different values of parameters α and β . Therefore, the average computational time will be decreased greatly and can well meet our requirements of the special zoning problem in the economic census.

5. Conclusions and Possible Further Work

This paper addresses a special zoning problem for economic census that is motivated by a real-world application. Firstly, we introduced the Constrained Delaunay Triangulation to solve several problems such as polygon-based graph construction problem and geographical barriers problem which are generally countered in the zoning problem. The difficulty of this NP-complete problem motivated us to propose the heuristic multikernel growth following a heuristic strategy to speed up the zoning process greatly in the premise of the required quality of zoning. According to real-world instances, we present problem formulations to optimize three criteria: contiguity, compactness, and balance of workload among zones.

Two special instances for economic census were performed, highlighting the applicability of this approach. They resulted in acceptable compact zones with considerable balance of workload. Test one showed the partition results of touching and nontouching polygons. Some experiments showed how to determine the values of α and β and the meanings of them; that is, the values of α determine workload composition while β determines balance and compactness object. Test one also showed that compactness would be better achieved if β is larger while balance object would become worse. Test two showed that our method could successfully solve zoning problem with barriers and still could achieve some reasonable results. Test three in Section 4.3 showed that our method could get optimized results within acceptable time cost.

The simple spatial heuristic approach in this paper makes a good performance in most of cases, but sometimes it maybe results in bad compactness objective. The reason is that the compactness objective is achieved by minimizing the maximum length of MST constructed in the zones when the kernels are growing without considering the shape of zones. In the future, we will do our efforts on the problem and maybe take into account the shape of zones for compactness.

Conflict of Interests

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

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant nos. 41222009 and 41271405) and the foundation of BNU (Project no. 111007022).

References

- [1] L. Muyldermans, D. Cattrysse, and D. Van Oudheusden, "District design for arc-routing applications," *Journal of the Operational Research Society*, vol. 54, no. 11, pp. 1209–1221, 2003.
- [2] M. Altman, "The computational complexity of automated redistricting: is automation the answer?" *Rutgers Computer & Technology Law Journal*, vol. 23, no. 1, pp. 81–142, 1997.
- [3] P. K. Bergey, C. T. Ragsdale, and M. Hoskote, "A simulated annealing genetic algorithm for the electrical power districting problem," *Annals of Operations Research*, vol. 121, no. 1, pp. 33–55, 2003.
- [4] S. W. Hess, J. B. Weaver, H. J. Siegfeldt, J. N. Whelan, and P. A. Zitlau, "Nonpartisan political redistricting by computer," *Operations Research*, vol. 13, no. 6, pp. 998–1006, 1965.
- [5] M. E. T. Horn, "Solution techniques for large regional partitioning problems," *Geographical Analysis*, vol. 27, no. 3, pp. 230–248, 1995.
- [6] A. Mehrotra, E. L. Johnson, and G. L. Nemhauser, "An optimization based heuristic for political districting," *Management Science*, vol. 44, no. 8, pp. 1100–1114, 1998.
- [7] F. Ricca and B. Simeone, "Local search algorithms for political districting," *European Journal of Operational Research*, vol. 189, no. 3, pp. 1409–1426, 2008.
- [8] C. Easingwood, "A heuristic approach to selecting sales regions and territories," *Operational Research Quarterly*, vol. 24, no. 4, pp. 527–534, 1973.
- [9] L. M. Lodish, "Sales territory alignment to maximize profit," *Journal of Marketing Research*, vol. 12, no. 1, pp. 30–36, 1975.
- [10] A. A. Zoltners and P. Sinha, "Integer programming models for sales resource allocation," *Management Science*, vol. 26, no. 3, pp. 242–260, 1980.
- [11] J. A. Ferland and G. Guenette, "Decision support system for the school districting problem," *Operations Research*, vol. 38, no. 1, pp. 15–21, 1990.
- [12] R. G. González-Ramírez, N. R. Smith, R. G. Askin, and V. Kalashinkov, "A heuristic approach for a logistics districting problem," *International Journal of Innovative Computing, Information and Control*, vol. 6, no. 8, pp. 3551–3562, 2010.
- [13] C. F. Daganzo, "The distance traveled to visit N points with a maximum of C stops per vehicle: an analytic model and an application," *Transportation Science*, vol. 18, no. 4, pp. 331–350, 1984.
- [14] G. F. Newell and C. F. Daganzo, "Design of multiple-vehicle delivery tours—I a ring-radial network," *Transportation Research Part B*, vol. 20, no. 5, pp. 345–363, 1986.
- [15] A. Langevin and F. Soumis, "Design of multiple-vehicle delivery tours satisfying time constraints," *Transportation Research Part B*, vol. 23, no. 2, pp. 123–138, 1989.
- [16] T. J. Cova and R. L. Church, "Contiguity constraints for singleregion site search problems," *Geographical Analysis*, vol. 32, no. 4, pp. 306–329, 2000.
- [17] J. C. J. H. Aerts and G. B. M. Heuvelink, "Using simulated annealing for resource allocation," *International Journal of Geographical Information Science*, vol. 16, no. 6, pp. 571–587, 2002.
- [18] J. C. Williams, "A zero-one programming model for contiguous land acquisition," *Geographical Analysis*, vol. 34, no. 4, pp. 330–349, 2002.
- [19] A. T. Murray and D. Q. Tong, "Coverage optimization in continuous space facility siting," *International Journal of Geographical Information Science*, vol. 21, no. 7, pp. 757–776, 2007.

- [20] L. Yuan, Z. Yu, W. Luo, L. Yi, and Y. Hu, "Pattern forced geophysical vector field segmentation based on Clifford FFT," *Computers & Geosciences*, vol. 60, pp. 63–69, 2013.
- [21] P. Moreno-Regidor, J. García López de Lacalle, and M.-Á. Manso-Callejo, "Zone design of specific sizes using adaptive additively weighted Voronoi diagrams," *International Journal of Geographical Information Science*, vol. 26, no. 10, pp. 1811–1829, 2012.
- [22] A. G. N. Novaes, J. E. Souza de Cursi, A. C. L. da Silva, and J. C. Souza, "Solving continuous location-districting problems with Voronoi diagrams," *Computers and Operations Research*, vol. 36, no. 1, pp. 40–59, 2009.
- [23] L. Svec, S. Burden, and A. Dilley, "Applying Voronoi diagrams to the redistricting problem," *The UMAP Journal*, vol. 28, no. 3, pp. 313–329, 2007.
- [24] T. Shirabe, "A model of contiguity for spatial unit allocation," *Geographical Analysis*, vol. 37, no. 1, pp. 2–16, 2005.
- [25] P. K. Bergey, C. T. Ragsdale, and M. Hoskote, "A decision support system for the electrical power districting problem," *Deci*sion Support Systems, vol. 36, no. 1, pp. 1–17, 2003.
- [26] J. Kalcsics, S. Nickel, and M. Schröder, "Towards a unified territorial design approach—applications, algorithms and GIS integration," *Top*, vol. 13, no. 1, pp. 1–74, 2005.
- [27] R. Z. Ríos-Mercado and E. Fernández, "A reactive GRASP for a commercial territory design problem with multiple balancing requirements," *Computers and Operations Research*, vol. 36, no. 3, pp. 755–776, 2009.
- [28] C.-I. Chou and S. P. Li, "Taming the gerrymander—statistical physics approach to political districting problem," *Physica A: Statistical Mechanics and Its Applications*, vol. 369, no. 2, pp. 799–808, 2006.
- [29] C.-I. Chou and S.-P. Li, "Spin systems and political districting problem," *Journal of Magnetism and Magnetic Materials*, vol. 310, no. 2, pp. 2889–2891, 2007.
- [30] M. S. Heschel, "Effective sales territory development," *Journal of Marketing*, vol. 41, no. 2, pp. 39–43, 1977.
- [31] L. M. Lodish, "'Vaguely right' approach to sales force allocations," *Harvard Business Review*, vol. 51, pp. 119–124, 1974.
- [32] L. D. Bodin, "A districting experiment with a clustering algorithm," *Annals of the New York Academy of Sciences*, vol. 219, pp. 209–214, 1973.
- [33] M. Blais, S. D. Lapierre, and G. Laporte, "Solving a home-care districting problem in an urban setting," *Journal of the Operational Research Society*, vol. 54, no. 11, pp. 1141–1147, 2003.
- [34] W. Macmillan, "Redistricting in a GIS environment: an optimisation algorithm using switching-points," *Journal of Geographical Systems*, vol. 3, no. 2, pp. 167–180, 2001.
- [35] S. J. D'Amico, S.-J. Wang, R. Batta, and C. M. Rump, "A simulated annealing approach to police district design," *Computers and Operations Research*, vol. 29, no. 6, pp. 667–684, 2002.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 316092, 12 pages http://dx.doi.org/10.1155/2015/316092

Research Article

Epidemic Spreading Characteristics and Immunity Measures Based on Complex Network with Contact Strength and Community Structure

Xueting Zhang, Bingfeng Ge, Qi Wang, Jiang Jiang, Hanlin You, and Yingwu Chen

¹College of Information System and Management, National University of Defense Technology, Changsha, Hunan 410073, China ²College of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan 410073, China

Correspondence should be addressed to Xueting Zhang; zhangxueting0807@foxmail.com

Received 3 July 2015; Accepted 22 October 2015

Academic Editor: Xiaobo Qu

Copyright © 2015 Xueting Zhang 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.

Middle East Respiratory Syndrome (MERS), bursting in South Korea from May 2015 and mainly spreading within the hospitals at the beginning, has caused a large scale of public panic. Aiming at this kind of epidemic spreading swiftly by intimate contact within community structure, we first established a spreading model based on contact strength and SI model, and a weighted network with community structure based on BBV network model. Meanwhile, the sufficient conditions were deduced to ensure the optimal community division. Next, after the verification by the real data of MERS, it is found that the spreading rate is closely related to the average weight of network but not the number of communities. Then, as the further study shows, the final infection proportion declines with the decreases both in isolation delay and in average weight; however, this proportion can only be postponed rather than decreased with respect to sole average weight reduction without isolation. Finally, the opportunities to take action can be found to restrain the epidemic spreading to the most extent.

1. Introduction

Middle East Respiratory Syndrome (MERS), first identified in Saudi Arabia in 2012, is a viral respiratory disease caused by a novel coronavirus. According to the World Health Organization (WHO), on May 14, 2013, there were 38 MERS cases, which grew further to 1150 after two years (i.e., May 31, 2015). For human-to-human transmission, the virus does not appear to pass easily from person to person unless they have close contact, such as providing unprotected care to or living together with infected patients.

The epidemic status in South Korea from May 20, 2015, caused national public panic and worldwide attention. At the beginning, it spread evidently amongst the infectors and the patients in the same sickroom. This paper thus focuses on the spreading characteristics of intimate contact with community structure.

Nowadays, the studies of epidemics spreading are twofold: the spreading model of differential equation and the complex network theory. There are three spreading models widely used in the study of virus transmission, namely, SIR model, SIS model, and SI model [1–4], and the solving algorithms are mainly based on percolation theory [5, 6], mean field theory [7, 8], and Markov chain theory [9, 10]. For the complex network, Erdös [11] proposed the random network, while Watts and Strogatz [12] presented the small world network model with smaller average shortest path length and bigger clustering coefficient. Barabási and Albert [13] put forward the scale-free network model with both adding points and preferential attachment. According to the real network, the connections in many networks are not merely binary entities (i.e., either present or not) but have associated weights that record their strengths relative to one another. Thus, Barrat et al. [14] created the BBV model where point weight and edge weight evolve dramatically.

With the further study on the network topology, it is widely recognized that closely connected nodes and communities in social networks play an important role in topological properties and functional dynamics of involved complex networks [15, 16]. As a result, there are many community

division algorithms and accuracy indices such as modularity [17–20]. Based on the aforementioned models, Liu and Hu [21] researched on the epidemic spreading through the network with small world effect by SIS model, while Smieszek et al. [22] studied that by SIR model. Salathé and Jones [23] focused on the influence of community structure to virus transmission. In order to know about the virus transmission mechanism better, the dynamic models endeavor to slow down the outbreak rate and control the spreading range. Many different immunization strategies are proposed, such as random immunization [1], target immunization [8], and acquaintance immunization [24].

SI model is often applied to study on the epidemic dynamics at the early outbreak stages [25]. At the beginning, MERS in South Korea were mainly concentrated in three hospitals with obvious characteristics of community structure. In reality, the infection probability increases with the raise of contact time between infectious people and susceptible people, which must be highlighted in the epidemic spreading models. Edge weight is essential to describe the contact intimacy. In order to study the epidemic spreading characteristics and the optimal opportunity to take measures of MERS in South Korea, it is supposed that there is a linear relationship between contact strength and contact time. In this paper, spreading model is based on SI model with contact strength, and weighted network with community structure is based on BBV network model. The spreading characteristics are obtained by simulation according to the aforementioned models. Hence, they are verified by the real data in the South Korea MERS epidemic.

The remainder of this paper is organized as follows. Section 2 demonstrates the spreading model based on SI model in view of contact strength. Section 3 establishes the weighted network with community structure based on BBV model and analyzes the characteristics. Then, the spreading characteristics are studied to find the effective factors in Section 4, where the epidemic spreading process is divided into five stages. Section 5 studies the controlling measures (such as how to execute isolation and reduce the average weight of network) and the optimal opportunity to carry them out. Finally, case study based on MERS in South Korea is demonstrated in Section 6 to verify the models and the measure effects. Overall, both theoretical analysis and simulation results focus on the spreading characteristics and measure effects.

2. Epidemic Spreading Model

In this section, we suppose that the longer time the susceptible person contacts with the infectious ones, the larger probability the susceptible person will get infectious, at the beginning stage of the epidemic outbreak. In order to study the epidemic spreading process, we propose the spreading model based on contact strength and the computer simulation flowchart.

2.1. SI Model Based on Contact Strength. During the study of epidemic diffusion theory, the models are always based

on some assumptions that the infectious unit is the node in the network and the epidemic can only spread through the links. The individuals are divided into 3 types: S (Susceptible) means the healthy state which is likely to be infectious, I (Infected) indicates the illness state which has already been infectious, and R (Removed) signifies the immune state which has always been recovered or dead.

In terms of some epidemic bursting suddenly without valid control, such as SARS, H1N1 [26], and especially MERS, SI model is often used to study the spreading characteristics at the beginning period of diffusion process. Overall, prompt prevention measures would reduce the detrimental influence, which are of theoretical value and reality significance.

When the contact strength to an infectious person is T_1 , the noninfectious probability of susceptible person is defined as

$$(1-\lambda)^{T_1/T_0}. (1)$$

Thus, when the contact strength to an infectious person is T_0 , the infectious probability of susceptible person is defined as

$$1 - (1 - \lambda)^{T_0/T_0} = \lambda.$$
(2)

One susceptible person contacts an infectious person with T_1 strength, then leaves for a while, and then contact the same infectious person with T_3 strength. If he does not get infection at the first time, he would seem as a healthy susceptible person at the second moment. That is, the two contacts are independent, and the noninfectious probability of susceptible person after second contact is

$$(1-\lambda)^{T_2/T_0} \cdot (1-\lambda)^{T_3/T_0} = (1-\lambda)^{(T_2+T_3)/T_0}.$$
 (3)

If one susceptible point contacts two infectious points with strengths T_4 and T_5 , respectively, the noninfectious probability of susceptible person after two contacts is

$$(1-\lambda)^{(T_4+T_5)/T_0}$$
. (4)

For susceptible point *i*, the infection probability λ_i is

$$\lambda_i = 1 - (1 - \lambda)^{\sum_j T_{ij}/T_0}, \tag{5}$$

where T_{ij} indicates the edge strength between point i and infectious point j and $\sum_j T_{ij}$ means the sum of edge strength between point i and its adjacent infectious points.

 $\langle \sum_j T_{ij} \rangle$ is defined as the average $\sum_j T_{ij}$ of each point in the whole network.

When the whole network is stable, the ratio of susceptible points is s(t), and the ratio of infectious points is i(t), then

$$s(t) + i(t) = 1,$$

$$\frac{\operatorname{d}i(t)}{\operatorname{d}t} = \left[1 - (1 - \lambda)^{\langle \sum_{j} T_{j} \rangle / T_{0}}\right] i(t) \left[1 - i(t)\right].$$
(6)

In the actual infection process, the longer time the susceptible person contacts the infectious person, the larger probability of susceptible person to get infection. In other

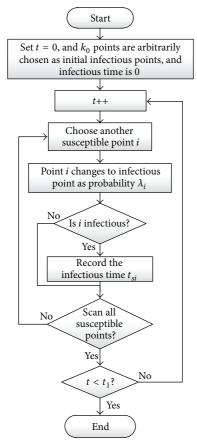


FIGURE 1: Flowchart of the spreading model simulation.

words, the edge strength between them is magnified. In conclusion, T_{ii} is a formula $T_{ii}(t)$ that depends on time t.

For simplicity, it is assumed that there is a linear relationship between $T_{ij}(t)$ and t:

$$T_{ii}\left(t\right) = \alpha_{ii} \cdot t,\tag{7}$$

where α_{ij} indicates the coefficient of contact strength, $\alpha_{ij} < 1$, $\langle \alpha \rangle$ means the average α_{ij} , and $\langle k \rangle$ signifies the average point degree of network. Thus,

$$\frac{\mathrm{d}i(t)}{\mathrm{d}t} = \left[1 - (1 - \lambda)^{(\langle k \rangle \cdot \langle \alpha \rangle t)/T_0}\right] i(t) \left[1 - i(t)\right], \tag{8}$$

where $a = (1 - \lambda)^{(\langle k \rangle \cdot \langle \alpha \rangle)/T_0}$.

The solution is

$$i(t) = \frac{1}{(i_0^{-1} - 1)e^{-t + \ln^{-1} a \cdot (a^t - 1)} + 1},$$
(9)

where i_0 is the ratio of infectious points at time 0, and when $t \to \infty$, i(t) = 1.

2.2. Simulation of Spreading Model. According to the above spreading model, the simulation flowchart is shown in Figure 1.

In terms of susceptible point i, the infectious probability λ_i is

$$\lambda_i = 1 - (1 - \lambda)^{(\sum_j \alpha_{ij} \cdot (t - t_{sj}))/T_0},$$
 (10)

where α_{ij} is the edge weight, t_{sj} is the infected moment for point j and T_0 and λ are the parameters related to the epidemic spreading characteristics.

During the simulation process, the time step length is always the same as T_0 .

Epidemic spreads in the human-to-human network. After the construction of the epidemic spreading model based on the contact among people, we need to construct the network model. Due to the characteristics of people social contact, the weighted network model with community structure is essential to be constructed.

3. Network Model

This section establishes the weighted network with community structure and proposes the simulation process as the flowchart and then analyzes the sufficient conditions for the optimal community division. After that, the weight distribution and the community division characteristics are obtained [27].

3.1. Generation of Weighted Network with Community Structure. BBV network model is a weighted scale-free network model provided by Barrat et al. [14], which allows the dynamical evolution of weights during the growth of the system.

It is assumed that point i is within the community X, A_i indicates the strength of point i, A_i^I means the inner strength of point i, and A_i^O signifies the outer strength of point i; thus,

$$A_{i}^{I} = \sum_{j \in X} \alpha_{ij},$$

$$A_{i}^{O} = \sum_{j \notin X} \alpha_{ij},$$

$$A_{i} = A_{i}^{I} + A_{i}^{O},$$
(11)

where α_{ij} means the edge weight between point i and point j; if i and j are disconnected, $\alpha_{ij} = 0$. If point i and point j are in the same community, it is called edge weight inner the community, while if they are in two different communities, it is called edge weight outer the community.

Therefore, the evolution process of advanced BBV network is as follows.

Step 1 (initial setup). There are S communities (i.e., X_1, X_2, \ldots, X_S) contributing to network G. The initial community has m_{r_0} nodes (all m_{r_0} are the same as m), connected by a small quantity of edges. And different communities are connected by several edges (without the inner links of points in the community), which constitutes the initial network. All initial edge weights are α_0 .

Step 2 (adding points). Add a new point i in each time step, and choose an arbitrary community (assumed as X_r) to join in. The new adding point is linked to the existing m_r ($m_r \le m_{r_0}$) points in the community as the probability $p_r^I(j)$ (as shown in formula (1)) and linked to the existing n ($n < \sum_s m_{s_0}$) points out of the community as the probability $p_r^O(k)$ (as shown in formula (2)):

$$p_r^{I}(j) = \frac{A_j^{I}}{\sum_{j \in X_r} A_j^{I}},$$

$$p_r^{O}(k) = \frac{A_k^{O}}{\sum_{k \in G} A_k^{O}}.$$
(12)

Step 3 (weight change). The new adding edge is endued with an initial weight α_0 , which leads to the weight change of the inner community of conjoint point j and the outer community of conjoint point k. The edge weights of the inner community are adapted to $\alpha_{jv} = \alpha_{jv} + \mu^I \alpha_{jv}/A_i^I$ (where i, j, and v are located at different communities), and the edge weights of the outer community are adapted to $\alpha_{kl} = \alpha_{kl} + \mu^O \alpha_{kl}/A_k^O$ (where i and k are located at different communities and k and k are located at different communities); μ^I and μ^O are the coefficients of weight enhancement.

Step 4 (weight normalization). After the construction of network, all the weights in the network are normalized. Assuming that the maximum edge weight of outer community is α_m^O , the maximum edge weight of inner community is α_m^I . Thus, the edge weights in the outer community of each point are normalized as α_i^O/α_m^O , and the edge weights in the inner community of each point are normalized as α_i^I/α_m^I .

Similar to the BBV network model, each community forms a BBV network; when *S* is big, all points in the whole network are connected to the edges out of the community to construct a BBV network. According to the mean filed theory [7, 8] and BBV network model [14], both the inner weight distribution and outer weight distribution of community are in accordance with the power law.

According to the above four steps, the network meets all restrictions that are likely to be created as the simulation process of the weighted network with community structure in Figure 2.

3.2. Sufficient Conditions for Community Division. After the network generation, further research needs to analyze whether this kind of community division is the best. In this section, we study the sufficient condition to reach the optimal community division by analyses of modularity.

Salathé and Jones [23] proposed the representation method of modularity Q in the weighted network:

$$Q = \sum_{r=1}^{S} \left[e_{rr} - (h_r)^2 \right],$$
 (13)

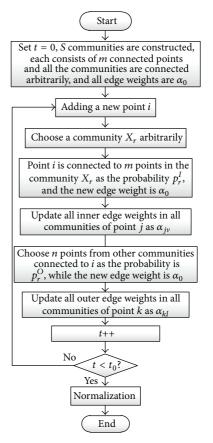


FIGURE 2: Flowchart of network model generation.

where

$$e_{rr} = \frac{1}{2A} \sum_{i,j \in X_r} \alpha_{ij},$$

$$h_r = \frac{1}{2A} \sum_{i \in X_r} A_i,$$

$$A = \frac{1}{2} \sum_{i,i \in G} \alpha_{ij}.$$
(14)

In the formula, A is the sum of all edge weights in network G; e_{rr} is the proportion of the sum edge weight of community r occupying the edge weight sum of whole network; h_r is the proportion of the sum point strength of community r taken up in the whole network.

Then, it is essential to verify the former network model generated by four steps, where the modularity is Q_0 . Two communities X_r and X_s are chosen arbitrarily from the network, and then one point ν in community X_s is assigned to community X_r . Thus, the network modularity changes to Q_1 , and the variation of modularity ΔQ is

$$\begin{split} \Delta Q &= Q_0 - Q_1 \\ &= \left[e_{rr} - \left(h_r \right)^2 + e_{ss} - \left(h_s \right)^2 \right] \\ &- \left[e_{rr}' - \left(h_r' \right)^2 + e_{ss}' - \left(h_s' \right)^2 \right], \end{split}$$

$$q_{0} = e_{rr} - (h_{r})^{2} + e_{ss} - (h_{s})^{2},$$

$$q_{1} = e'_{rr} - (h'_{r})^{2} + e'_{ss} - (h'_{s})^{2},$$

$$q_{1} = \left[\frac{2A \cdot e_{rr} + 2A^{O}_{v,r}}{2A} - \left(\frac{2A \cdot h_{r} + A_{v}}{2A}\right)^{2}\right]$$

$$+ \left[\frac{2A \cdot e_{ss} - 2A^{I}_{v}}{2A} - \left(\frac{2A \cdot h_{s} - A_{v}}{2A}\right)^{2}\right],$$

$$\Delta Q = q_{0} - q_{1}$$

$$= \frac{A^{I}_{v}}{A} - \frac{A^{O}_{v,r}}{A} + 2\left(\frac{A_{v}}{2A}\right)^{2} + 2\left(h_{r} - h_{s}\right)\frac{A_{v}}{2A},$$

where $A_{v,r}^O$ is the sum edge weight of the adjacent edges between point v and inner points of community X_r , which names after the sum outer weight of community r of point v, recorded as $\max_r (A_{v,r}^O)$.

During the network generation, all communities are at the same level, but when the network scale is bigger, generally speaking, $h_r \approx h_s$. Thus,

$$\Delta Q \approx \frac{A_{\nu}^{I}}{A} - \frac{A_{\nu,r}^{O}}{A} + 2\left(\frac{A_{\nu}}{2A}\right)^{2}.$$
 (16)

(15)

During the whole process, it is assured that $\Delta Q \ge 0$.

When $A_{\nu}^{I}/A - A_{\nu,r}^{O}/A > 0$, it is known that $\Delta Q > 2(A_{\nu}/2A)^{2} > 0$. Thus, if $A_{\nu}^{I} > A_{\nu,r}^{O}$, it is indubitable that the former community division methodology always gets the optimal result. If $A_{\nu}^{I} > A_{\nu}^{O}$ (ν is the arbitrary point), it is surely satisfied to the optimal restrictions. During the network construction period, all points need to be connected to more inner points of community with higher contact strength and fewer outer points of community with lower contact strength.

After strict theoretical derivation and analyses, it is concluded that the sufficient condition to reach the optimal community division is $A^I_{\nu} > A^O_{\nu,r}$. Plenty of simulations should be done to analyze the characteristics.

3.3. Characteristics of the Network

3.3.1. Characteristics of the Edge Weight. After the weighted networks with community structure are generated, the characteristics of network are studied further. It focuses on the relationship of four characteristics and the community amount, or the coefficient of weight growth, separately. The four characteristics are the inner weight distribution of community, the outer weight distribution of community, the average inner weight of community, and the average outer weight distribution of community.

In Figure 3, there are 5000 points: $m = m_0 = 3$, $n = (S - 1) \times m$, and $\mu^I = \mu^O = 1$.

It is demonstrated in Figure 3 as log-log coordinate that the inner weight distributions of community obey the power law, which is $p(\alpha) \sim \alpha^{-\gamma}$. In terms of the inner weight distribution of community, γ slightly increases with the rise

Table 1: Relationship of the average edge weight and the community amount.

S	50	100	150	200	250
Mean (α^I)	0.111	0.182	0.206	0.216	0.238
Mean (α^{O})	0.062	0.089	0.097	0.145	0.135

of S; while in terms of the outer weight distribution of community, γ is generally high when S is large; anyway it is not strictly monotonous.

In Figure 4, there are 5000 points: S = 100, $m = m_0 = 3$, $n = (S - 1) \times m$, and $\mu^I = \mu^O = \mu$. The less the value of μ is, the larger the value of γ is which related to the inner weight of the community or the outer weight of the community.

Figures 3 and 4 demonstrate that every BBV network is isolated independently with different communities, because the inner weight distribution of each community obeys the law power. However, Figures 3(a) and 4(a) are both the statistical results of the inner weight within the community, and thus there is deviation between the simulation result and the power-law distribution.

Supposing that mean(α^I) indicates the mean of inner weighs, mean(α^O) signifies the mean of outer weight.

In Table 1, there are 5000 points: S = 100, $m = m_0 = 3$, $n = (S - 1) \times m$, $\mu^I = \mu^O = 1$.

With the increase of the community amount, the average inner weights of community get higher. Compared to the network with fewer communities, the average outer weight of community is relatively higher when the community amount is larger.

In Table 2, there are 5000 points: S = 100, $m = m_0 = 3$, $n = (S - 1) \times m$, and $\mu^I = \mu^O = \mu$.

It is demonstrated in Table 2 that with the growth of μ , both the inner and the outer weight distributions of communities have the trend of decrease.

After the simulation, it is concluded that the weight distributions obey the power law and satisfy the reality well. Moreover, the power-law distribution characteristics rely on the parameters obviously, hence by control of which, any type of networks can be built.

3.3.2. Characteristics of the Community Structure. In this section, we will analyze the sufficient condition to satisfy the optimal community division and explain whether the network has a clear community structure by calculating the value of the modularity.

In Figure 5, there are 5000 points: S = 100, $m = m_0 = 3$, $n = (S - 1) \times m$, and $\mu^I = \mu^O = 1$.

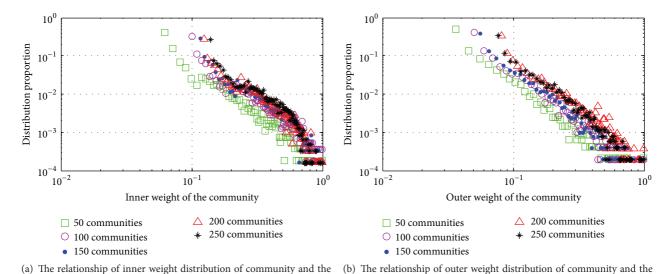
It is demonstrated that $A^I_{\nu}>\max_r(A^O_{\nu,r})$, and each point meets the condition that $A^I_{\nu}>A^O_{\nu,r}$. In conclusion, the community division with this method is optimal.

In Table 3, there are 5000 points: S = 100, $m = m_0 = 3$, $n = (S - 1) \times m$, and $\mu^I = \mu^O = 1$. And in Table 4, there are 5000 points: S = 100, $m = m_0 = 3$, $n = (S - 1) \times m$, and $\mu^I = \mu^O = \mu$.

In reality, the value range of Q in the network with community is [0.3, 0.7], and it is demonstrated that there is

0.5 0.55 0.6 0.65 0.7 0.8 0.9 1.0 1.5 Mean (α^I) 0.2806 0.2848 0.2885 0.2493 0.2519 0.2117 0.1846 0.1816 0.1067 Mean (α^{O}) 0.1929 0.1352 0.1098 0.0894 0.1858 0.1832 0.1431 0.1206 0.0611

TABLE 2: Relationship of the average edge weights and the coefficient of weight growth.



community amount community amount

FIGURE 3: Relationship of weight distribution and the community amount.

TABLE 3: Relationship of Q value and community amount.

S	50	100	150	200	250
Q	0.6439	0.6728	0.6950	0.6147	0.6640

obvious community structure in the network from Tables 3 and 4.

4. Characteristics of Epidemic Spreading

This section studies the specific characteristics of the epidemic spreading based on the aforementioned models, including the infection proportion of epidemic spreading in different stages, and the relationship of transmission rate and the community amount, or the average network weight, or the strength of the initial infectious point, separately.

4.1. Influence of Community Amount to Epidemic Spreading. It is defined that the epidemic transmission rate $d\eta$ means the increase rate of the infection proportion. During the simulation process, the transmission rate $\varphi(t)$ is defined as the amount of new infectious points in each simulation time step.

According to Figure 6, in terms of the infection proportion variation curve for 200 communities, given 4 infection proportions, namely, $\eta(t_1)$, $\eta(t_2)$, $\eta(t_3)$, $\eta(t_4)$, we know that $\eta(0) < \eta(t_1) < \eta(t_2) < \eta(t_3) < \eta(t_4) < \eta(\infty)$, corresponding to t_1, t_2, t_3, t_4 ($t_1 < t_2 < t_3 < t_4$), respectively. $\varphi(t_2)$ and $\varphi(t_3)$ are the transmission rate at t_2 and t_3 , given

 $\varphi_1 = \min(\varphi(t_2), \varphi(t_3))$, respectively. Therefore, the epidemic spreading process can be divided into 5 stages.

Stage 1 (the initial stage of spreading). $\eta(0) \leq \eta(t) < \eta(t_1)$, where $\eta(t)$ is small, and the transmission rate is slow while the infection people are gathered in a small range.

Stage 2 (the initial stage of outbreak). $\eta(t_1) \leq \eta(t) < \eta(t_2)$, and the transmission rate increases constantly but $\varphi(t) < \varphi(t_2)$.

Stage 3 (the middle stage of outbreak). $\eta(t_2) \le \eta(t) < \eta(t_3)$, and the infection rate is larger than some value $\varphi(t) \ge \varphi_1$.

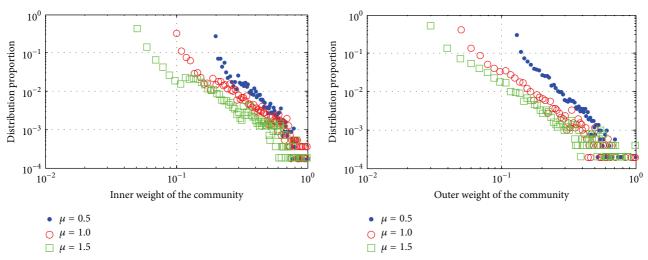
Stage 4 (the end stage of outbreak). $\eta(t_3) \le \eta(t) < \eta(t_4)$, and the transmission rate decreases constantly but $\varphi(t) < \varphi(t_3)$.

Stage 5 (the end stage of spreading). $\eta(t) > \eta(t_4)$, the infection proportion is larger than some value, and at this stage, the whole network is nearly infectious.

In comparison with the multiple simulations, the relationships between the infection proportion with different community amounts and the increase of time step length are shown in Figure 6.

In Figure 6, there are 5000 points, while the average inner weights of community are normalized as 0.1, and the average outer weights of community are normalized as 0.05.

The four curves are shown in Figure 6, when the average weights are analogous, the four curves have similar variation trend at Stages 2 to 5, and these four stages are irrelative to



(a) Relationship of inner weight distribution of community and the coefficient of weight growth

(b) Relationship of outer weight distribution of community and the coefficient of weight growth

FIGURE 4: Relationship of community weight distribution and the coefficient of weight growth.

TABLE 4: Relationship of Q value and the coefficient of weight growth.

μ	0.50	0.55	0.60	0.65	0.70	0.80	0.90	1.0	1.5
O	0.6090	0.6054	0.6158	0.6376	0.6554	0.6358	0.6281	0.6728	0.6397

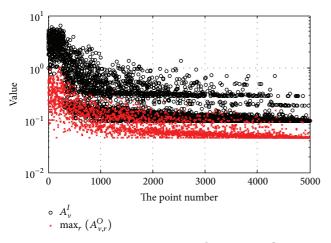


FIGURE 5: Comparison graph of A_v^I and $\max_r (A_{vr}^O)$.

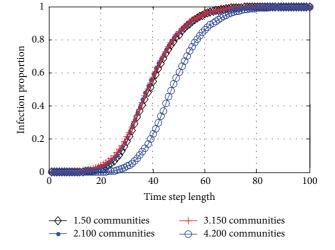


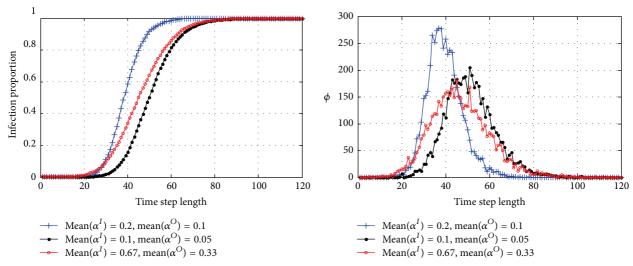
FIGURE 6: Simulation graph of infection conditions with different community amounts.

the community amount without any control measurements. Compared to curves 2 and 4, it is found that there is randomness in the initial spreading stage related to the sum weight of initial infectious points.

4.2. Influence of the Average Weight to Epidemic Spreading. In Figure 7, there are 5000 points and S = 100.

The later four stages are related to the average weight of the whole network, and the bigger the mean is, the larger the maximum infection proportion would be, and the transmission rate reaches the highest at the initial stage of outbreak. It is the middle stage of outbreak that the average weight has most effects on; therefore, reducing the weights in the network can postpone the outbreak to some extent.

The parameters related to the infection ability affect the five stages. People barely have any effective measures to master the epidemic during the initial stage of spreading. This paper focuses on the control of large-scale epidemic spreading and the effective measures of nodes isolation and



- (a) The relationship of the infection proportion and the average weights
- (b) The relationship of the infection rate and the average weights

FIGURE 7: Relationship between the average weights and the infection proportion or the infection rate.

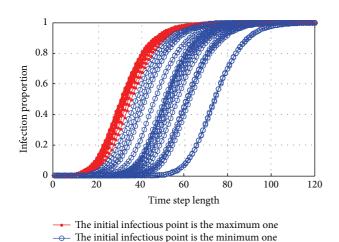


FIGURE 8: Relationship of infection proportion and the strength of initial infectious point.

edge weights, although the impact of these parameters on epidemic spreading is not further studied in detail.

4.3. Influence of Initial Infection Strength to Epidemic Spreading. In Figure 8, there are 5000 points, S = 100, mean(α^I) = 0.1, and mean(α^O) = 0.05. It has been simulated 16 times when the strength of the initial infectious point is the maximum and another 16 times when the strength of the initial infectious point is the minimum.

Figure 8 shows that when the strength of initial infectious point is large, the continuous time at the first stage of epidemic spreading is shorter and more centralized after several times of simulation; when the strength of initial infectious point is small, the continuous time is longer and distributed more randomly.

Through the simulation, it is concluded that the epidemic transmission rate has explicit relationship with the average weight of the network other than the community amount, especially at the later four stages of epidemic spreading. There is randomness at the first stage of epidemic spreading, while the larger the initial infectious point strength is, the less the randomness will be. Moreover, the first stage duration time is longer when the strength of the initial infectious point is less.

5. Measures to Control Epidemic Spreading

- 5.1. Measures to Isolate the Infectious Points. In reality, at the very beginning of the epidemic spreading process, it is hard to attract the public attention due to the minority of infectors [28]. However, we hope to control the epidemic at the initial stage of epidemic outbreak. Assuming there is a constant parameter related to the infection ability, it is concluded after several simulations that
 - (1) from the initial stage of outbreak to the end stage of epidemic spreading, the infection proportion is stable around the value i_{∞} ($i_{\infty} \le 1$);
 - (2) the stable i_{∞} ($i_{\infty} \le 1$) is related to t_0 and λ .

The simulation result is demonstrated as in Figure 9. There are 5000 points, S = 100, $m = m_0 = 3$, $n = (S-1) \times m = 6$, $\mu^I = \mu^O = 1$, $T_0 = 1$, and $\lambda = 0.1$. The amount of the initial infectious points is 10, and there are simulation curves at each t_0 .

It is defined that the extreme isolation delay t_0 is signed as m_{t0} which makes the final stable infection proportion i_{∞} satisfy $i_{\infty} \leq 0.9$, in terms of the infection probability λ at $T_0 = 1$.

In Table 5, there are 5000 points, S = 100, $m = m_0 = 3$, $n = (S - 1) \times m = 6$, and $\mu^I = \mu^O = 1$.

This section focuses on the influence of isolation delay on epidemic spreading and obtains that timely isolation can

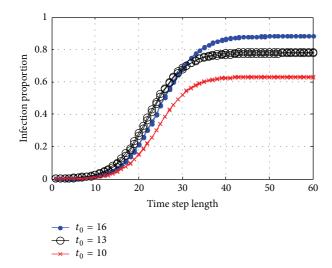


Figure 9: Relationship of η and the step length with different t_0 .

Table 5: The relationship of m_{t_0} and λ .

λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
m_{t_0}	16	11	9	8	7	6	5	4

reduce the final infection proportion i_{∞} , which indicates that there is direct relationship between i_{∞} and the isolation delay.

5.2. Measures to Reduce the Average Weight of Network. In reality, during the control process at the initial stage of the infectious epidemic, it exists infectivity for some viruses without any explicit symptoms [29]. Coupled with the isolation efficiency, it is difficult to put all infectious points isolated at the initial stage; in other words, it is hard to make t_0 less. Accordingly, if only taking measurement as point isolation, the epidemic infection proportion i_{∞} eventually reaches a large value, which is not conducive to control the epidemic spreading.

If the isolation measures are not taken in time, it is also required to control the whole network by other measures in order to better control the epidemic spreading. In reality, it is often acceptable that the healthy person should enhance the self-protection conscience and cut down the opportunities to contact with the infectious ones, which equals reducing the average weight of points and canceling some edges in the network model [30].

Nowadays, the widely acceptable methods to control the whole network include keeping away from public places and minimizing direct contact with each other as much as possible [31, 32]. However, if measures are carried out too early or too much, it will change people's living habits, increase the economic loss, and cause large-scale social panic, but if it is too late, it will lose the best opportunity to control the epidemic spreading.

It is demonstrated in Figure 7 that it can only postpone the epidemic outbreak other than cut down the final spreading proportion by reducing the average weight of the network. However, it is not useful to lessen the average weight of the

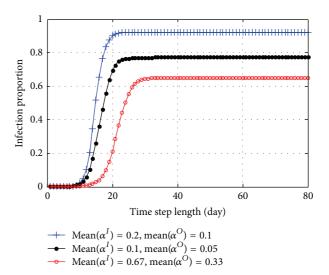


FIGURE 10: Relationship of infection proportion and time step length after points isolation.

network at Stage 1 (the initial stage of epidemic spreading) but to make point isolation. At Stage 2 (the initial stage of outbreak), if the isolation delay t_0 cannot meet the requirements even when the infection proportion becomes i_{∞} , it is essential to take measures to decrease the average weight of the network.

Figure 10 depicts the influence of different network average weights have on i_{∞} after the point isolation.

In Figure 10, there are 5000 points, S=100, $m=m_0=3$, $n=(S-1)\times m=6$, $T_0=1$, $\lambda=0.6$, and $t_0=6$. When the isolation measures are taken and the isolation delay t_0 is constant, the less the average weight of the network is, the smaller the value of i_{∞} would be.

In reality, once the epidemic breaks out, the isolation measures should be taken; then, we can analyze the characteristics of epidemic spreading by statistical method and study the isolation delay. It is necessary to take some measures to cut down the average weight of the network at Stage 2 (the initial stage of outbreak). When it comes to Stage 3 (the middle stage of outbreak), it is essential to decrease the average weight of the network, in order to make it come into Stage 4 (the end stage of outbreak) and Stage 5 (the end stage of spreading) as soon as possible.

6. Case Study on MERS

6.1. Case Study of Epidemic Spreading Characteristics. Epidemic spreading model based on contact strength was proposed in Section 2, while weighted network model with community structure was proposed in Section 3. However, real data is essential to verify whether the epidemic spreading result based on this model is persuasive.

From the first MERS infection case that was diagnosed definitely in South Korea, there have been 166 people infected until 20 June 2015. The change trend of infection population during these 31 days is shown in Figure 11.

It is not until June 7th when the South Korea government actually issued a response to MERS; in consequence, the 19 days from May 20th to June 7th should be regarded as the initial stage of the epidemic outbreak without interference measures, and the infection population is 87. The epidemic in this period mainly occurred in 3 hospitals, namely, Pyongtaek Mary Hospital, Konyang University Hospital, and Daejeon Hospital. They are considered as 3 communities (S=3), and the sum of points is regarded as the total amount of contact points in epidemic spreading network with the three communities. Then the fitting result of the simulation data and the real-world data are demonstrated in Figure 12, where the infection proportion η means the ratio of infectious points in the total amount of points in the network.

From May 20th to June 7th, the population is about 500–1000 in the contact network of the three hospitals. Hence, the simulation parameters in Figure 12 are as follows: in terms of the network model, the total amount of points is 600, S=3, $m=m_0=3$, $n=(S-1)\times m=6$, $\mu^I=0.2$, and $\mu^O=0.6$; in terms of the epidemic spreading model, $T_0=2.4$, $\lambda=0.1$, the initial amount of infectious points is 1. The real data curve is drawn by the MERS infectious population in South Korea divided into 600.

Overall, due to the fitting result in Figure 12, the real initial stage of epidemic outbreak can be explained well by the network model and the epidemic spreading model proposed in this paper.

In this section, the accuracy of models in this paper is verified by comparison of the real data.

6.2. Case Study of Control Measures. Figure 13 shows the comparison about the simulation result of infection proportion change with the increase of time step length and the real data after June 7th when the isolation measures were taken by South Korea government.

In Figure 13, there are 600 points, S=3, $m=m_0=3$, $n=(S-1)\times m=6$, $\mu^I=0.2$, and $\mu^O=0.6$, and in the epidemic spreading model, $T_0=2.4$, $\lambda=0.1$, the initial infection population is 1, and the isolation measure is taken at the 19th step, $t_0=4$. The real data curve is drawn by the MERS infectious population divided into 600.

It is demonstrated in Figure 13 that the curve can fit the real infection proportion statistics curve when the isolation delay $t_0=4$. However, the South Korea government took measure at Stage 2 (the initial stage of outbreak) actually, which was a little later. The simulation result is shown as in Figure 14 if the measures were taken at Stage 1 (the initial stage of epidemic spreading).

In Figure 14, the network model and epidemic spreading model are the same as Figure 13, $t_0=4$, but the isolation measure is taken at the first step. The 6 simulation result curves are the 6 results with maximum i_{∞} among the 100 times simulation.

Figure 14 demonstrates that if we could take measures at the very beginning of epidemic spreading, the final infection proportion i_{∞} would be reduced significantly even with the same isolation delay. However, due to the abruptness of epidemic outbreak, it is difficult to master control at

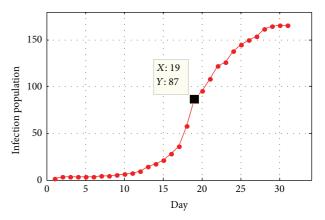


FIGURE 11: Statistics of MERS infection population in South Korea.

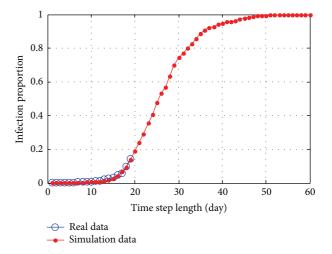


FIGURE 12: Fitting result of the simulation infection proportion and the real-world data.

the early stage. Consequently, it is essential to reduce the average weight of the network by some auxiliary measures, for example, dissolving some organizations with evident community structure (such as schools), decreasing the stay time in public places, or canceling some public activities.

7. Conclusion

MERS viruses are spreading swiftly at the very beginning period among the intimate human-to-human contact communities. The control measurements and the optimal opportunity to take action are proposed to study the spreading characteristics. Assuming that the contact strength has a linear relationship with time, the epidemic spreading model is established based on the SI model, which is the foundation of analytical solution research. The weighted network model with community structure is constructed on the basis of the BBV network model, deducing that the generated network meets the sufficient conditions for the optimal community division. The simulation results show that the weight distributions within the community and among the communities both obey the power-law distribution.

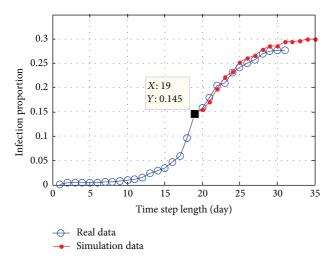


FIGURE 13: Comparative graph of simulation result and real data after the measures.

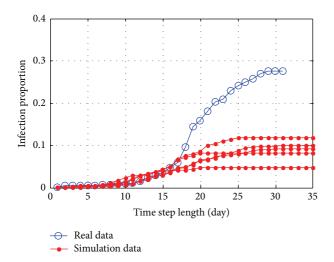


FIGURE 14: Simulation result when the isolation measures are taken at Stage 1.

According to the epidemic spreading network model, this research focuses on the characteristics of virus transmission. Hence, the epidemic spreading process would be divided into five stages: the initial stage of spreading, the initial stage of outbreak, the middle stage of outbreak, the end stage of outbreak, and the end stage of spreading. Then the simulation results show that the duration time of Stage 1 is related to the sum weight of the initial infection nodes, while those of Stages 2 to 5 are only related to the average weight instead of the community amount. However, without isolation measures, the decrease of the average weight only postpones spreading without interdicting it.

There is always a time delay before patients are isolated; thus, we study on the effects and optimal opportunity to master the epidemic spreading by two approaches: the point isolation and the average weight of network reduction. In conclusion, shorter time delay would reduce the final infection proportion, and with the same time delay, the

decline of the average weight would further reduce the final infection proportion. The simulation data matches the real data of MERS in South Korea well. Therefore, the measures are adopted as follows: firstly isolate the infectious points as soon as possible; if the infection proportion is higher than the setting threshold, then the average weight should be cut down to the corresponding value (e.g., canceling public activities, dissolving some organizations, and keeping away from the crowded places).

The models in this paper are also appropriate for the diffusion of terrorist forces or evil forces with intimate community structures. Since the proposed epidemic spreading model is based on the SI model, it is hard to simulate the condition in which infectious people recover or die after a period of epidemic outbreak. Moreover, it is supposed that there is a linear relationship between contact strength and contact time, which cannot mimic the reality very well. In the future, we will study on different diseases with different epidemic spreading models on the proposed weighted network model with community structure after the period of large-scale infection or immunization.

Conflict of Interests

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

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant nos. 71331008 and 71501182.

References

- [1] R. M. Anderson, R. M. May, and B. Anderson, "The mathematics of infection. (Book reviews: infectious diseases of humans. dynamics and control)," *Science*, vol. 254, no. 2, pp. 591–592, 1991.
- [2] N. Bailey, "The mathematical theory of infectious diseases and its applications," *Immunology*, vol. 34, no. 5, pp. 955–956, 1978.
- [3] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.
- [4] M. E. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [5] C. Li, W. Hu, and T. Huang, "Stability and bifurcation analysis of a modified epidemic model for computer viruses," *Mathematical Problems in Engineering*, vol. 2014, Article ID 784684, 14 pages, 2014.
- [6] C. Fraser, C. A. Donnelly, S. Cauchemez et al., "Pandemic potential of a strain of influenza A (H1N1): early findings," *Science*, vol. 324, no. 5934, pp. 1557–1561, 2009.
- [7] M. E. J. Newman, S. Forrest, and J. Balthrop, "Email networks and the spread of computer viruses," *Physical Review E*, vol. 66, no. 3, Article ID 035101, 2002.
- [8] R. Pastor-Satorras and A. Vespignani, "Epidemic dynamics in finite size scale-free networks," *Physical Review E*, vol. 65, no. 3, Article ID 035108, 2002.
- [9] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading in real networks: an eigenvalue viewpoint,"

- in *Proceedings of the 22nd International Symposium on Reliable Distributed Systems (SRDS '03)*, pp. 25–34, IEEE, Florence, Italy, October 2003.
- [10] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," ACM Transactions on Information and System Security, vol. 10, no. 4, pp. 1311–1326, 2008.
- [11] P. Erdös, "On the structure of linear graphs," *Israel Journal of Mathematics*, vol. 1, no. 3, pp. 156–160, 1963.
- [12] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [13] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [14] A. Barrat, M. Barthélemy, and A. Vespignan, "Modeling the evolution of weighted networks," *Physical Review E*, vol. 70, no. 6, Article ID 066149, 12 pages, 2004.
- [15] 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.
- [16] P. M. Gleiser and L. Danon, "Community structure in jazz," Advances in Complex Systems, vol. 6, no. 4, pp. 565–573, 2003.
- [17] M. E. J. Newman, "Communities, modules and large-scale structure in networks," *Nature Physics*, vol. 8, no. 1, pp. 25–31, 2012
- [18] J.-P. Zhang and Z. Jin, "Epidemic spreading on complex networks with community structure," *Applied Mathematics and Computation*, vol. 219, no. 6, pp. 2829–2838, 2012.
- [19] B. Serrour, A. Arenas, and S. Gómez, "Detecting communities of triangles in complex networks using spectral optimization," *Computer Communications*, vol. 34, no. 5, pp. 629–634, 2011.
- [20] D. Zhang, F. Xie, Y. Zhang, F. Dong, and K. Hirota, "Fuzzy analysis of community detection in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 22, pp. 5319–5327, 2010.
- [21] Z. Liu and B. Hu, "Epidemic spreading in community networks," *Europhysics Letters*, vol. 72, no. 2, pp. 315–321, 2005.
- [22] T. Smieszek, L. Fiebig, and R. W. Scholz, "Models of epidemics: When contact repetition and clustering should be included," *Theoretical Biology and Medical Modelling*, vol. 6, article 11, 2009.
- [23] M. Salathé and J. H. Jones, "Dynamics and control of diseases in networks with community structure," *PLoS Computational Biology*, vol. 6, no. 4, pp. 387–395, 2010.
- [24] C.-H. Li, C.-C. Tsai, and S.-Y. Yang, "Analysis of epidemic spreading of an SIRS model in complex heterogeneous networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 4, pp. 1042–1054, 2014.
- [25] Y. Wang, G. Xiao, J. Hu, T. H. Cheng, and L. Wang, "Imperfect targeted immunization in scale-free networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 388, no. 12, pp. 2535–2546, 2009.
- [26] W. Zhao, J. Li, and X. Meng, "Dynamical analysis of SIR epidemic model with nonlinear pulse vaccination and lifelong immunity," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 848623, 10 pages, 2015.
- [27] P. Mundur, M. Seligman, and G. Lee, "Epidemic routing with immunity in Delay Tolerant Networks," in *Proceedings of IEEE Military Communications Conference (MILCOM '08)*, pp. 1–7, San Diego, Calif, USA, November 2008.
- [28] K. Gao and D. Hua, "Effects of immunity on global oscillations in epidemic spreading in small-world networks," *Physics Procedia*, vol. 3, no. 5, pp. 1801–1809, 2010.

- [29] C.-Y. Xia, Z. Wang, J. Sanz, S. Meloni, and Y. Moreno, "Effects of delayed recovery and nonuniform transmission on the spreading of diseases in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 7, pp. 1577–1585, 2013
- [30] N. Gupta, A. Singh, and H. Cherifi, "Community-based immunization strategies for epidemic control," in *Proceeding of the 7th IEEE International Conference on Communication Systems and Network*, pp. 1–6, Chengdu, China, 2015.
- [31] X. Fu, M. Small, D. M. Walker, and H. Zhang, "Epidemic dynamics on scale-free networks with piecewise linear infectivity and immunization," *Physical Review E*, vol. 77, no. 3, Article ID 036113, pp. 420–427, 2008.
- [32] S. Cauchemez, A. Bhattarai, T. L. Marchbanks et al., "Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 7, pp. 2825–2830, 2011.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 538482, 8 pages http://dx.doi.org/10.1155/2015/538482

Research Article

An Efficient Top-*k* **Query Processing with Result Integrity Verification in Two-Tiered Wireless Sensor Networks**

Ruiliang He, 1 Hua Dai, 1 Geng Yang, 1 Taochun Wang, 2 and Jingjing Bao 3

¹College of Computer Science & Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Correspondence should be addressed to Hua Dai; daihua@njupt.edu.cn

Received 9 April 2015; Revised 13 July 2015; Accepted 27 July 2015

Academic Editor: Mark Leeson

Copyright © 2015 Ruiliang He 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 two-tiered wireless sensor networks, storage nodes take charge of both storing the sensing data items and processing the query request issued by the base station. Due to their important role, storage nodes are more attractive to adversaries in a hostile environment. Once a storage node is compromised, attackers may falsify or abandon the data when answering the query issued by the base station, which will make the base station get incorrect or incomplete result. This paper proposes an efficient top-k query processing scheme with result integrity verification named as ETQ-RIV in two-tiered sensor networks. According to the basic idea that sensor nodes submit some encoded message containing the sequence relationship as proof information for verification along with their collected sensing data items, a data binding and collecting protocol and a verifiable query response protocol are proposed and described in detail. Detailed quantitative analysis and evaluation experiments show that ETQ-RIV performs better than the existing work in both communication cost and query result redundancy rate.

1. Introduction

Since the traditional multihop architecture is not suitable for large-scale wireless sensor networks (WSNs), a novel twotiered architecture has been proposed. The two-tiered wireless sensor networks (TWSNs) introduce storage nodes that are abundant in energy, memory, and computing power to traditional multihop WSNs. In such a two-tiered architecture, the storage node serves as an intermediate tier between the base station and the sensor nodes, which is shown in Figure 1. The whole network is partitioned into several cells, each of which consists of a storage node and some sensor nodes nearby. The sensing data are sent to and stored in the storage node in the same cell after being collected by the sensor nodes. After receiving the queries issued by the base station, the storage node processes the query over the data items that are received from the sensor nodes and then returns the query result to the base station. The two-tiered architecture is also known to be indispensable for increasing network capacity

and scalability, reducing system complexity, and prolonging network lifetime [1–3].

In TWSNs, storage nodes not only store all the sensing data of all the sensor nodes in the same cell, but also respond to the query requests issued by the base station, which makes the storage nodes more attractive to adversaries. Once a storage node is compromised, attackers may falsify or abandon the data when the storage node is answering the query issued by the base station, which will make the base station get incorrect or incomplete result. In an application that depends on the query result, an incomplete result may lead to wrong decisions. Therefore, it is of great significance to construct a verifiable query processing scheme, by which authenticity and integrity of query results can be verified by the base station.

Top-k query is frequently used in WSNs aiming at getting the k highest or lowest data in a specified region during a specified time epoch. For instance, "to get the 5 highest temperature data of the second warehouse from 12:00 to

²College of Mathematics and Computer Science, Anhui Normal University, Wuhu 241003, China

³Tongda College of Nanjing University of Posts and Telecommunications, Nanjing 210003, China

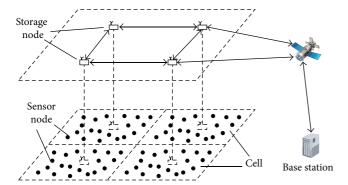


FIGURE 1: Network model of TWSN.

13:00" is a typical top-k query that could be used as fire detection. This paper proposes an efficient top-k query processing scheme with result integrity verification in TWSNs, named as ETQ-RIV, the basic idea of which is as follows. The sensor nodes submit some encoded message containing the sequence relationship along with their collected sensing data items to the storage nodes. When the base station issues queries, storage nodes send all the satisfied data items along with the encoded message as proof information to the base station. Then the base station can compute the result and verify the authenticity and integrity of it.

ETQ-RIV is an improvement scheme of EVTQ proposed in our previous work [4]. The improvements are as follows.

- (i) In the data collecting process, each sensor node does not need to submit the out-bound HMAC, which further reduces the in-cell communication.
- (ii) With processing query, the storage node needs to send one HMAC for each unqualified sensor node in EVTQ. However, in ETQ-RIV, only one HMAC for all noncontributed sensor nodes is required, which will greatly reduce the query communication cost.

The main contributions of this paper are as follows.

- (i) We propose a sequence relationship encoding based method, which makes storage nodes unable to falsify or omit data items without being noticed.
- (ii) We give the two concrete protocols, named as *data* binding and collecting protocol and verifiable query response protocol, to achieve ETQ-RIV, which make the base station capable of verifying the authenticity and integrity of the top-k query result.
- (iii) We present the quantitative analysis of communication cost and redundancy rate of *ETQ-RIV* in detail and conduct experiments to evaluate the performance of this work compared to present methods.

The rest of this paper is organized as follows. The related works are presented in Section 2. Section 3 presents the preliminaries including related models, problem description, and performance evaluation index. Section 4 presents the two protocols proposed in this paper. Section 5 presents the theoretical analysis of the communication cost and

redundancy rate. Section 6 presents the result and analysis of the simulation experiments. In last section, we make a conclusion for this paper.

2. Related Works

The traditional multihop model is not suitable for large scale wireless sensor networks; thus a novel two-tiered architecture was proposed, which is indispensable for increasing network capacity and scalability, reducing system complexity, and prolonging network lifetime [1–3].

In TWSNs, the storage nodes play an important role and are more attractive to adversaries, which may bring serious data security problems. Therefore, data privacy and security have been widely discussed in recent research work.

Attentions have been paid to secure top-k query processing in two-tiered sensor networks recently. A novel verifiable fine-grained top-k query processing scheme was proposed by Zhang et al. [5], which is the first work of verifiable topk query scheme in two-tiered sensor networks. The basic idea of Zhang's scheme is that each sensor node generates hashed message authentication code (HMAC) [6] for every three consecutive data items to make query results verifiable. However, the long bits of HMACs result in high cost of communications. Liao and Li proposed a secure top-k query scheme named PriSecTopk [7] based on order preserving encryption [8] and message authentication code (MAC) [9], which could not lower the communication cost either. Ma et al. proposed a novel fine-grained verification scheme for top-k queries named VSFTQ [10] which uses symmetric encryption instead of message authentication. VSFTQ reduces the communication cost to a certain extent. On the basis of Zhang's scheme and Ma's scheme, Dai et al. proposed an efficient verifiable top-k query processing scheme in our previous work [4], which has a better performance in communication cost. Yu et al. proposed a dummy readingbased anonymization framework [11, 12], under which the query results can be guaranteed by verifiable top-k query (VQ) schemes they proposed. However, the VQ scheme requires each sensor node to send the neighboring sensor node IDs and hashes corresponding to individual genuine encrypted readings, which may lead to high communication cost of sensor nodes. Moreover, to obtain the top-k query result, the storage node is required to submit a top- $(\eta - 1 + k)$ query result due to the dummy readings, where η is a system parameter denoting the difference between the maximum and minimum encrypted readings within an epoch.

There is a special kind of top-*k* query when *k* equals one, which is also known as max/min query. For max/min query processing, Yao et al. proposed a preliminary privacy-preserving scheme [13] based on prefix membership verification (PMV) mechanism [14, 15] to compute the maximum or minimum value over encrypted data items. Dai et al. proposed an energy-efficient privacy-preserving MAX/MIN query processing solution [16] based on 0-1 encoding technique [17] which has a better performance compared to [13].

Besides top-k query, secure range query has been widely discussed recently [18–23]. The schemes proposed in [18–23]

ask for data items falling into specified ranges without data privacy disclosure and make query result verifiable.

3. Models and Problem Statement

3.1. Models. In this paper we consider the two-tiered network model as shown in Figure 1. The whole network is partitioned into *cells*, each of which consists of a storage node M and n homogeneous sensor nodes $\{s_1, s_2, \ldots, s_n\}$. M is abundant in resources such as energy, storage, and computation, which is in charge of not only storing of all sensing data collected by the sensor nodes in the same cell but processing and answering of the query request issued by the base station. Limited in resource, the sensor nodes just collect sensing data and transmit them to the storage node in the cell. As in [4], we take the assumption that storage nodes know the topological information of the whole cell, while a sensor node knows the locations of its neighboring nodes in one hop as well as the storage node's location of affiliated cells. In addition, the base station knows the topological information of the whole network.

In TWSN, a top-k query can be denoted as a three-element tuple Q = (c, t, k), where c represents the query region, t represents the time epoch of the query, and k is the quantity of required data items. For simplicity of description, we discuss the specific top-k query that covers only one cell in one time epoch. It is easy to extend the proposed method to achieve queries including multiple cells over multiple time epochs.

3.2. Problem Statement. The threat model considered in this paper is similar to [4]. We assume that an arbitrary number of storage nodes could be compromised and instructed to respond with falsified and/or incomplete data as a top-k query result to the base station. In TWSN, both sensor nodes and storage nodes could be compromised. In general, a sensor node only has very little information and the vast majority in the whole network are uncompromised ones. However, the storage node plays a more important role in the network, for it stores all data items collected in the whole cell. Once a storage node is compromised, the adversary may falsify or abandon the data when the storage node answering the query issued by the base station, which will make the base station get incorrect or incomplete result and mislead the user into making wrong decisions. Therefore, storage nodes tend to be more attractive and vulnerable to adversaries. Although a compromised storage node may lead to leakage of data privacy, it is not concerned in this paper. In practice, in some application of WSNs, it is the data integrity instead of data privacy that is more important. For instance, video surveillance in a sensor network for building security is known to adversaries thus requiring no privacy. Therefore, we focus on the verifiable top-*k* query in this paper.

Given a top-k query Q = (c, t, k), we assume the storage node M returns a data set R as query result. The problem of interest is how the base station can verify whether R satisfies both the *authenticity* and *integrity* requirements, which means the following two rules must hold.

- (1) *Authenticity Rule*. All data items in *R* are surely collected by sensor nodes in the query cell *c* during time epoch *t*.
- (2) *Integrity Rule*. There are exactly *k* data items that are indeed the *k* highest data items among what are collected by sensor nodes in the query cell *c* during time epoch *t*, or equivalently

$$\forall r_i \in R_t \land \forall d_i \in D_t \setminus R_t \longrightarrow r_i > d_i, \tag{1}$$

where D_t represents the data set consisting of all data items collected by sensor nodes in the query cell c during time epoch t.

- *3.3. Evaluation Metrics.* In this paper, the following two metrics are used to evaluate the performance of the proposed scheme.
 - (1) Communication cost: including in-cell communication cost C_I and query processing communication cost C_Q . C_I represents the size in bits of data items transmitted from all the sensor nodes to the storage in a cell in the data binding and collecting procedure, while C_Q stands for the size of data items transmitted from the storage node in the query cell to the base station during the verifiable query response procedure.
 - (2) Redundancy rate of query result: the proportion of the total size in bits of additional proof information used to enable verifiable top-k queries to the total size in bits of response message in final query result. This rate, denoted as γ , indicates the efficiency of the query processing. A lower redundancy rate means less additional communication cost required for verification. The redundancy rate will be calculated by the following equation:

$$\gamma = \frac{l_s - l_r}{l_s} \times 100\%,\tag{2}$$

where l_s is the size in bits of the total data items returned by the storage node, while l_r is the total data size of final query result.

4. Verifiable Top-k Query Processing

- 4.1. Assumptions and Definition. Given a top-k query $Q_t = (c,t,k)$, we assume that the query covers just a cell c in a time epoch t, which we have mentioned in previous sections. There are n sensor nodes in the cell c, which can be denoted as $\Gamma = \{s_1, s_2, \ldots, s_n\}$. Assume that an arbitrary sensor node s_i $(1 \le i \le n)$ collects N data items denoted as $D_i = \{d_{i,1}, d_{i,2}, \ldots, d_{i,N}\}$ in each time epoch and sorts them into descending order. s_i shares a secret key K_i with the base station, which is used to encode the proof information for each data item by a HMAC function.
- 4.2. Data Binding and Collecting Protocol. In each time epoch, sensor node s_i collects data and encodes them before sending to the storage node M in the same cell for storage. The detailed procedure is as follows.

- (1) s_i sorts the N data items $D_i = \{d_{i,1}, d_{i,2}, \dots, d_{i,N}\}$ that it collects in time epoch t. Without loss of generality, we assume $d_{i,1} > d_{i,2} > \dots > d_{i,N}$ after being sorted.
- (2) According to the sequence of the sensing data items, s_i computes the message verification code $V(d_{i,j})$ for each data item. $V(d_{i,j})$ of the jth data item $d_{i,j}$ $(1 \le j \le N)$ can be computed as follows:

$$V(d_{i,j}) = \text{HMAC}_{K_i}(t \parallel d_{i,1} \parallel d_{i,2} \parallel \cdots \parallel d_{i,j}),$$
 (3)

where $\mathrm{HMAC}_{K_i}(\cdot)$ denotes encoding the corresponding data using a HMAC function with key K_i and \parallel is the concatenation operator.

(3) s_i constructs a data collecting message MSG_C according to the following format and sends to the storage node M:

$$MSG_{C} = \langle id(s_{i}), t, d_{i,1}, V(d_{i,1}), d_{i,2}, V(d_{i,2}), \dots, d_{i,N}, V(d_{i,N}) \rangle.$$
(4)

(4) M receives and stores the data collecting message of s_i. We denote the data set consisting of all the data items from all the sensor nodes in the cell as

$$DS_t = \bigcup_{s_i \in \Gamma} \left\{ d_{i,1}, \dots, d_{i,N} \right\}.$$
 (5)

4.3. Verifiable Query Response Protocol. Query processing and verification requires collaboration of the base station and the storage node, which works as follows. According to the query issued by the base station, the storage node M processes this query and responds with the corresponding data items. At last, the base station computes the query result and verifies the authenticity and integrity of the result. The protocol is described in detail as follows.

Phase 1 (query request transmission). The base station sends the query $Q_t = (t, k, c)$ to M and waits for its feedback.

Phase 2 (query message feedback). (1) After Q_t has been received, M computes the k highest data items, denoted as $topk(DS_t)$, according to all the sensing data DS_t which was sent during the time epoch t from the sensor nodes $\{s_1, s_2, \ldots, s_n\}$ in cell C. $topk(DS_t)$ satisfies the following condition:

$$|\operatorname{top} k(\mathrm{DS}_t)| = k \wedge \forall d_i,$$

$$d_i (d_i \in \operatorname{top} k(\mathrm{DS}_t) \wedge d_i \in \mathrm{DS}_t \setminus \operatorname{top} k(\mathrm{DS}_t)) \longrightarrow d_i > d_i.$$
(6)

Assuming that there are δ_i data items sent by s_i within $topk(DS_t)$, that is to say, $\{d_{i,1}, d_{i,2}, \ldots, d_{i,\delta_i}\} \subseteq topk(DS_t)$, then the following condition holds:

$$\delta_i = |D_i \cap \text{top}(DS_t)|, \quad 0 \le \delta_i \le \min\{N, k\}.$$
 (7)

(2) M constructs the following response message according to δ_i .

(i) Given a sensor node s_i where $\delta_i > 0$, we call it the contributed node, the response message of which is named as $msg_1(s_i)$ and it should be computed as

 $msg_1(s_i)$

$$=\begin{cases} \left\langle \operatorname{id}\left(s_{i}\right), d_{i,1}, \dots, d_{i,\delta_{i}+1}, V\left(d_{i,\delta_{i}+1}\right)\right\rangle, & 0 < \delta_{i} \leq N-1, \quad (8) \\ \left\langle \operatorname{id}\left(s_{i}\right), d_{i,1}, \dots, d_{i,N}, V\left(d_{i,N}\right)\right\rangle, & \delta_{i} = N, \end{cases}$$

where d_{i,δ_i+1} is the maximum data item collected by s_i that is not in $topk(DS_t)$ and we call it the out-bound of the s_i . If all the data items collected by s_i are in $topk(DS_t)$, its out-bound does not exist. Otherwise, there will be one and only one out-bound as for s_i .

(ii) We call the sensor node where $\delta_i = 0$ noncontributed node. There is only one response message msg_0 for all noncontributed nodes which should be computed as

$$\operatorname{msg}_{0} = \left\langle \left\{ \operatorname{id}\left(s_{i}\right), d_{i,1} \mid \delta_{i} = 0 \land s_{i} \in \Gamma \right\},$$

$$\bigoplus_{\delta_{i} = 0 \land s_{i} \in \Gamma} V\left(d_{i,1}\right) \right\rangle,$$
(9)

where \bigoplus is the exclusive or operator.

(3) M summarizes all the response message generated in step (2), constructs the query feedback message MSG_Q , and sends it to the base station:

$$MSG_O = \langle \{ msg_1(s_i) \mid \delta_i > 0 \land s_i \in \Gamma \}, msg_0 \rangle.$$
 (10)

Phase 3 (query result computation and verification). (1) Upon receiving the query feedback message from M, the base station will do the preprocessing to confirm all the message $\{ \text{msg}_1(s_i) \mid \delta_i > 0 \land s_i \in \Gamma \}$ of each contributed node and the message msg_0 of all noncontributed nodes.

- (2) The base station sorts all data items in the response message and gets the k highest data items, which is the top-k query result. We denote the top-k query result by R_t , the minimum data item of which is denoted as $\min(R_t)$.
- (3) The base station checks whether the following conditions hold in sequence. If and only if all conditions are satisfied, the query result R_t is an authentic and complete result. Otherwise, the query result is abnormal.

Condition 1. All sensor ID in message $\{ \text{msg}_1(s_i) \mid \delta_i > 0 \land s_i \in \Gamma \}$ and msg_0 construct a set, named Ω , so that $\Omega = \{ \text{id}(s_i) \mid s_i \in \Gamma \}$ holds.

Condition 2. Assume that s_i is an arbitrary sensor node that contributes data items to message $\{ \operatorname{msg}_1(s_i) \mid \delta_i > 0 \land s_i \in \Gamma \}$, and the response message of s_i is $\langle \operatorname{id}(s_i), d_{i,1}, d_{i,2}, \ldots, d_{i,\mu_i}, H_i \rangle$, where $d_{i,1} > d_{i,2} > \cdots > d_{i,\mu_i}$. The base station then computes $\operatorname{HMAC}_{K_i}(t \parallel d_{i,1} \parallel \cdots \parallel d_{i,\mu_i})$, where K_i is a distinct key known only to s_i and the base station. Then one of the following two conditions must hold:

(1)
$$\mu_i = N \wedge H_i = H_{k_i}(t \parallel d_{i,1} \parallel \cdots \parallel d_{i,N} \parallel \Psi) \wedge d_{i,l_i} \ge \min(R_t);$$

(2)
$$2 \le \mu_i < N \land H_i = H_{k_i}(t \parallel d_{i,1} \parallel \cdots \parallel d_{i,N} \parallel \Psi) \land d_{i,k_i}$$

Condition 3. Let $\{s_1, s_2, \ldots, s_p\}$ be all the noncontributed nodes and their response message msg_0 is $\langle \operatorname{id}(s_1), d_1, \operatorname{id}(s_2), d_2, \ldots, \operatorname{id}(s_p), d_p, H_0 \rangle$, in which all the data items construct a set $\operatorname{DS}' = \{d_1, d_2, \ldots, d_p\}$. The base station computes $\operatorname{HMAC}_{K_1}(t \parallel d_1), \operatorname{HMAC}_{K_2}(t \parallel d_2), \ldots, \operatorname{HMAC}_{K_p}(t \parallel d_p)$ using the key K_1, K_2, \ldots, K_p shared with s_1, s_2, \ldots, s_p , respectively, and the following condition holds:

$$\left(H_{0} = \bigoplus_{d_{j} \in DS'} HMAC_{K_{j}} (t \parallel d_{j})\right)$$

$$\wedge \left(\forall d_{j} \in DS' \longrightarrow d_{j} < \min(R_{t})\right).$$
(11)

As described in previous two protocols, the sensing data items collected by a sensor node are sorted in descending order and encoded with a HMAC function. Each sensor node shares a secret key only with the base station, and the storage nodes know nothing about the keys. In such a case, it is computationally infeasible for the storage nodes to falsify the message verification code, as long as we choose a considerably complex HMAC algorithm such as SHA-1 [24]. Thus, it is impossible for storage nodes to falsify or conceal data items in the query result without being detected by the base station. Therefore, the scheme proposed in this paper enables authenticity and integrity verification of the top-*k* query result.

5. Protocol Analysis

5.1. Communication Cost Analysis. From data binding and collecting protocol and verifiable query response protocol we learn that there are two kinds of communication costs in two-tiered sensor networks, named as in-cell communication cost C_I and query processing communication cost C_Q . Assume that there are n sensor nodes in a cell and each node ID has $l_{\rm id}$ bits. Each sensor node collects N data items in every time epoch, each data item has l_d bits, and each HMAC code number has l_h bits. In addition, each time epoch number has l_t bits. Assume that there are L hops between each sensor node and the storage node on average. We assume that there are μ contributed nodes and n- μ noncontributed nodes, and λ of the μ contributed nodes contributes all the N data items to the query result. Obviously, we have $0 \le \lambda \le N$ and $\lambda \le k$ hold. According to such two protocols, we have

$$C_{I} = n \cdot (l_{id} + l_{t} + N \cdot (l_{d} + l_{h})) \cdot L,$$

$$C_{Q} = (\mu \cdot l_{id} + (k + \mu - \lambda) \cdot l_{d} + \mu \cdot l_{h})$$

$$+ ((n - \mu) \cdot l_{id} + (n - \mu) \cdot l_{d} + l_{h})$$

$$= n \cdot l_{id} + (n + k - \lambda) \cdot l_{d} + (\mu + 1) \cdot l_{h}.$$
(12)

5.2. Redundancy Rate Analysis. We define the redundancy rate γ of query result as the proportion of the total size of additional proof information used for query result verification to the total size of response message in final query result.

Table 1: Default evaluation parameters.

Para.	n	l _{id} (bit)	l_t (bit)	l_d (bit)	l_h (bit)	N	k
Val.	250	32	32	32	128	20	10

We take the same assumption as Section 5.1. According to our proposed protocols, we have

$$\gamma = \frac{C_{Q} - (\mu \cdot l_{id} + k \cdot l_{d})}{C_{Q}} \times 100\%$$

$$= \frac{(n - \mu) \cdot l_{id} + (n - \lambda) \cdot l_{d} + (\mu + 1) \cdot l_{h}}{n \cdot l_{id} + (n + k - \lambda) \cdot l_{d} + (\mu + 1) \cdot l_{h}} \times 100\%.$$
(13)

6. Performance Evaluation

In this section, we evaluate the performance of the scheme ETQ-RIV proposed in this paper and compare it with the schemes VSFTQ, EVTQ, and AD-VQ, which are proposed in [10], [4], and [11, 12], respectively. The four schemes VSFTQ, EVTQ, ETQ-RIV, and AD-VQ are implemented on the simulator of [25] with random sensor data items. We compare the in-cell communication cost C_I , query processing communication cost C_Q , and the query result redundancy rate γ of the four schemes, respectively. We also assume that the packet transmissions are both collision-free and error-free in our experiments.

We take the assumption that we carry out the top-k query just in one cell with one storage node and n sensor nodes. The n sensor nodes are distributed uniformly over a two-dimensional region which covers an $80*80\,\mathrm{m}^2$ area. Each sensor node collects N data items during each time epoch and its communication radius is 10 meters. The bit lengths of sensor node ID, time epoch number, sensing data item, and HMAC code are represented by l_{id} , l_t , l_d , and l_h , respectively. The default parameters are listed in Table 1, which are used in the evaluations unless otherwise specified.

In each measurement, we generate 20 different networks with different network IDs. In each network, the sensor nodes are distributed randomly with different topology. The measurement result is the average of 20 networks.

6.1. In-Cell Communication Cost Evaluation. With default parameters, we evaluate the in-cell communication costs of VSFTQ, EVTQ, ETQ-RIV, and AD-VQ in different networks, the result of which is shown in Figure 2(a). Figure 2(a) indicates that our proposed scheme ETQ-RIV takes a little less communication costs than VSFTQ and AD-VQ, while EVTQ takes the highest communication costs. Compared with EVTQ, AD-VQ, and VSFTQ, ETQ-RIV saves about 3.77%, 2.86%, and 0.97% of in-cell communication costs on average, respectively. The reason is that in each scheme sensor nodes are required to submit every data item along with some proof information. EVTQ, AD-VQ, and ETQ-RIV use HMAC as proof information. In EVTQ, each sensor node submits one more out-bound HMAC besides every data item's proof information. In AD-VQ, each sensor requires submitting some information of neighboring nodes besides

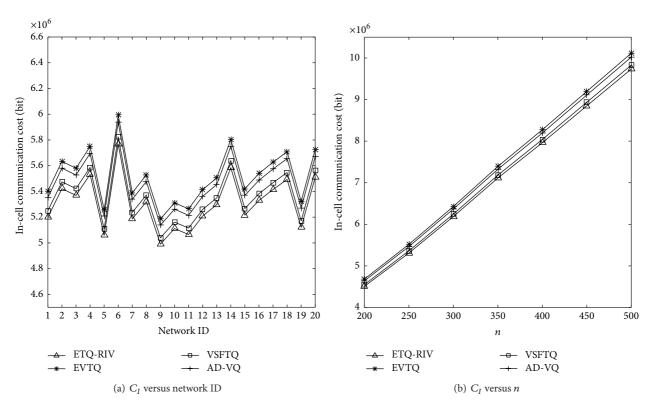


FIGURE 2: Evaluation results of in-cell communication costs.

the HMAC information. In ETQ-RIV, the sensor nodes do not need the out-bound HMAC and neighboring nodes information, so ETQ-RIV performs better than EVTQ and AD-VQ. Different from the other three schemes, VSFTQ replaces the HMAC with symmetric encryption of data item's order, score, and the time epoch as proof information. In our evaluation, the total bit length of proof information in VSFTQ is more than ETQ-RIV. As a result, ETQ-RIV performs better than VSFTQ.

Figure 2(b) indicates that the in-cell communication costs of the three schemes increase as the sensor node number *n* increases. ETQ-RIV has the best performance as before and saves about 3.77%, 2.86%, and 0.97% of in-cell communication costs on average compared with EVTQ, AD-VQ, and VSFTQ, respectively. The reason is same as the description in the previous section.

6.2. Query Communication Cost Evaluation. We evaluate top-k query communication costs of VSFTQ, EVTQ, ETQ-RIV, and AD-VQ as n as well as k increases. Figure 3(a) indicates that the query communication costs of the four schemes all increase as n increases. The reason is obvious: the larger n, the more noncontributed sensor nodes, and the more verification information for noncontributed nodes required to be submitted. Figure 3(a) also shows that ETQ-RIV has the lowest query communication cost, followed by VSFTQ, EVTQ, and then AD-VQ. In detail, the query communication cost of ETQ-RIV is about 98.67% lower than AD-VQ, 64.30% lower than EVTQ, and 47.69% lower than VSFTQ. In AD-VQ, to obtain the top-k query result,

the storage node is required to submit a top- $(\eta - 1 + k)$ query result due to the dummy readings, where η is a system parameter denoting the difference between the maximum and minimum encrypted readings within an epoch. That is why AD-VQ has the highest query communication cost. In EVTQ, the storage node is required to send a HMAC as proof information for each unqualified sensor node that has no data item satisfying the query, so its query communication cost is high. In VSFTQ, the storage node needs to send symmetric encryption of data item's order, score, and the time epoch as proof information, which takes more communication cost than ETQ-RIV. However, in ETQ-RIV, only one HMAC is needed for all noncontributed sensor nodes, which greatly reduces the query communication cost.

Figure 3(b) shows that the query communication costs of the four schemes all increase as *k* increases, which is because the larger *k*, the more data items that need to be returned by the storage node. Similar to Figure 3(a), ETQ-RIV still has the lowest query communication cost, which is about 99.31% lower than AD-VQ, 51.70% lower than EVTQ, and 38.44% lower than VSFTQ. The reason is as mentioned before: only one HMAC is required for all noncontributed sensor nodes in ETQ-RIV.

6.3. Redundancy Rate Evaluation. We evaluate top-k query result redundancy rate of VSFTQ, EVTQ, ETQ-RIV, and AD-VQ as k increases. Figure 4 indicates that the redundancy rates of the four schemes all decrease as k increases. Furthermore, AD-VQ has the highest redundancy rate followed by EVQT, VSFTQ, and ETQ-RIV. The redundancy rate of

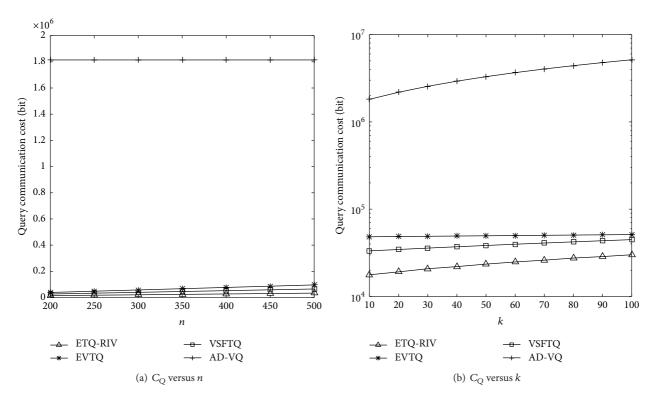


FIGURE 3: Evaluation results of query communication costs.

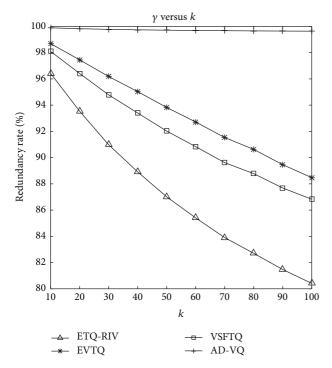


FIGURE 4: Evaluation results of query result redundancy rate.

ETQ-RIV is about 12.68% lower than AD-VQ, 6.77% lower than EVTQ, and 5.19% lower than VSFTQ. The reason is as follows. A top-k query result can be divided into two parts:

the satisfied data items of query result and the verification information. A larger k implies more satisfied data items and a lower redundancy rate. Compared with the other three schemes, only one HMAC is required for all noncontributed sensor nodes in ETQ-RIV, which makes the ETQ-RIV lowest redundancy rate.

According to the above evaluations and analysis, we can conclude that our proposed ETQ-RIV has better performance than the existing works [4, 5, 7, 10–12] both in communication costs and redundancy rate.

7. Conclusions

In this paper, we focus on the problem of verifiable top-k query in two-tiered wireless sensor networks and propose an efficient top-k query processing scheme with result integrity verification which is denoted as ETQ-RIV. To make the query result verifiable, each sensor node should submit some encoded message containing the sequence relationship as proof information for verification along with their collected sensing data items. Evaluation results show that ETQ-RIV can decrease the redundancy rate of query result and thus decrease both in-cell and query communication costs and performs better than the existing works in communication costs.

Conflict of Interests

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

Acknowledgments

This research was supported by the National Natural Science Foundation of China under the Grants nos. 61300240, 61402014, 61472193, 61373137, 61373138, 61272084, and 61201163, the Natural Science Foundation of Jiangsu Province under the Grants nos. BK20151511 and BK20141429, the Project of Natural Science Research of Jiangsu University under Grants nos. 11KJA520002 and 14KJB520027, the Postdoctoral Science Foundation of China under Grant no. 2013M541703, and the Postdoctoral Science Foundation of Jiangsu Province under Grant no. 1301042B.

References

- [1] O. Gnawali, K. Y. Jang, J. Paek et al., "The tenet architecture for tiered sensor networks," in *Proceedings of the 4th ACM Conference on Embedded Networked Sensor Systems*, pp. 153–166, Boulder, Colo, USA, 2006.
- [2] P. Desnoyers, D. Ganesan, and P. Shenoy, "TSAR: a two tier sensor storage architecture using interval skip graphs," in Proceedings of the 5th ACM Conference on Embedded Networked Sensor Systems, pp. 39–50, San Diego, Calif, USA, November 2005.
- [3] Y. Diao, D. Ganesan, G. Mathur, and P. J. Shenoy, "Rethinking data management for storage-centric sensor networks," in *Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research (CIDR '07)*, pp. 22–31, Asilomar, Calif, USA, January 2007.
- [4] H. Dai, G. Yang, F. Xiao, and Q. Zhou, "EVTQ: an efficient verifiable top-k query processing in two-tiered wireless sensor networks," in *Proceedings of the 9th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN '13)*, pp. 206–211, IEEE, Dalian, China, December 2013.
- [5] R. Zhang, J. Shi, Y. Liu et al., "Verifiable fine-grained top-k queries in tiered sensor networks," in *Proceedings of the 29th IEEE International Conference on Computer Communications*, pp. 1199–1207, San Diego, Calif, USA, 2010.
- [6] H. Krawczyk, R. Canetti, and M. Bellare, "HMAC: keyed-hashing for message authentication," Tech. Rep. RFC 2104, Internet Society, Reston, Va, USA, 1997.
- [7] X. Liao and J. Li, "Privacy-preserving and secure top-k query in two-tier wireless sensor network," in *Proceedings of the Global Communications Conference*, pp. 335–341, Anaheim, Calif, USA, 2012.
- [8] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order-preserving encryption for numeric data," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 563–574, Paris, France, June 2004.
- [9] R. Rivest, "The MD5 message-digest algorithm," Tech. Rep. RFC 1321, Internet Society, Reston, Va, USA, 1992.
- [10] X. Ma, H. Song, J. Wang, J. Gao, and G. Min, "A novel verification scheme for fine-grained top-k queries in two-tiered sensor networks," *Wireless Personal Communications*, vol. 75, no. 3, pp. 1809–1826, 2014.
- [11] C. M. Yu, N. K. Guo, Y. Chen et al., "Top-k query result completeness verification in sensor networks," in *Proceeding of* the IEEE Communications Workshops (ICC '13), pp. 1026–1030, Budapest, Hungary, 2013.
- [12] C. M. Yu, N. K. Guo, Y. Chen et al., "Top-k query result completeness verification in tiered sensor networks," *IEEE*

- Transactions on Information Forensics and Security, vol. 9, no. 1, pp. 109–124, 2014.
- [13] Y. Yao, N. Xiong, J. H. Park, L. Ma, and J. Liu, "Privacy-preserving max/min query in two-tiered wireless sensor networks," *Computers & Mathematics with Applications*, vol. 65, no. 9, pp. 1318–1325, 2013.
- [14] J. Cheng, H. Yang, S. H. Wong, P. Zerfos, and S. Lu, "Design and implementation of cross-domain cooperative firewall," in Proceedings of the IEEE International Conference on Network Protocols (ICNP '07), pp. 284–293, Beijing, China, October 2007.
- [15] A. X. Liu and F. Chen, "Collaborative enforcement of firewall policies in virtual private networks," in *Proceedings of the 27th ACM Symposium on Principles of Distributed Computing*, pp. 95–104, ACM, August 2008.
- [16] H. Dai, G. Yang, and X. Qin, "EMQP: an energy-efficient privacy-preserving MAX/MIN query processing in tiered wireless sensor networks," *International Journal of Distributed Sen*sor Networks, vol. 2013, 11 pages, 2013.
- [17] H. Y. Lin and W. G. Tzeng, "An efficient solution to the millionaires' problem based on homomorphic encryption," in Proceedings of the 3rd International Conference on Applied Cryptography and Network Security, pp. 97–134, New York, NY, USA, 2005.
- [18] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in *Proceeding of the 27th IEEE International Conference on Computer Communications*, pp. 46–50, IEEE, Phoenix, Ark, USA, April 2008.
- [19] B. Sheng and Q. Li, "Verifiable privacy-preserving sensor network storage for range query," *IEEE Transactions on Mobile Computing*, vol. 10, no. 9, pp. 1312–1326, 2011.
- [20] J. Shi, R. Zhang, and Y. Zhang, "Secure range queries in tiered sensor network," in *Proceedings of the 28th IEEE International Conference on Computer Communications*, pp. 945–953, Rio de Janeiro, Brazil, 2009.
- [21] J. Shi, R. Zhang, and Y. Zhang, "A spatiotemporal approach for secure range queries in tiered sensor networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 1, pp. 264–273, 2011
- [22] F. Chen and A. X. Liu, "SafeQ: secure and efficient query processing in sensor networks," in *Proceedings of the 29th IEEE International Conference on Computer Communications* (INFOCOM '10), pp. 1–9, IEEE, San Diego, Calif, USA, March 2010.
- [23] Y. Yi, R. Li, F. Chen, A. X. Liu, and Y. Lin, "A digital watermarking approach to secure and precise range query processing in sensor networks," in *Proceedings of the International Conference on Computer Communications (IEEE INFOCOM '13)*, pp. 1950–1958, Turin, Italy, April 2013.
- [24] D. Eastlake and P. Jones, "US secure hash algorithm 1 (SHA1)," RFC 3174, Internet Society, Reston, Va, USA, 2001.
- [25] A. Coman, J. Sander, and M. A. Nascimento, "Adaptive processing of historical spatial range queries in peer-to-peer sensor networks," *Distributed and Parallel Databases*, vol. 22, no. 2-3, pp. 133–163, 2007.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 521289, 20 pages http://dx.doi.org/10.1155/2015/521289

Research Article

A Comprehensive Sensitivity Analysis of a Data Center Network with Server Virtualization for Business Continuity

Tuan Anh Nguyen, 1,2 Dugki Min, 1 and Jong Sou Park 2

Correspondence should be addressed to Tuan Anh Nguyen; anhnt2407@gmail.com

Received 8 April 2015; Revised 21 August 2015; Accepted 30 August 2015

Academic Editor: Supeng Leng

Copyright © 2015 Tuan Anh Nguyen 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

Sensitivity assessment of availability for data center networks (DCNs) is of paramount importance in design and management of cloud computing based businesses. Previous work has presented a performance modeling and analysis of a fat-tree based DCN using queuing theory. In this paper, we present a comprehensive availability modeling and sensitivity analysis of a DCell-based DCN with server virtualization for business continuity using stochastic reward nets (SRN). We use SRN in modeling to capture complex behaviors and dependencies of the system in detail. The models take into account (i) two DCell configurations, respectively, composed of two and three physical hosts in a DCell₀ unit, (ii) failure modes and corresponding recovery behaviors of hosts, switches, and VMs, and VM live migration mechanism within and between DCell₀s, and (iii) dependencies between subsystems (e.g., between a host and VMs and between switches and VMs in the same DCell₀). The constructed SRN models are analyzed in detail with regard to various metrics of interest to investigate system's characteristics. A comprehensive sensitivity analysis of system availability is carried out in consideration of the major impacting parameters in order to observe the system's complicated behaviors and find the bottlenecks of system availability. The analysis results show the availability improvement, capability of fault tolerance, and business continuity of the DCNs complying with DCell network topology. This study provides a basis of designing and management of DCNs for business continuity.

1. Introduction

Cloud computing based businesses have been demanding a rapid escalation of IT infrastructures with efficient resources organization and high level of continuity. To endure business continuity, data centers have drastically evolved in their size and architecture design to host a variety of cloud computing applications and services such as online social networking, e-commerce services, scientific computing, and big data processing. Nevertheless, a data center becomes a centric single point of failure in the cloud infrastructure in the way that the failures of components (e.g., links, switches, and servers) may result in the overall failures of a set of connected components [1]. Internet enterprises may incur millions of dollar per hour [2] due to their service outage,

since their business operations require constantly connected and online services. It is demanding to avoid such risks and improve safety of DCN. And thus designing of a DCN for fault tolerance and business continuity is critical and a sharp focus of both academia and industry areas.

Recent work has attempted to design and organize a data center's resources in the networking manner in which a large number of physical hosts are interconnected in a specific network topology called data center network (DCN), for instance, fat-tree [3], DCell [4], and BCube [5]. Thus, DCN topologies are the communication backbone in a data center [6]. The critical requirements in designing a DCN are scalability and efficiency to connect tens or even hundreds of thousands physical hosts [3, 4]. In the perspective of system users, the metrics of interest for a DCN, however, are

¹Department of Computer, Information & Communications Engineering, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, Seoul 05029, Republic of Korea

²Department of Computer Engineering, Korea Aerospace University, 76 Hanggongdaehang-ro, Deogyang-gu, Goyang-si, Gyeonggi-do 412-791, Republic of Korea

the overall system availability and continuity of their hosted services and applications [7]. In this context, DCell proposed by Guo et al. [4] has emerged as an appropriate solution for DCN architecture in which the architecture is extremely scalable to millions of servers in data centers [8] by recursively constructing higher level DCells based on a DCell₀ as the fundamental building block. The DCell network architecture allows avoiding any single point of failure and thus is able to tolerate different types of failures such as node failures, link failures, and network device failures. Furthermore, to enhance system availability and capability of fault tolerance, one may employ server virtualization [7, 9–11] into a DCN. The approach creates virtual computing machines (VMs) on each physical host of the DCN. Along with the nature of the DCell topology, the VMs become the core elements of the network to deliver high availability and fault tolerance in the way that a VM is able to be migrated from a host to another host [12, 13] and from a DCell₀ to another DCell₀ in the DCN [14] in order to avoid any hardware failures and thus to assure business continuity of system users. The DCellbased DCN with server virtualization is our sharp focus in this paper.

There are a number of papers on presentation and description of DCN topologies [3-5]. Some other work concerned with different aspects of DCN including fault tolerance characteristics [1, 15], structural robustness of DCN topologies [16], or connectivity of DCNs [6]. Nevertheless, none of these papers presented a quantitative assessment of system behaviors using stochastic models [17]. One of the previous works [18] attempted to model and analyze a simple configuration of a two-computer network with redundancy of network devices/links for fault tolerance. To the best of our knowledge, only a recent paper [19] delivered a thorough performance modeling and analysis of a fattree based DCN using queuing theory. Thus we find that modeling and analysis of a virtualized DCN using stochastic models are still a preliminary endeavor. This motivates us to model and analyze a virtualized DCell-based DCN using SRN.

We summarize the main contributions of our work as follows:

- (i) Modeled a DCell-based DCN for business continuity under two configurations, respectively, consisting of two and three virtualized servers in a DCell₀ in a complete manner using SRN.
- (ii) Incorporated failure modes and recovery behaviors of hosts, switches, and VMs along with VM live migration within and between DCell₀s for the sake of fault tolerance.
- (iii) Captured the featured dependencies between components in the system architecture in detail: (a) between hosts and VMs and (b) between switches and VMs.
- (iv) Performed detailed analyses of the constructed SRN models in terms of steady state availability, downtime cost, modeling complexity, and sensitivity with respect to major parameters.

Through modeling and analysis, we have found the following:

- (i) A virtualized DCell-based DCN with greater number of hosts in a DCell₀ can enhance the level of continuity and availability. The DCN with three hosts in a DCell₀ can deliver a level of availability over tier 4 in HA standards for a data center [20].
- (ii) In a virtualized DCN based on DCell network topology, the recovery of hardware (hosts) and software (VMs) subsystems contributes a major impact on system availability. As the size of the DCN increases, the recovery of software subsystem exposes a more important role versus that of hardware subsystem.
- (iii) A bigger size of the VMs in a DCN causes a declining tendency of system availability. Nevertheless, in a more complicated DCN, the influence of VM image size is mitigated.
- (iv) The cross-links between the pairs of hosts in different DCell₀s in a DCN and their bandwidth are necessary elements to tolerate switch failures, to mitigate system downtime, to improve system overall availability, and thus to secure system operation for business continuity.
- (v) The modeling and analysis of the virtualized DCNs in this paper help guide the system design and management of a DCN:
 - (a) A thorough adoption of software fault tolerance is necessary in the design of a DCN.
 - (b) The effectiveness and readiness of the repair and maintenance services in a DCN need more attention and improvement.
 - (c) The trade-off between system availability and performance and the overall cost of networking [21] in a DCN is an important metric in system design.

The rest of this paper is organized as follows. Related work is presented in Section 2. Section 3 introduces a virtualized DCN. Section 4 presents SRN models for the DCNs. The numerical analysis and discussion are presented in Section 5. Finally, Section 6 concludes the paper.

2. Related Work

Design of data center infrastructure is critically demanding in research and development both from academia and industry to deliver cloud-based online apps and services with highest availability. Server-network architecture within a data center thus plays an important role in enhancing agility and reconfigurability of interconnecting different infrastructure resources. In that context, topologies of a DCN contribute major impact on system performance, availability, and scalability to deliver changeable application demands and service requirements [22]. Current DCNs adopt three-layer network topology in which physical servers are interconnected into a rack by a top of rack (ToR) switch, the ToR switches are

networked through end of rack (EoR) switches, and core switches connect these EoR switches together to the external network providers [15]. The topology however confronts a variety of critical issues as a nature: (i) fault-dependency propagation in which a failure of an upper-level switch causes the complete disconnection and unavailability of a number of dependent switches and servers connected to the failed switches and (ii) significant bandwidths that are required to maintain efficient connectivity. To avoid and eliminate the current issues, a number of network topologies have been proposed for alternatives including (i) tree based topologies, such as fat-tree [3, 23, 24] and Clos Network [25] and (ii) recursive-based topologies, such as DCell [4], FiConn [26], BCube [5], and Hyper-BCube [27]. In the work [15], Liu et al. conducted a detailed comparison between the DCNs and concluded that no single topology outperforms the others in all aspects and there will always be trade-offs among cost, performance, and reliability. Among the network topologies, the DCell comes out as a candidate to satisfy the requirements of robustness and connectivity [6], fault tolerance, and scalability in a data center even though the aggregated bottleneck throughput is comparatively low [15]. In this paper, we focus on the DCell-based network topology in consideration of fault tolerance and network availability, which were not considered in the previous work [15]. A DCell [4] is recursively constructed based on the most basic element DCell₀ as follows:

- (i) A DCell₀ consists of *n* physical servers all connected to *n*-port switch.
- (ii) A DCell₁ is composed of n+1 DCell₀s. Each server of a DCell₀ in a DCell₁ has two links. One connects to its switch; the other connects to a corresponding server in another DCell₀, complying with a predetermined DCell routing algorithm. Consequently, every pair of DCell₀s in a DCell₁ has exactly a unique link between each other.
- (iii) A DCell_k is a level-k of DCell_{k-1}.

In this paper, we study two DCell-based DCNs at the level of DCell₁ which, respectively, consist of two and three servers in a DCell₀. We will show through availability as the measure of interest that such DCell-based DCNs expose better ability to tolerate node and switch failures.

High availability (HA) and business continuity (BC) are the key factors in designing enterprise computing systems for a cloud-based business to be successful [28]. Nevertheless, as computing systems with high level of complexity and dependency have been coming out such as Infrastructure-as-a-Service (IaaS), software defined data center (SDDC), and software defined network (SDN), the systems are likely prone to a variety of failures. Severely, a failure of a component may cause a cascading failure or unavailability of a group of other components. For instance, the failure of a switch connecting to a number of physical servers causes the unavailability of the set of servers at the same time. To achieve predetermined levels of HA and BC, which are indicated in service level agreement (SLA) [29, 30] between customers and system owner, the system design has to tolerate any single point

of failure in both hardware and software subsystems. The ANSI/TIA-942 [20] presents four tiers; each specifies basic requirements of availability and downtime minutes per year as follows: (i) tier 1 (basic): 99.671% for availability and 1729.224 downtime minutes in a year; (ii) tier 2 (redundant components): 99.74% for availability and 1361.304 downtime minutes in a year; (iii) tier 3 (concurrently maintainable): 99.982% for availability and 94.608 downtime minutes in a year; and (iv) tier 4 (fault tolerant): 99.995% for availability (four nines) and 26.28 downtime minutes in a year. In order to achieve the above levels of HA standards, one may need to adopt server virtualization [11, 31–34] on nodes and apply VM live migration [12, 35, 36] as the means of fault tolerance for nodes and switches. Server virtualization creates and fosters a plurality of VMs on each physical server. With VM live migration mechanism, a VM is not only able to be migrated from a failed host to another host in the same DCell₀ as soon as a host's failure occurs but also it can be migrated from a running host in a DCell₀ to another host in other DCell₀s if the switch in the former DCell₀ fails. In this paper, we will show a comprehensive modeling and analysis of a virtualized DCell-based DCN for high availability and continuity. The analysis results shown in Section 5 reflect that a DCN complying with DCell network topology along with server virtualization and VM live migration mechanism can achieve HA whereas the DCN with a standalone DCell₀ cannot. Furthermore, the overall system availability of the DCell-based DCN surpasses tier 4 (which is a high available and fault tolerant system) in the ANSI/TIA-942 HA standard for a data center.

Sensitivity analysis [10, 18, 37, 38] is used popularly to provide a selection basis and help design system parameters by observing system characteristics and responses with respect to predetermined valuables in order to identify the most impacting factors as well as to detect bottlenecks in system availability. One may adopt two types of sensitivity analysis: (i) nonparametric sensitivity analysis [39], which studies the system responses upon the component addition/removal or modifications of system model and (ii) parametric sensitivity analysis [40], which observes the system behaviors with respect to the variations of given input parameters. The parametric sensitivity analysis has been adopted to assess system performance and reliability/availability upon the effect of changes of given parameters in different systems. Nguyen et al. [10] presented a comprehensive sensitivity analysis of system steady state availability for a virtualized servers system. The thorough study of availability sensitivity with respect to the intervals of software rejuvenation on VMMs and VMs provides a design basis on how to improve availability of a virtualized system in a wiser manner by combining both software rejuvenations at VM and VMM. In the work [18], Matos Jr. et al. applied parametric sensitivity analysis on a redundant computer network system with respect to MTTF and MTTR of every network component to figure out the important and influent factors of network availability. In another work [38], Matos et al. implemented four different sensitivity analysis techniques to identify the parameters with greatest impact on the availability of a mobile cloud computing system. Accordingly, a sensitivity analysis can be conducted to assess the importance of parameters in the following approaches:

- (i) Repeatedly vary one selected parameter at a time while the others are kept constant and observe the system behaviors on the measures of interest with respect to the varying parameter. This approach studies the system responses upon a broad value range of the parameters in consideration.
- (ii) For differential sensitivity analysis, compute partial derivatives of the measure of interest with respect to each system parameter. This approach is useful in the case that input values of parameters are assigned in a continuous domain. The differential sensitivity of the system availability *A* with respect to variable is defined as in (1) or (2) for a scaled sensitivity:

$$S_{\tau}(A) = \frac{\partial A}{\partial \tau},$$
 (1)

$$SS_{\tau}(A) = \frac{\partial A}{\partial \tau} \left(\frac{\tau}{A}\right).$$
 (2)

- (iii) Calculate the percentage difference in the variation of a parameter from its minimum to maximum values. This technique is designed for integer-valued parameters which are not properly evaluated by the differential sensitivity analysis approach.
- (iv) For design of experiments (DOE) [41], simultaneously examine individual and interactive effects of factors on the output measures.

In this paper, we adopt approaches (i) and (ii) to study system behaviors and responses with respect to parameters at the default values and in a broad value domain. The analyses (i) help find the major impact factors on system availability and (ii) study system characteristics upon the variations of parameters and thus (iv) help design system parameters and (v) guide to adopt proactively different tolerance techniques to achieve optimized overall system availability.

There are a few works on sensitivity modeling and analysis of availability for DCNs. Matos Jr. et al. in the work [18] modeled and analyzed the availability of a small-scale computer network using continuous time Markov chain (CTMC). This very first work studied the impact of the failures of network devices (switches and routers) and network links on system availability. The study took into account the redundancy of either network devices or links as a measure to tolerate the aforementioned failures and to improve system availability. The contributions of this work suggest the approach of adopting stochastic models to analyze a DCN. However, as the system scale increases, the CTMC models (where each state in the model is the combination of all states of the components in consideration) likely confront largeness problem or state-space explosion as well as intractable presentation of the model. Furthermore, unplanned redundancy of physical devices is a costly solution especially for DCNs as the number of machines increases vastly. It is needed to

organize the physical components in a well-designed network topology for either fault tolerance or high availability and/or performance. Alshahrani and Peyravi in a very first work [19] on modeling and analysis of DCNs attempted to adopt queuing theory to model a typical topology of DCN, fattree [3]. This work proposed a detailed analytical model to assess performance metrics of interest (e.g., throughput and delay) of a fat-tree based DCN. The work nevertheless did not take into consideration any type of failures. The system architecture is composed of only network devices for simplification of theoritical formulation. This preliminary work raises a need to conduct further studies on various attributes (e.g., availability, reliability, and performability [42]) of dependability of different DCN topologies. Several other works studied various essential issues of contemparory DCNs. Liu et al. [1, 15] studied fault tolerance characteristics of renowned network topologies of DCNs. Among different topologies of DCNs presented in the works [3-5, 23, 25, 27, 43], DCell topology is pointed out as a typical topology with high scalability and fault tolerance capability [1, 4, 15] with more and more interest in practice. Several works [44-46] presented large-scale emperical studies of failures in typical data centers. The works have charaterized a variety of failures in DCNs such as failures of servers (e.g., hard disk, memory and raid controller failures) and failures of network devices (top-of-rack switch, aggregation switch, and router failures). Some other works [7, 9, 11, 47-50] showed the adoption of renowned virtualization technology and VM migration in computing systems is of paramount importance to achieve high availability and to tolerate unexpected risks or failures.

Based on the above literature review, we find that the modeling and analysis of a DCN are still in initial steps. Previous work attempted to model and analyze DCNs using either stochastic models or queuing theory without an adequate consideration of different system failures and corresponding fault tolerant techniques. Moreover, the system architectures did not incorporate contemporary virtualization technology and VM migration techniques for high availability and effecient fault tolerance in virtualized environment. Our focus in this paper is on the ability of DCell-based DCNs to tolerate any type of risks in the system in order to ensure the system's safety in terms of system operation and availability. We are more interested in risks and measures to tolerate risks for DCNs to achieve high availability (which are critically demanding in design of network in data centers) rather than other attibutes of DCN's dependability. Therefore we choose to study DCell network topology for DCNs which allows us to enhance the system capability of fault tolerance. We attempted to use SRN with a variety of modeling funtionalities in order to capture the dependency between upper and lower level components in the system architecture (for instance, between a physical server and its hosted VMs or between a switch and its connected servers). Furthermore, the SRN models are tractable (literally compared to CTMC) and thus enable us to incorporate various behaviors (failures, VM migration, and interaction between submodels). We will be using SRN to model typical DCell-based DCNs in the next sections.

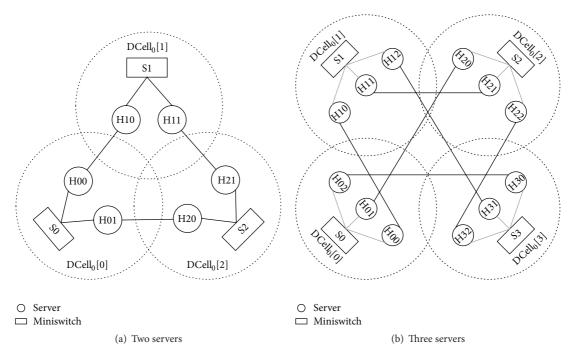


FIGURE 1: DCN system architectures.

3. A Virtualized Data Center Network

3.1. System Architectures. The system architectures of a DCell-based DCN are depicted in Figure 1. Figure 1(a) shows the architecture of a DCell-based DCN with two servers in a DCell₀ (called DCN2 from now on) whereas Figure 1(b) depicts the architecture of the DCell-based DCN with three servers in a DCell₀ (called DCN3). Both DCN2 and DCN3 comply with DCell configuration and routing [4]. Accordingly, DCN2 consists of three DCell₀s: DCell₀[0], DCell₀[1], and DCell₀[2]. Each of DCell₀s comprises two servers (also called host) and a gigabit switch to help clients use the servers. All hosts are virtualized to run a number of virtual machines (VMs). Within a DCell₀, gigabit links for high speed data transactions connect the hosts and the corresponding switch. For easy understanding of the SRN models of the system (to be presented in the next sections), we apply a naming convention in which characters represented for a component accompany specific numbers. In particular, the DCell₀[0] consists of the switch S0, the hosts H00 and H01, and their respective virtual machines VM00 and VM01. The naming convention is applied in the same way for $DCell_0[1]$ which comprises S1, {H10, H11}, and {VM10, VM11} and for DCell₀[2] which is also composed of S2, {H20, H21}, and {VM20, VM21}. In the same way of the above descriptions, DCN3 is composed of four DCell₀s from DCell₀[0] to DCell₀[3]. Each DCell₀ in DCN3 consists of one switch and three hosts and each host in turn runs a number of VMs. The notation of components in DCN3 complies with the naming convention aforementioned in DCN2. The network routing is formed between hosts among different DCell₀s. Particularly, internal network links in DCN2 are formed between the following pairs of hosts: {H00, H10}; {H01, H20};

and {H11, H21}. Whereas, in DCN3 the links are formed between the pairs: {H00, H10}; {H01, H20}; {H02, H30}; {H11, H21}; {H12, H31}; and {H22, H32}, which comply with DCell routing tactics [4]. We will use these architectures to model and analyze the system availability in the next sections.

3.2. System Behaviors and Assumptions

(i) Operational States. A host and a switch can fail and recover upon the states of their hardware components as similar as the blade server described in [51]. And a VM may go through a variety of complicated states as in [10, 44, 45]. But capturing the different operational states of hosts, VMs, and switches in modeling in a complete manner is beyond our focus and could lead to largeness problem [52] of the models. Hence, the two-states model (up and down states) is used to represent the basic operational states of system subsystems.

(ii) VM Live Migration. The VM live migration technique is employed to tolerate unexpected failures of hosts and switches. In a $\mathrm{DCell_0}$, if a host fails, all VMs running on the failed host are immediately migrated onto the remaining hosts with good consideration of load balancing. Moreover, if a switch fails, all VMs operating on the hosts in the $\mathrm{DCell_0}$ of that failed switch are instantly migrated to the other hosts of all other remaining $\mathrm{DCell_0}$ s. For instance, in the case of the DCN2, if the host H00 fails, the VMs running on H00 are live-migrated onto the host H01. When the switch S0 goes down, the live migration processes are triggered instantly to migrate the VMs running on H00 and H01, respectively, onto the host H10 of the $\mathrm{DCell_0}$ [1] and onto the host H20 of the $\mathrm{DCell_0}$ [2]. The descriptions can be applied in the same way for other $\mathrm{DCell_0}$ s in $\mathrm{DCN2}$ and $\mathrm{DCN3}$. The above VM live

migration mechanisms are used to prevent the VMs from unexpected downtime due to failures of hosts and switches; thus system availability is improved and business continuity is endured. In order to reduce the complexity of system models, it is necessary to assume that the VM live migration processes do not confront any unexpected failures such as data loss and memory errors, during the migration period as captured in some work [35, 53].

(iii) Virtual Machine Monitor (VMM). Hypervisor or VMM is in charge of creating and maintaining virtualization environment to operate the upper VMs. Thus, the operational states of VMs are dependent on the operations of the underlying VMM. The detailed dependencies of a host, a VMM, and a VM are captured in a number of works [10, 44, 45]. Nevertheless, we do not take into account the VMM in modeling for simplification and our focus is on the states of VMs, since user's applications and services run on VMs. We consider the up and down states of VMM as one part of the host's up and down states.

(iv) High Availability. Our systems are designed to deliver HA services to the user. In HA system, the availability is assigned to the cases that more components are in up states. Thus, we assume that repair and maintenance services in data centers are good enough to recover the failed hardware components in advance of the remaining component failures. Particularly in DCN2, we assume the following:

- (i) If a host fails in a DCell₀, the remaining host only fails after the recovery of the aforementioned host.
- (ii) If the number of failed hosts in the DCN2 is larger or equal to the haft of the total number of hosts, the remaining hosts can fail after one of the failed hosts is recovered.
- (iii) If two switches fail, the repair person is summoned and repairs subsequently the failed switch before the failure of the remaining switch.
- (iv) A switch and a host are in operation and can fail if there is at least a VM running in the host of the same DCell₀.

In the case of DCN3, we assume the following:

- (i) If two hosts fail in a DCell₀, the remaining can fail as either one or both of the failed hosts are recovered.
- (ii) If the number of failed hosts in DCN3 is larger or equal to the haft of the total number of hosts, the remaining hosts can fail as one of the failed hosts is recovered.
- (iii) If three out of four switches fail, the remaining switch can fail after the recovery of one of the failed switches.
- (iv) A switch and a host are in operation and can fail if there is at least a VM running in the host of the same DCell₀. The purposes of these assumptions are to reduce the largeness of the model and to mitigate the cases with very low probability to occur in HA system.

(v) Distributions. The time to occurrence of any event in actual computing system may follow different types of probability distribution [54]. However, we can make appropriate distribution assumptions for every transition so that the analytical system model is closer to the practical system. In this paper, we choose to use exponential distribution for simplification in modeling and analysis as a common option in a large number of papers [10, 44, 45].

The above system behaviors and assumptions are all taken into consideration in the modeling of the DCNs to be carried out and described in detail in Section 4.

4. Stochastic Reward Net Models

4.1. System Models. The SRN system models of DCN2 and DCN3 are, respectively, depicted in Figures 2 and 3. The system models are composed of partial models including host models, switch models, and VM subsystem model as chronologically named from Figure 2(a to g) in SRN system model for DCN2 and from Figure 3(a to q) in SRN system model for DCN3. In consideration of system availability measures, we use two-states model (up and down states) to model hosts, switches, and VMs for the sake of modeling simplification. Our sharp focus in modeling is on the dependency between components and fault tolerant behaviors for high availability in the case of any component's failure. We will describe the model of a specific component as an example to refer to the other similar components. Then the model integration and dependency will be presented subsequently. The transitions of tokens within the models are conducted stochastically by enabling/disabling the timed/immediate transitions based on the predefined behaviors. The combinations of all tokens' locations in the models represent the system's respective states. We list down all the states of every component and possible locations of tokens in the models as in Table 1. In order to capture exact predefined system behaviors we apply a set of guard functions [55–58] attached to every transitions in the models to control the transitions of tokens.

4.2. SRN Models of Hosts, VMs, and Switches. Figure 4 shows the two-state SRN models of selected host, VM, and switch. In the assumption, we mentioned that the characteristics and configurations of all hosts, VMs, and switches are assumed to be identical initially. Thus, we can use the two-state SRN models to capture up and down states of the component in regard of availability measures. We describe the modeling of the host H00, the VM00, and the switch S0 as the examples to refer to the modeling of the other identical hosts, VMs, and switches in both SRN system models of DCN2 and DCN3.

Figure 4(a) depicts the modeling of a host with repair actions. Initially, a host is considered in running state depicted by a token in up state $P_{\rm H00up}$. A virtualized host in DCN may undergo an expected failure or maintenance period after a specific time with MTTF $1/\lambda_{\rm H}$. In this case, the transition $T_{\rm H00f}$ is triggered to fire and the token in $P_{\rm H00up}$ is removed and deposited in $P_{\rm H00dn}$. As the host goes down, a repair person is summoned to recover the host. After the repair, the transition $T_{\rm H00r}$ is enabled and the token in

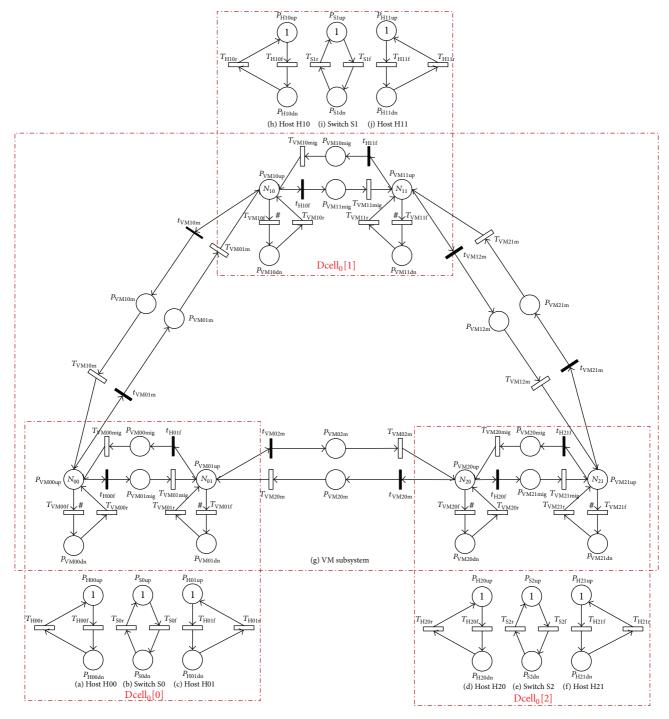


FIGURE 2: SRN model of a DCN with two servers in a DCell₀.

 $P_{
m H00dn}$ is removed and deposited in $P_{
m H00up}$. The host returns operational.

Figure 4(b) captures the behaviors of VMs running on host H00. Assume that there are initially N_{00} in running states. As time goes by, a VM can fail with a failure rate $\lambda_{\rm VM}$. The transition is fired subsequently and one token in $P_{\rm VM00up}$ is removed and deposited in $P_{\rm VM00dn}$. The VM goes down. Because of the competition between the VMs in up state to fail, the failure rate of the running VMs at a time depends on

the number of VMs or, in other word, the number of tokens in the place $P_{\rm VM00up}.$ Therefore, we apply marking dependence on the transition $T_{\rm VM00f}$ represented by the marker "#." The VMs in downstate are repaired in sequence by software or by a repair person. The repaired VM restarts to healthy state. This repair action is captured by firing the transition $T_{\rm VM00r};$ then a token in $P_{\rm VM00dn}$ is taken out and deposited in $P_{\rm VM00up}.$

Figure 4(c) presents the failure and repair action of a switch in modeling. At the beginning the switch is considered

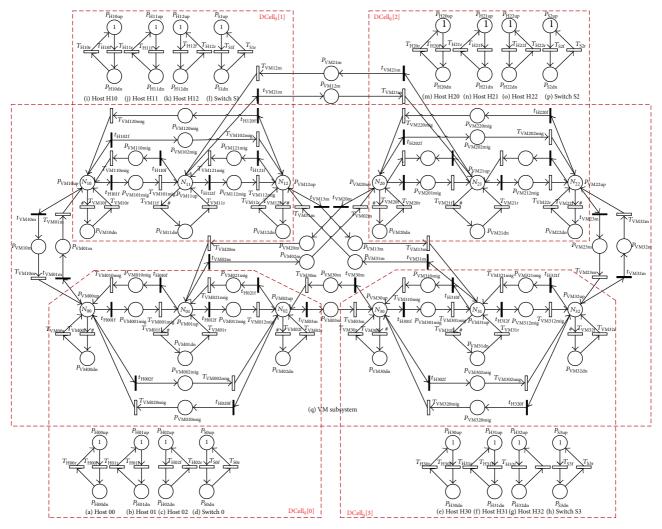


FIGURE 3: SRN model of a DCN with three servers in a DCell₀.

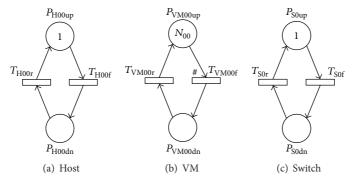


FIGURE 4: SRN models of hosts, VMs, and switches.

in healthy state depicted by a token in $P_{\rm S0up}$. After some time, it may fail, the transition $T_{\rm S0f}$ is triggered to fire, and the token in $P_{\rm S0up}$ is taken out and deposited in $P_{\rm S0dn}$. The switch fails consequently. After repairing the failed switch, the transition $T_{\rm S0r}$ is enabled and the token in $P_{\rm S0dn}$ is removed and deposited in $P_{\rm S0up}$. The switch starts running normally.

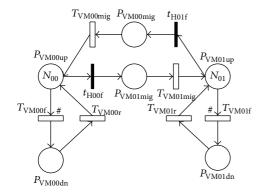
4.3. SRN Models of a Standalone DCell₀. Figures 5 and 6, respectively, depict the SRN models of the DCell₀s comprising two and three hosts (hereinafter called DCN0 and DCN1), which are the basic units to construct the DCellbased DCN2 and DCN3. The DCN0 and DCN1 are actually the DCell₀[0] taken out for an example of modeling

Table 1: Description of places in SRN system models of DCN2 and DCN3.

Place	Description
$\begin{array}{l} P_{\rm H00up}, P_{\rm H01up}, P_{\rm H02up}, P_{\rm H10up}, P_{\rm H11up}, P_{\rm H12up}, P_{\rm H20up}, \\ P_{\rm H21up}, P_{\rm H22up}, P_{\rm H30up}, P_{\rm H31up}, \text{and } P_{\rm H32up} \end{array}$	Running states of respective hosts: H00, H01, H02, H10, H11, H12, H20, H21, H22, H30, H31, and H32.
$\begin{array}{l} P_{\rm H00dn}, P_{\rm H01dn}, P_{\rm H02dn}, P_{\rm H10dn}, P_{\rm H11dn}, P_{\rm H12dn}, P_{\rm H20dn}, \\ P_{\rm H21dn}, P_{\rm H22dn}, P_{\rm H30dn}, P_{\rm H31dn}, \text{ and } P_{\rm H32dn} \end{array}$	Down states of respective hosts: H00, H01, H02, H10, H11, H12, H20, H21, H22, H30, H31, and H32.
$P_{\text{S0up}}, P_{\text{S1up}}, P_{\text{S2up}}, \text{ and } P_{\text{S3up}}$	Running states of respective switches: S0, S1, S2, and S3.
P_{S0dn} , P_{S1dn} , P_{S2dn} , and P_{S3dn}	Down states of respective switches: S0, S1, S2, and S3.
$\begin{aligned} &P_{\text{VM00up}}, P_{\text{VM01up}}, P_{\text{VM02up}}, P_{\text{VM10up}}, P_{\text{VM11up}}, P_{\text{VM12up}}, \\ &P_{\text{VM20up}}, P_{\text{VM21up}}, P_{\text{VM22up}}, P_{\text{VM30up}}, P_{\text{VM31up}}, \text{ and } P_{\text{VM32up}} \end{aligned}$	Running states of respective VMs: VM00, VM01, VM02, VM10, VM11, VM12, VM20, VM21, VM22, VM30, VM31, and VM32.
$\begin{array}{l} P_{\text{VM00dn}}, P_{\text{VM01dn}}, P_{\text{VM02dn}}, P_{\text{VM10dn}}, P_{\text{VM11dn}}, P_{\text{VM12dn}}, \\ P_{\text{VM20dn}}, P_{\text{VM21dn}}, P_{\text{VM22dn}}, P_{\text{VM30dn}}, P_{\text{VM31dn}}, \text{ and } P_{\text{VM32dn}} \end{array}$	Down states of respective VMs: VM00, VM01, VM02, VM10, VM11, VM12, VM20, VM21, VM22, VM30, VM31, and VM32.
$P_{\rm VM00mig}$ and $P_{\rm VM01mig}; P_{\rm VM10mig}$ and $P_{\rm VM11mig}; P_{\rm VM20mig}$ and $P_{\rm VM21mig}$	Intermediate states of VM migration processes in DCell ₀ [0], DCell ₀ [1], and DCell ₀ [2] in DCN2, respectively, from H01 to H00 and from H00 to H01, from H11 to H10 and from H10 to H11, and from H21 to H20 and from H20 to H21.
$P_{ m VM001mig}$ and $P_{ m VM010mig}$; $P_{ m VM012mig}$ and $P_{ m VM021mig}$; $P_{ m VM002mig}$ and $P_{ m VM002mig}$	Intermediates states of VM migration processes in $DCell_0[0]$ in DCN3, respectively, from H00 to H01 and from H10 to H00, from H01 to H02 and from H02 to H01, and from H00 to H02 and from H02 to H00
$P_{ m VM101mig}$ and $P_{ m VM110mig}$; $P_{ m VM112mig}$ and $P_{ m VM121mig}$; $P_{ m VM102mig}$ and $P_{ m VM120mig}$	Intermediate states of VM migration processes in $DCell_0[1]$ in DCN3, respectively, from H10 to H11 and from H11 to H10, from H11 to H12 and from H12 to H11, and from H10 to H12 and from H12 to H10.
$P_{\rm VM201mig}$ and $P_{\rm VM210mig}; P_{\rm VM212mig}$ and $P_{\rm VM221mig};$ $P_{\rm VM202mig}$ and $P_{\rm VM220mig}$	Intermediate states of VM migration processes in $DCell_0[2]$ in DCN3, respectively, from H20 to H21 and from H21 to H20, from H21 to H22 and from H22 to H21, and from H20 to H22 and from H22 to H20.
$P_{ m VM301mig}$ and $P_{ m VM310mig}$; $P_{ m VM312mig}$ and $P_{ m VM321mig}$; $P_{ m VM302mig}$ and $P_{ m VM320mig}$	Intermediate states of VM migration processes in $DCell_0[3]$ in DCN3, respectively, from H30 to H31 and from H31 to H30, from H31 to H32 and from H32 to H31, and from H30 to H32 and from H32 to H30.
$P_{ m VM01m}$ and $P_{ m VM10m}$	Intermediate states of VM migration processes between $\mathrm{DCell}_0[0]$ and $\mathrm{DCell}_0[1]$ in DCN2 and DCN3, respectively, from H00 to H10 and from H10 to H00.
$P_{ m VM12m}$ and $P_{ m VM21m}$	Intermediate states of VM migration processes between $DCell_0[1]$ and $DCell_0[2]$ in DCN2 and DCN3, respectively, from H11 to H21 and from H21 to H11.
$P_{ m VM20m}$ and $P_{ m VM02m}$	Intermediate states of VM migration processes between $DCell_0[2]$ and $DCell_0[0]$ in DCN2 and DCN3, respectively, from H20 to H01 and from H01 to H20.
$P_{ m VM03m}$ and $P_{ m VM30m}$	Intermediate states of VM migration processes between $DCell_0[0]$ and $DCell_0[3]$ in DCN3, respectively, from H02 to H30 and from H30 to H02.
P_{VM13m} and P_{VM31m}	Intermediate states of VM migration processes between $\mathrm{DCell}_0[1]$ and $\mathrm{DCell}_0[3]$ in DCN3, respectively, from H12 to H31 and from H31 to H12.
$P_{ m VM23m}$ and $P_{ m VM32m}$	Intermediate states of VM migration processes between $DCell_0[2]$ and $DCell_0[3]$ in DCN3, respectively, from H22 to H32 and from H32 to H22.

description, respectively, in the DCN2 and DCN3. The SRN model of DCN0 in Figure 5 consists of host models of the hosts H00 (Figure 5(a)) and H01 (Figure 5(c)), switch model of the switch S0 (Figure 5(b)), and VM models of the VM00 and VM01 (Figure 5(d)). The modeling of these partial components can be referred to the description of

the corresponding models in Figure 4. Henceforth we describe the dependency of the VM model upon the host and switch models. In particular, we apply VM live migration as a fault tolerant technique to avoid the downtime of VMs because of their host's failures. Initially, all components are in up state depicted by the tokens in $P_{\rm H00up}$, $P_{\rm H01up}$, $P_{\rm S0up}$,



(d) VM subsystem

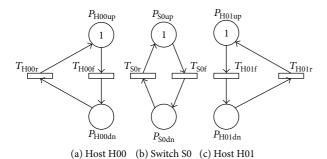


FIGURE 5: SRN models of a DCell₀ with two servers.

 P_{VM00up} , and P_{VM01up} . At a certain time, the host H00 may fail, which is represented by a token in $P_{\rm H00dn}$. The failure and repair transitions $T_{\rm VM00f}$ and $T_{\rm VM00r}$ are disabled. All the VMs running on the host H00 consequently are triggered to undergo a live migration process. This behavior is captured by enabling the immediate transition $t_{\rm H00f}$. The tokens in $P_{\rm VM00up}$ are taken out and deposited in the intermediate place P_{VM01mig} . At this point, the VM migration processes start by enabling the transition $T_{\rm VM01mig}$. A token in the place P_{VM01mig} is removed and deposited in the place P_{VM01up} one after another. Thus, the VMs in the failed host H00 are all live-migrated onto the operational host H01 in the same DCell₀. In the case that the host H01 fails in the progress of VM migration, the migration processes are interrupted and halted until the failed host H01 is recovered completely. The VMs' image files and related data are stored in the DCell₀'s storage, represented by the tokens in the intermediate place $P_{
m VM01mig}$. Hence, the transition $T_{
m VM01mig}$ is disabled until the completion of the failed host H01's recovery processes. Based on the above description, we can refer to the case of the host H01's failure. As soon as the host H01 fails (represented by a token in $P_{\rm H01dn}$) but the host H00 still runs (one token in $P_{\rm H00up}$), the VMs running on the host H01 are live-migrated onto the host H00 with the same aforementioned processes. The immediate transition $t_{\rm H01f}$ is triggered to fire. All the tokens in P_{VM01up} are taken out and deposited in P_{VM10mig} . The migration process of VMs is carried out in sequence as long as the host H00 is operational. If the host H00 fails during the migration of VMs from the failed host H01, the transition $T_{\rm VM10mig}$ is disabled and the VM migration

processes stop until the host H00 is recovered. Furthermore, if both hosts H00 and H01 go down along with each other, the running VMs' image file and related data are stored on a shared memory, which is captured by a number of VMs in the places P_{VM01mig} and P_{VM10mig} previously taken out from the respective places P_{VM00up} and P_{VM01up} . Then all the transitions in the VM subsystem model (Figure 5(d)) are disabled to stop completely the VMs' operations. In addition, if the switch S0 fails (a token resides in P_{S0dn}), the running VMs on the hosts H00 and H01 are live-migrated to the respective hosts in the other DCell₀s upon the network routing presented in the system architecture in Figure 1(a). This behavior is presented in detail in the next section. Modeling description of the DCN1 model in Figure 6 is carried out in detail as the above description of the DCN0 model in Figure 5 with good consideration on the notation alteration. The DCN1 model consists of the host models of the hosts H00, H01, and H02; the switch model of the switch S0 and the VM models of the VM00, VM01, and VM02, respectively, hosted on the aforementioned hosts. The dependency and behaviors of VM subsystem upon the operational states of the hosts and switch are similar as described in the DCN0 model. The VM live migration processes are conducted between the two among three hosts. If a host fails, the running VMs on the failed host are live-migrated to the two remaining hosts in consideration of balancing the number of VMs on each host. If a switch goes down, the running VMs on each host are live-migrated to the corresponding hosts in the other DCell₀s through the cross-links between DCell₀s according to the network routing showed in the DCN3 system architecture in Figure 1(b).

4.4. System Model Integration. The models of DCN2 in Figure 2 and of DCN3 in Figure 3 are made of, respectively, three DCN0s and four DCN1s complying with the DCellbased network routing topologies as in the system architecture in Figure 1. The modeling descriptions of every component and DCell₀ units are carried out based on the detailed descriptions of partial component models in Figures 4, 5, and 6. In this section, we show the features of DCell-based DCN upon system model integration. In consideration of a standalone DCell₀, if its switch undergoes a downtime period because of unexpected failure or planned maintenance, the communication between computing machines in the DCell₀ and system users is disconnected as a result. To avoid this adverse situation, in DCN2 and DCN3 the computing VMs are live-migrated to other DCell₀s through the cross-links between hosts from different DCell₀. In particular, in the DCN2, the $DCell_0[0]$ connects to the $DCell_0[1]$ via the link between the hosts H00 and H10 and to the DCell₀[2] via the link between the hosts H00 and H20. In turn, the DCell₀[1] connects to the DCell₀[2] via the link between the hosts H11 and H21. In the DCN3, the above description goes in similar way in which a DCell₀ connects to three remaining DCell₀s via different links between the pairs of specific hosts. We take the failure of the switch S0 in the DCell₀[0]-DCN2 as an example to describe the system behaviors and interactions between DCell₀s upon the failure of switches. As the switch S0 fails depicted by a token in the place P_{S0dn} in Figure 2, all

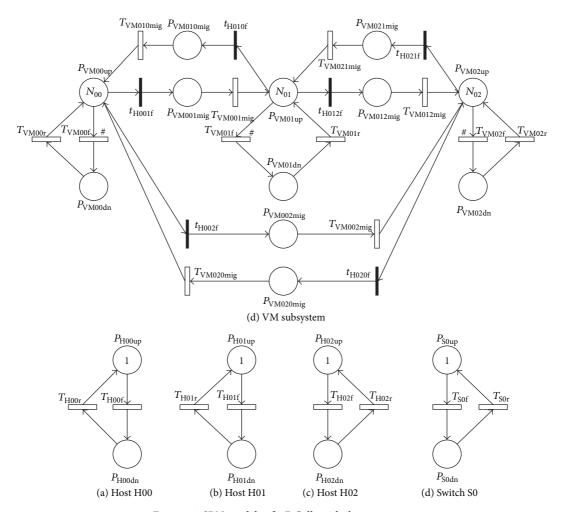


Figure 6: SRN models of a $DCell_0$ with three servers.

the running VMs on the hosts H00 and H01 (represented by the tokens residing in the places P_{VM00up} and P_{VM01up}) are live-migrated, respectively, to the hosts $H\dot{1}0$ in $DCell_0[\dot{1}]$ and H20 in DCell₀[2]. To capture these behaviors in modeling, the immediate transitions t_{VM01m} and t_{VM02m} are triggered to fire as soon as the token in $P_{\rm S0up}$ is removed and deposited in P_{S0dn} . Subsequently, the tokens in the places P_{VM00up} and P_{VM01up} are taken out and deposited, respectively, in the places P_{VM01m} and P_{VM02m} . At this point, the VMs' image files and related data stored in the local memory system of the $DCell_0[0]$ are organized to be migrated from $DCell_0[0]$ to $DCell_0[1]$ and $DCell_0[2]$. The timed transitions T_{VM01m} and $T_{\rm VM02m}$ are enabled to start migration processes. After the completion of the VM migration processes, the VMs hosted on the host H00 in DCell₀[0] with the failure of the switch S0 now operate on the host H10 in DCell₀[1] and the VMs hosted on the host H01 in DCell₀[0] with the failure of the switch S0 now run on the host H20 inDCell₀[2]. Under the same reasoning, we can describe the live migration mechanism of VMs from DCell₀[1] to DCell₀[0] and DCell₀[2] upon the failure of the switch S1 and also from DCell₀[2] to DCell₀[0] and DCell₀[1] upon the failure of the switch S2. In the DCN3 system model, as soon as the switch S0 in DCell₀[0] fails,

the tokens running on the hosts H00, H01, and H02 (depicted by the tokens in the places $P_{\rm VM00up}, P_{\rm VM01up}, P_{\rm VM02up})$ are, respectively, migrated to the hosts H10 in $\rm DCell_0[1]$, H20 in $\rm DCell_0[2]$, and H30 in $\rm DCell_0[3]$ (captured by, resp., depositing the tokens in $P_{\rm VM10up}, P_{\rm VM20up}$, and $P_{\rm VM30up}$). Based on the above detailed description, the migration of VMs from the other $\rm DCell_0s$ upon switch failures can be conducted accordingly.

5. Numerical Results

The SRN models of the DCNs are implemented in Stochastic Petri Net Package (SPNP) [57]. SPNP provides two ways to implement the SRN models: (i) raw input language for SPNP called CSPL (C-based SPN Language), an extension of the C programming language with a variety of application programming interfaces (API) for easier description of SRN models; and (ii) a Graphical User Interface (GUI) for intuitive specification of the SRN models, which later is converted into CSPL automatically by the software itself. The models are converted into Markov Reward Model (MRM) and then solved by using analytic-numeric methods with regard to specific metrics of interest. We use GUI to construct and

verify the correctness of the SRN models and CSPL input source to solve the models and generate various numerical analysis results as well as to investigate the complexity of those analyses. Our metrics of interest for analyses include (i) steady state availability (SSA), (ii) downtime cost, and (iii) sensitivity of SSA with respect to major impacting parameters. Default values of parameters used in modeling are provided in Table 2 based on previous works [10, 50, 58, 59].

To investigate the capability of the DCNs to assure business continuity, we initiate one VM to run on each host of the $DCell_0[0]$ in either DCN2 or DCN3 at the beginning; and none of VMs is initialized on all the other hosts. In general, the DCell-based DCNs can maintain business operations on the aforementioned VMs even in the case of switch failures by migrating the VMs onto the other hosts of all the other $DCell_0s$. Thus the overall system availability

is improved apparently. These features of the DCell-based DCNs are shown by the numerical analysis results in the next subsections.

5.1. Steady State Analysis. The steady state analyses are carried out along with the downtime and cost analyses for four case studies from (I) to (IV) as in Tables 3 and 4. In order to compute the measures of interest using SPNP, we define the requirements of our systems' availability as follows: (i) there is at least a VM running in a certain DCell₀ and (ii) the switch in the DCell₀ stays in operational state. The requirements are to ensure that there is at least a connection between system users and running computing units. Based on the predetermined requirements, we define the reward functions to compute the system availability for the four DCNs as follows:

$$A_{\text{DCN0}} = \begin{cases} 1, & \text{if } \left(\#P_{\text{VM00up}} + \#P_{\text{VM01up}} > 0 \right) \&\& \left(\#P_{\text{S0up}} == 1 \right), \\ 0, & \text{otherwise,} \end{cases}$$

$$A_{\text{DCN1}} = \begin{cases} 1, & \text{if } \left(\#P_{\text{VM00up}} + \#P_{\text{VM01up}} + \#P_{\text{VM02up}} > 0 \right) \&\& \left(\#P_{\text{S0up}} == 1 \right), \\ 0, & \text{otherwise,} \end{cases}$$

$$A_{\text{DCN2}} = \begin{cases} 1, & \text{if } \left\{ \left(\#P_{\text{VM00up}} + \#P_{\text{VM01up}} > 0 \right) \&\& \left(\#P_{\text{S0up}} == 1 \right) \right\} \\ & \parallel \left\{ \left(\#P_{\text{VM10up}} + \#P_{\text{VM11up}} > 0 \right) \&\& \left(\#P_{\text{S1up}} == 1 \right) \right\}, \\ 0, & \text{otherwise,} \end{cases}$$

$$A_{\text{DCN3}} = \begin{cases} 1, & \text{if } \left\{ \left(\#P_{\text{VM00up}} + \#P_{\text{VM01up}} + \#P_{\text{VM02up}} > 0 \right) \&\& \left(\#P_{\text{S0up}} == 1 \right) \right\}, \\ & \parallel \left\{ \left(\#P_{\text{VM10up}} + \#P_{\text{VM11up}} + \#P_{\text{VM02up}} > 0 \right) \&\& \left(\#P_{\text{S1up}} == 1 \right) \right\}, \\ & \parallel \left\{ \left(\#P_{\text{VM20up}} + \#P_{\text{VM11up}} + \#P_{\text{VM12up}} > 0 \right) \&\& \left(\#P_{\text{S2up}} == 1 \right) \right\}, \\ & \parallel \left\{ \left(\#P_{\text{VM30up}} + \#P_{\text{VM31up}} + \#P_{\text{VM32up}} > 0 \right) \&\& \left(\#P_{\text{S3up}} == 1 \right) \right\}, \\ 0, & \text{otherwise.} \end{cases}$$

The numerical results of the steady state analyses and downtime cost analyses with default parameters are shown in Table 5. We assume that a minute of system downtime incurs a penalty of 16,000 USD on the system owner according the SLA signed with customers [60]. The number of nines (a correspondence to availability, nines = $-\log(1 - A)$ [58]) is used to present the improvement and change of steady state availability in an intuitive way. The results show that the adoption of DCell-based architectures improves significantly the system availability and thus decreases vastly the downtime and the corresponding downtime cost. Particularly, the DCN of a DCell0 with two hosts (DCN0 in Figure 5) has the state availability at correspondingly about 2.55 of nines; thus the downtime in a year is at a huge number of 1450.4 minutes and the system owner must bear 23,206,942 USD per year

for this system's performance. If we adopt the DCell-based architecture in Figure 3 (DCN3), the system's steady state availability improves vastly with the corresponding number of nines at about 5.19 (almost double compared to DCN0), the system downtime drops off at about 3.4 minutes in a year, and thus the incurred cost now is only 53,959 USD per year. This analysis results reflect the efficiency of the DCell-based DCN in terms of fault tolerance to achieve high availability and mitigate system downtime in comparison with the normal DCN without the adoption of DCell topology. In comparison, the DCNs with three hosts in a DCell₀ (e.g., DCN3 and DCN1) gain relatively higher availability than the respective DCNs with two hosts in a DCell₀ (e.g., DCN2 and DCN0). This is to say that an increase of number of VMs can benefit the system owner to provide higher availability to customers.

Table 2: Parameter default values used in the analyses.

Parameters	Description	Assigned transitions	Mean time/values
$\lambda_{ m H}$	Host failure rate	$T_{\rm H00f}, T_{\rm H01f}, T_{\rm H02f}, T_{\rm H10f}, T_{\rm H11f}, T_{\rm H12f}, T_{\rm H20f}, T_{\rm H21f}, T_{\rm H22f}, \\ T_{\rm H30f}, T_{\rm H31f}, \text{ and } T_{\rm H32f}$	800 hours
$\mu_{ m H}$	Host repair rate	$T_{ m H00r}, T_{ m H01r}, T_{ m H02r}, T_{ m H10r}, T_{ m H11r}, T_{ m H12r}, T_{ m H20r}, T_{ m H21r}, T_{ m H22r}, T_{ m H22r}, T_{ m H31r}, T_{ m H32r}$	9.8 hours
$\lambda_{ m VM}$	VM failure rate	$T_{\text{VM00f}}, T_{\text{VM01f}}, T_{\text{VM02f}}, T_{\text{VM10f}}, T_{\text{VM11f}}, T_{\text{VM12f}}, T_{\text{VM20f}}, T_{\text{VM21f}}, T_{\text{VM22f}}, T_{\text{VM30f}}, T_{\text{VM31f}}, \text{and } T_{\text{VM32f}}$	4 months
$\mu_{ m VM}$	VM repair rate	$T_{ m VM00r}, T_{ m VM01r}, T_{ m VM02r}, T_{ m VM10r}, T_{ m VM11r}, T_{ m VM12r}, T_{ m VM20r}, T_{ m VM20r}, T_{ m VM21r}, T_{ m VM22r}, T_{ m VM30r}, T_{ m VM31r}, { m and} \ T_{ m VM32r}$	30 mins
$\lambda_{_{ m S}}$	Switch failure rate	$T_{ m S0f}$, $T_{ m S1f}$, and $T_{ m S2f}$	1 year
μ_{S}	Switch repair rate	T_{S0r} , T_{S1r} , and T_{S2r}	24 hours
$\omega_{ m mig}$	Network bandwidth within a DCell ₀	$T_{\rm VM00mig}, T_{\rm VM01mig}, T_{\rm VM10mig}, T_{\rm VM11mig}, T_{\rm VM20mig}, \\ T_{\rm VM21mig} \\ T_{\rm VM001mig}, T_{\rm VM010mig}, T_{\rm VM002mig}, T_{\rm VM020mig}, T_{\rm VM012mig}, \\ T_{\rm VM021mig} \\ T_{\rm VM101mig}, T_{\rm VM110mig}, T_{\rm VM102mig}, T_{\rm VM120mig}, T_{\rm VM112mig}, \\ T_{\rm VM121mig} \\ T_{\rm VM201mig}, T_{\rm VM210mig}, T_{\rm VM202mig}, T_{\rm VM220mig}, T_{\rm VM212mig}, \\ T_{\rm VM201mig} \\ T_{\rm VM201mig}, T_{\rm VM310mig}, T_{\rm VM302mig}, T_{\rm VM320mig}, T_{\rm VM312mig}, \\ and T_{\rm VM301mig}$	1 Gb/s
$\omega_{ m m}$	Network bandwidth between two DCell _o s	$T_{\rm VM01m}, T_{\rm VM10m}, T_{\rm VM02m}, T_{\rm VM20m}, T_{\rm VM03m}, T_{\rm VM30m}, \\ T_{\rm VM12m}, T_{\rm VM21m}, T_{\rm VM13m}, T_{\rm VM31m}, T_{\rm VM23m}, \text{and } T_{\rm VM32m}$	256 Mb/s
$S_{ m VM}$	Memory size of a VM		10 GB
N_{00}, N_{01}, N_{02}	Number of VMs running on respective hosts H00, H01, and H02		1
$N_{10}, N_{11}, N_{12}, \\ N_{20}, N_{21}, N_{22}, \\ N_{30}, N_{31}, N_{32}$	Number of VMs running on respective hosts H10, H11, H12, H20, H21, H22, H30, H31, and H32		0

TABLE 3: Case studies in steady state analyses.

Case	Description
I	A standalone DCell ₀ with two servers (DCN0)
II	A standalone DCell ₀ with three servers (DCN1)
III	A DCN with two servers in a DCell ₀ (DCN2)
IV	A DCN with three servers in a DCell ₀ (DCN3)

To observe the impact of the number of VMs $(n_{\rm VM})$ on the steady state availability of DCNs, we conduct the analyses with different values of $n_{\rm VM}$ (from 1 to 6) until the SPNP suffers unexpected memory computation errors (m.e). Table 5 shows the analysis results of steady state availabilities and their corresponding number of nines. In all cases, the increase of $n_{\rm VM}$ slightly gains higher but not significantly system availability for the DCNs.

The adoption of DCell network topology and the increase of number of VMs in DCNs do achieve significantly higher

availability for the systems. Nevertheless, it is costly and time-consuming to model and analyze such complicated systems. Table 6 points out the complexity of the analyses using two measures: (i) number of tangible markings and (ii) number of marking-to-marking transitions. As shown clearly, the number of VMs exposes a major influence on the system complexity in modeling and analysis, especially for the systems under the adoption of DCell network topology (DCN2 and DCN3). For DCN0 and DCN1 (without adoption of DCell), the system complexity increases from tens or hundreds to about hundreds or thousands of markings and transitions as the $n_{\rm VM}$ increases from 1 to 6. Whereas in the cases of DCN2 and DCN3, the system complexity boosts up from tens to tens of millions of markings and marking transitions as $n_{\rm VM}$ increases. The vast increase of the system complexity quickly causes memory errors in computation. The DCN2 SRN model suffers unexpected memory errors as the number of marking transitions is at tens of millions. The memory errors in analysis of the DCN3 SRN model occur

TABLE 4: Steady state and downtime cost analyses.

Case	Туре	Steady state availability	Number of nines	Downtime per year (minutes)	Downtime cost per year (USD)
I	DCN0	0.997240422469	2.55	1450.4	23,206,943
II	DCN1	0.997259841407	2.56	1440.2	23,043,637
III	DCN2	0.999950276761	4.30	26.1	418,152
IV	DCN3	0.999993583541	5.19	3.4	53,959

TABLE 5: Impact of total number of VMs on system steady state availability.

$n_{ m VM}$	DC	N0	DC	CN1	DC	N2	DC	N3
1	0.997064755072	2.532356	0.997077756809	2.5343	0.999773875854	3.646	0.999803564319	3.71
2	0.997240422469	2.559157	0.997257682983	2.5619	0.999950276761	4.303	0.999989752473	4.99
3	0.997240488479	2.559168	0.997259841407	2.5622	0.999950574780	4.306	0.999993583541	5.19
4	0.997240519634	2.559173	0.997261106490	2.5624	0.999950839446	4.308	m.e	m.e
5	0.997240550678	2.559178	0.997261943932	2.5626	0.999951101800	4.311	m.e	m.e
6	0.997240759564	2.559210	0.997262539658	2.5627	m.e	m.e	m.e	m.e

m.e: memory error.

TABLE 6: Analysis complexity of tangible markings and marking-to-marking transitions.

Case	Туре		1	2	2	;	3	
Casc	Туре	(i)	(ii)	(i)	(ii)	(i)	(ii)	
I	DCN0	26	54	70	180	150	440	
II	DCN1	80	198	272	820	664	2270	
III	DCN2	91	196	5546	23015	64005	342328	
IV	DCN3	365	1050	104567	610066	4224477	32140034	
Case	Tymo	Tymo		4	5	5	(6
Case	Type	(i)	(ii)	(i)	(ii)	(i)	(ii)	
I	DCN0	280	900	476	280	900	476	
II	DCN1	1408	5248	2710	1408	5248	2710	
III	DCN2	403859	2485761	1862923	403859	2485761	1862923	
IV	DCN3	n	ı.e	m	.e	n	ı.e	

m.e: memory error.

Table 7: Sensitivity of system availability with default parameters.

Parameters	DCN2	DCN3
λ_{H}	$6.43 \cdot E - 08$	$6.12 \cdot E - 09$
$\mu_{ m H}$	$-2.47 \cdot E - 06$	$-2.08 \cdot E - 08$
$\lambda_{ m VM}$	$1.66 \cdot E - 11$	$2.17 \cdot E - 11$
$\mu_{ m VM}$	$-9.46 \cdot E - 08$	$-1.24 \cdot E - 07$
$\lambda_{_{ m S}}$	$1.16 \cdot E - 09$	$6.32 \cdot E - 10$
$\mu_{ m S}$	$-1.30 \cdot E - 08$	$-2.54 \cdot E - 09$
$\omega_{ m mig}$	$9.63 \cdot E - 09$	$4.55 \cdot E - 09$
$\omega_{ m m}$	$3.85 \cdot E - 08$	$2.15 \cdot E - 08$
$S_{ m VM}$	$-9.87 \cdot E - 07$	$-5.51 \cdot E - 07$

as n_{VM} is larger than 3 and thus the complexity could reach hundreds of millions of markings and transitions.

5.2. Sensitivity Analysis. The major purposes of sensitivity analysis in this study are (i) to optimize system design

and (ii) to pinpoint the bottlenecks regarding availability, performance, and performability of the systems. Therefore, we conduct a variety of parametric sensitivity analyses of the DCN2 and DCN3 SRN models with respect to the major parameters in Table 2. The analysis results are shown in Table 7. We see that the parameters $\mu_{\rm H}$ and $\mu_{\rm VM}$ assume the greatest importance in system steady state availability of both DCN2 and DCN3, since they present highest absolute values. A major impact upon any change in the value of these parameters bears on the system availability in opposite directions. Sensitivities with respect to these two parameters are negative, since the smaller values the repair times of hosts and VMs get, the higher availability the DCN can achieve. This result reminds the system owner to improve the performance and readiness of the repair and maintenance services in a data center to mitigate recovery time of failed components. Nevertheless, in comparison between the cases of DCN2 and DCN3, the absolute value of the sensitivity with respect to the parameter $\mu_{\rm H}$ is greater in the case of DCN2 than it is in the case of DCN3. However the absolute value of the sensitivity with respect to the parameter μ_{VM} is higher in the case of DCN3 compared to it in the case of DCN2. The results imply that, in the DCN with more hosts in a DCell₀ and more DCell₀ units in the network (DCN3 in comparison with DCN2), the recovery of software subsystems (VMs) plays a more important role compared to the recovery of hardware systems (hosts). Thus in the DCell-based DCN with higher number of VMs and DCell₀ units, a failure of a host does not cause a significant impact on the operations of a VM compared to the failure of the VM itself, since the VMs have more chances to be migrated onto other hosts in other DCell₀s. Therefore, the DCN system designer ought to consider the thorough adoption of software fault tolerance on VMs in a DCN. In Table 7 we also see that the parameter S_{VM} contributes a significant impact on the system availability. The negative values of the sensitivities with respect to the parameter S_{VM} in both cases of DCN2 and DCN3 say that the bigger size of a VM in storage system causes a declining tendency of system availability, since the VM migration processes between hosts within or between DCell₀ units last longer to complete. Furthermore, the sensitivity with respect to the parameter $\omega_{\rm m}$ has higher value than it with respect to the parameter $\omega_{\rm mig}$, and both are positive. This is to say that an increase in network speed leads to a corresponding increase of system availability, since the time to migrate VMs could be reduced. Also, the link bandwidth of the pairs of hosts between DCell₀s ($\omega_{\rm m}$) reveals a more important contribution on the system availability than that of the pairs of hosts within a DCell₀ ($\omega_{\rm mig}$). The reason is that the cross-links of the hosts between DCell₀s are to tolerate the switch failures (which disconnect the communication between system users and VMs in a DCell₀) so that a VM is migrated from a DCell₀ to others upon any failure of the switch in the DCell₀. However, the cross-links could cause the high complexity of network routing and the requirements of high speed links could lead to a huge amount of overall system cost. Thus the system design has to be aware of the trade-offs between system availability and performance and the overall cost of networking.

Figure 7 shows the sensitivity analysis results with respect to the major impacting parameters in both DCN2 and DCN3. The analyses are carried out by altering the value of a parameter of interest as the other parameters remained constant.

Figures 7(a) and 7(b) show the analysis results with respect to MTTFs of host (λ_H), VM (λ_{VM}), and switch (λ_S). There are several similarities of the graphs in which (i) in the early period (0–1000] hours the system availability increases quickly as long as the MTTFs increase and (ii) the system availability slowly increases and approaches a steady value as the MTTFs get greater values in the late period (over 1000 hours). Also, switches in each DCell show its major impact on the system availability. If the MTTF of switches gets a low value in the early period (the switches fails more frequently), the system availability is severely pulled down in comparison with the sensitivity analysis results of system availability with respect to MTTFs of host and VM. The MTTFs of host and VM only contribute a little impact on system availability in the early period (showed by declining vertical graphs with circle and star markers) but mostly do not

cause a great impact on system availability in the late period (depicted by approximately horizontal graphs with circle and star markers). This is to say that a DCN is likely prone to switches' failures. Since the switches are the key components to connect a number of physical hosts in DCell₀s, a failure of a switch severely causes a failure of the whole DCell₀ (unable to connect system user to the DCell₀).

Figures 7(c) and 7(d) present the results of availability sensitivity analysis with respect to MTTRs of host (μ_H), VM (μ_{VM}) , and switch (μ_S) . The figures apparently reflect the significant impact of the MTTR of software system (VMs) onto the overall system availability. In both DCN2 and DCN3, as the MTTR of VMs increases to get higher values, the system availability slides down very quickly depicted by the graph marked with stars. This is because the user's applications run on VMs hence the VMs' up or down states decisively influence the system availability. Moreover, in the DCN2 with less number of hosts, the increase of MTTR of hosts can decrease the system availability as shown by the graph with circle marker in Figure 7(c). But in the DCN3 with more numbers of hosts, the value of MTTR of hosts does not significantly impact the system availability as shown in Figure 7(d). The reason is that if a host fails, the VM running on that host can be migrated to other hosts in the same DCell₀. In the DCN2 with less number of hosts, the longer time the repair of hosts spends, the less chance the system can have to be available. In the DCN3 with more numbers of hosts and under the assumption of a high available system where a host can be recovered before the last host's failure, the MTTR of hosts mostly does not impact the system availability. At last, the MTTR of switches does not affect the system availability as depicted by the graphs with triangle markers in Figures 7(c)and 7(d), since, as long as a switch fails, all VMs running on the DCell₀ of that switch are migrated to the other DCell₀s.

Figures 7(e) and 7(f) depict the availability sensitivity with respect to network bandwidths within a DCell₀ ($\omega_{\rm mig}$) and between DCell₀s ($\omega_{\rm m}$). If the network speeds within or between DCell₀s surpass a specific value at about 400 Mb/s, the system can achieve high availability. However if the speeds get slower, the system availability is pulled down quickly. The figures also reflect the importance of network bandwidth within a DCell₀ compared to that between DCell₀s. The low value of the network bandwidth within a DCell₀ pulls down the system availability more severely (as depicted by the vertical slope of the star-marked graph) than that between DCell₀s does. The reason is that the connection between hosts in a DCell₀ is to tolerate hosts' failures which are more frequent to occur but the connection between hosts among different DCell₀ is to tolerate switches' failures which happen less frequently.

Figure 7(g) shows the availability sensitivity with respect to VM image sizes ($S_{\rm VM}$). Under the default values of parameters, the size of VM image files affects the system availability in a negative manner. As the size increases, the system availability slides down quickly. Furthermore, the VM image size has greater influence on the system availability in the DCN with less number of hosts in a DCell $_0$ (DCN2) than that in the DCN with more numbers of hosts (DCN3) does. The bigger size of VM can pull down the system availability

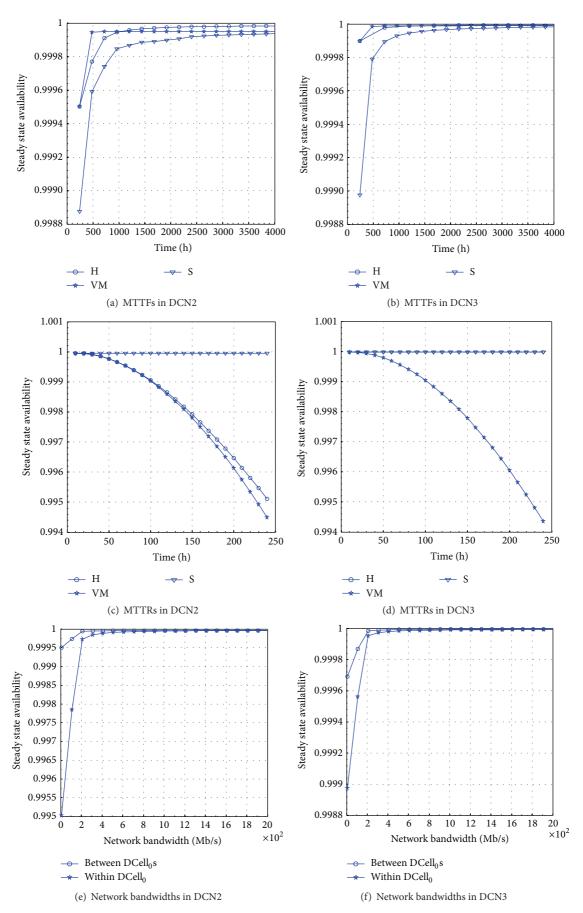


FIGURE 7: Continued.

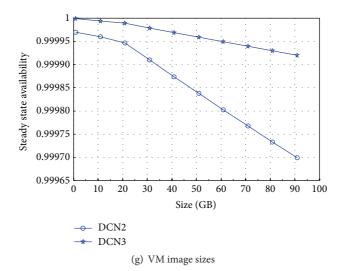


FIGURE 7: Sensitivity analysis of DCell-based DCNs.

more quickly in DCN2 in comparison with that in DCN3. This is depicted by the smaller slope of the star-marked graph (for DCN3) compared to that of the circle-marked graph (DCN2). This result implies that, in a DCell-based DCN with higher number of physical hosts in a DCell₀, the system can have better ability to tolerate hardware and software failures and thus be able to deliver bigger size of VM image files.

5.3. Discussion. A practical DCell-based DCN system comprises tens to hundreds of thousands of hardware components (hosts, switches, links, etc.) and thus hosts even an enormous number of VMs in a very complicated topology of networking and routing as described in [1, 4]. An effort to model and analyze such DCN system is critically important to help provide a guide basis for design and management of both hardware and software subsystems. We find this a fruitful topic for further work on system scalability. Nevertheless, the endeavor to build a complete and monolithic model to capture the whole system behaviors also confronts largeness problem (also known as state-space explosion) in modeling. To deal with this issue, one may adopt different modeling techniques and methodologies such as state truncation [61], state aggregation [62], model decomposition [63, 64], state exploration [65, 66], and model composition [67, 68]. Other different methodologies have been also adopted popularly in literature, which are also appropriate to deal with scalability and largeness problems of modeling a large DCN system such as (i) hierarchical models, which partition a complex model into a hierarchy of submodels [69] or combine combinatorial models and state-space models [70–72], (ii) interactive models [22, 73, 74], which divide a large monolithic model into a number of smaller scale models with comprehensive interactions and dependencies, (iii) fixed-point iterative models [75], and (iv) discrete-event simulation [76]. Thus this is a broad future research avenue to scale up system configuration and to resolve the largeness problem in modeling a DCN system. In this paper, our sharp focus is on system capability of fault tolerance and business continuity through availability

modeling and analysis. We have shown that the DCell-based DCN can have higher availability to assure business continuity even in the presence of severe failures of components. Nevertheless, it is necessary to observe and study the system in different perspectives including reliability [77], survivability [78], performability [79] for instance. These topics are still open for future endeavor.

6. Conclusion

This paper has presented a comprehensive availability modeling and sensitivity analysis of a DCell-based DCN. Our work studied two typical DCell configurations of the DCN, respectively, comprising two and three hosts in a DCell₀. Our focus is on the fault tolerant capability and business continuity of the DCNs; thus the VM live migration mechanisms are incorporated in detail to tolerate failures of switches and hosts. The modeling captured the distributed fault tolerant routing protocol designed in system architectures. A variety of analyses were carried out thoroughly in consideration of different measures of interest. The steady state availability analyses have shown that the DCell-based DCNs can assure HA and business continuity, tolerate hardware failures of switches and hosts, and enhance vastly the system's overall availability. The increasing number of VMs in a DCN slightly improves the system availability but causes a high complexity and largeness problems in modeling and analysis. The comprehensive sensitivity analyses of system steady state availability were also performed in order to observe the system characteristics and behaviors upon any change of major impacting parameters. The sensitivity analysis results have pointed out that (i) recovery actions of hosts and VMs are significantly important to mitigate system downtime, (ii) recovery actions of software subsystem (VMs) in a DCellbased DCN cause major impacts on system availability in comparison with those of hardware subsystems (hosts and switches), and (iii) network bandwidth of the link between DCell₀s is a critical parameter to obtain and maintain high availability of the system. This study brings about a guide basis to help manage and operate a DCN in data centers in terms of (i) maintenance and repair readiness, (ii) awareness of software fault tolerance in DCNs, and (iii) selection basis of network performance and availability and cost to avoid potential risks as well as tolerate faults.

Conflict of Interests

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

Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2015-H8501-15-1011) supervised by the IITP (Institute for Information & Communications Technology Promotion). The authors would like to thank Professor Dr. Kishor Trivedi, Professor of Duke University, United States, for providing SPNP.

References

- [1] Y. Liu and J. Muppala, "Fault-tolerance characteristics of data center network topologies using fault regions," in *Proceedings of the 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN '13)*, pp. 1–6, 2013.
- [2] D. A. Patterson, "A simple way to estimate the cost of downtime," in *Proceedings of the 16th USENIX Conference on System Administration*, pp. 185–188, 2002.
- [3] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," ACM SIGCOMM Computer Communication Review, vol. 38, no. 4, pp. 63–74, 2008.
- [4] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, and S. Lu, "Dcell: a scalable and fault-tolerant network structure for data centers," in *Proceedings of the ACM SIGCOMM Conference on Data Communication (SIGCOMM '08)*, pp. 75–86, Seattle, Wash, USA, August 2008.
- [5] C. Guo, G. Lu, D. Li et al., "BCube: a high performance, server-centric network architecture for modular data centers," in *Proceedings of the ACM SIGCOMM Conference on Data Communication (SIGCOMM '09)*, pp. 63–74, ACM, Barcelona, Spain, August 2009.
- [6] M. Manzano, K. Bilal, E. Calle, and S. U. Khan, "On the connectivity of data center networks," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2172–2175, 2013.
- [7] R. Mikkilineni and G. Kankanhalli, "Using virtualization to prepare your data center for 'real-time assurance of business continuity," in *Proceedings of the 19th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE '10)*, pp. 76–81, June 2010.
- [8] K. Bilal, S. U. R. Malik, O. Khalid et al., "A taxonomy and survey on green data center networks," *Future Generation Computer Systems*, vol. 36, pp. 189–208, 2014.
- [9] S. Loveland, E. M. Dow, F. LeFevre, D. Beyer, and P. F. Chan, "Leveraging virtualization to optimize high-availability system

- configurations," *IBM Systems Journal*, vol. 47, no. 4, pp. 591–604, 2008.
- [10] T. A. Nguyen, D. S. Kim, and J. S. Park, "A comprehensive availability modeling and analysis of a virtualized servers system using stochastic reward nets," *The Scientific World Journal*, vol. 2014, Article ID 165316, 18 pages, 2014.
- [11] M. F. Bari, R. Boutaba, R. Esteves et al., "Data center network virtualization: a survey," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 2, pp. 909–928, 2013.
- [12] C. Clark, K. Fraser, S. Hand et al., "Live migration of virtual machines," in *Proceedings of the 2nd Conference on Symposium* on Networked Systems Design & Implementation (NSDI '05), vol. 2, pp. 273–286, USENIX Association, Berkeley, Calif, USA, 2005
- [13] R. W. Ahmad, A. Gani, S. H. A. Hamid, M. Shiraz, A. Yousafzai, and F. Xia, "A survey on virtual machine migration and server consolidation techniques for cloud data centers," *Journal of Network and Computer Applications*, vol. 52, pp. 11–25, 2015.
- [14] E. Rodriguez, G. Alkmim, D. M. Batista, and N. L. S. Da Fonseca, "Live migration in green virtualized networks," in *Proceedings of the IEEE International Conference on Communications (ICC '13)*, pp. 2262–2266, June 2013.
- [15] Y. Liu, D. Lin, J. Muppala, and M. Hamdi, "A study of fault-tolerance characteristics of data center networks," in *Proceedings of the IEEE/IFIP 42nd International Conference on Dependable Systems and Networks Workshops (DSN-W '12)*, pp. 1–6, June 2012.
- [16] K. Bilal, M. Manzano, S. U. Khan, E. Calle, K. Li, and A. Y. Zomaya, "On the characterization of the structural robustness of data center networks," *IEEE Transactions on Cloud Computing*, vol. 1, no. 1, p. 1, 2013.
- [17] K. S. Trivedi, D.-S. Kim, and R. Ghosh, "System availability assessment using stochastic models," *Applied Stochastic Models* in *Business and Industry*, vol. 29, no. 2, pp. 94–109, 2013.
- [18] R. D. S. Matos Jr., A. P. Guimarães, K. M. A. Camboim, P. R. M. Maciel, and K. S. Trivedi, "Sensitivity analysis of availability of redundancy in computer networks," in *Proceedings of the 4th International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ '11)*, pp. 115–121, April 2011.
- [19] R. Alshahrani and H. Peyravi, "Modeling and simulation of data center networks," in *Proceedings of the 2nd ACM SIGSIM/PADS Conference on Principles of Advanced Discrete Simulation (SIGSIM-PADS '14)*, pp. 75–82, May 2014.
- [20] ADC Telecommunications, "TIA-942 data center standards overview," White Paper, ADC Telecommunications, Minneapolis, Minn, USA, 2006.
- [21] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: research problems in data center networks," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 1, p. 68, 2008.
- [22] F. Longo, R. Ghosh, V. K. Naik, and K. S. Trivedi, "A scalable availability model for Infrastructure-as-a-Service cloud," in Proceedings of the IEEE/IFIP 41st International Conference on Dependable Systems and Networks (DSN '11), pp. 335–346, June 2011.
- [23] R. N. Mysore, A. Pamboris, N. Farrington et al., "PortLand: a scalable fault-tolerant layer 2 data center network fabric," in *Proceedings of the ACM SIGCOMM Conference on Data Communication (SIGCOMM '09)*, pp. 39–50, August 2009.
- [24] M. Al-fares, S. Radhakrishnan, B. Raghavan, N. Huang, and A. Vahdat, "Hedera: dynamic flow scheduling for data center

- networks," in *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation (NSDI '10)*, p. 19, Boston, Mass, USA, March 2010.
- [25] A. Greenberg, J. R. Hamilton, N. Jain et al., "VL2: a scalable and flexible data center network," ACM SIGCOMM Computer Communication Review, vol. 39, no. 4, p. 51, 2009.
- [26] D. Li, C. Guo, H. Wu, K. Tan, Y. Zhang, and S. Lu, "FiConn: using backup port for server interconnection in data centers," in *Proceedings of the 28th Conference on Computer Communica*tions (IEEE INFOCOM '09), pp. 2276–2285, April 2009.
- [27] D. Lin, Y. Liu, M. Hamdi, and J. Muppala, "Hyper-BCube: a scalable data center network," in *Proceedings of the IEEE International Conference on Communications (ICC '12)*, pp. 2918–2923, June 2012.
- [28] T. Lumpp, J. Schneider, J. Holtz et al., "From high availability and disaster recovery to business continuity solutions," *IBM Systems Journal*, vol. 47, no. 4, pp. 605–619, 2008.
- [29] P. Patel, A. Ranabahu, and A. Sheth, Service Level Agreement in Cloud Computing, Kno.e.sis Publications, 2009.
- [30] S. K. Garg, A. N. Toosi, S. K. Gopalaiyengar, and R. Buyya, "SLA-based virtual machine management for heterogeneous workloads in a cloud datacenter," *Journal of Network and Computer Applications*, vol. 45, pp. 108–120, 2014.
- [31] T. Thein, S.-D. Chi, and J. S. Park, "Improving fault tolerance by virtualization and software rejuvenation," in *Proceedings of the 2nd Asia International Conference on Modelling & Simulation (AMS '08)*, pp. 855–860, May 2008.
- [32] J. Daniels, "Server virtualization architecture and implementation," *Crossroads*, vol. 16, no. 1, pp. 8–12, 2009.
- [33] P. Barham, B. Dragovic, K. Fraser et al., "Xen and the art of virtualization," in *Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP '03)*, vol. 37, p. 164, ACM, Bolton Landing, NY, USA, October 2003.
- [34] R. Y. Ameen and A. Y. Hamo, "Survey of server virtualization," International Journal of Computer Science and Information Security, vol. 11, no. 3, pp. 65–74, 2013.
- [35] M. Melo, P. Maciel, J. Araujo, R. Matos, and C. Araujo, "Availability study on cloud computing environments: Live migration as a rejuvenation mechanism," in *Proceedings of the 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN '13)*, pp. 1–6, June 2013.
- [36] P. G. J. Leelipushpam and J. Sharmila, "Live VM migration techniques in cloud environment—a survey," in *Proceedings* of the IEEE Conference on Information and Communication Technologies (ICT '13), pp. 408–413, April 2013.
- [37] J. T. Blake, A. L. Reibman, and K. S. Trivedi, "Sensitivity analysis of reliability and performability measures for multiprocessor systems," in *Proceedings of the ACM SIGMETRICS Conference* on Measurement and Modeling of Computer Systems (SIGMET-RICS '88), vol. 16, pp. 177–186, ACM, Santa Fe, NM, USA, May 1988.
- [38] R. Matos, J. Araujo, D. Oliveira, P. Maciel, and K. Trivedi, "Sensitivity analysis of a hierarchical model of mobile cloud computing," *Simulation Modelling Practice and Theory*, vol. 50, pp. 151–164, 2015.
- [39] M. P. Bendsøe and N. Kikuchi, "Generating optimal topologies in structural design using a homogenization method," Computer Methods in Applied Mechanics and Engineering, vol. 71, no. 2, pp. 197–224, 1988.
- [40] P. M. Frank and M. Eslami, "Introduction to system sensitivity theory," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 10, no. 6, pp. 337–338, 1980.

- [41] R. Y. Al-Jaar, "Book review: the art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling by Raj Jain (John Wiley & Sons 1991)," *ACM SIGMETRICS Performance Evaluation Review*, vol. 19, no. 2, pp. 5–11, 1991.
- [42] K. S. Trivedi, D. S. Kim, A. Roy, and D. Medhi, "Dependability and security models," in *Proceedings of the 7th International Workshop on the Design of Reliable Communication Networks*, pp. 11–20, October 2009.
- [43] T. Wang, Z. Su, Y. Xia, Y. Liu, J. Muppala, and M. Hamdi, "SprintNet: a high performance server-centric network architecture for data centers," in *Proceedings of the 1st IEEE International Conference on Communications (ICC '14)*, pp. 4005–4010, June 2014.
- [44] X. Wu, D. Turner, G. Chen et al., "NetPilot: automating datacenter network failure mitigation," ACM SIGCOMM Computer Communication Review, vol. 42, no. 4, pp. 443–454, 2012.
- [45] P. Gill, N. Jain, and N. Nagappan, "Understanding network failures in data centers: measurement, analysis, and implications," in *Proceedings of the ACM SIGCOMM Conference*, pp. 350–361, ACM, Ontario, Canada, August 2011.
- [46] K. V. Vishwanath and N. Nagappan, "Characterizing cloud computing hardware reliability," in *Proceedings of the 1st ACM Symposium on Cloud Computing (SoCC '10)*, pp. 193–204, June 2010.
- [47] M. G. Rabbani, M. F. Zhani, and R. Boutaba, "On achieving high survivability in virtualized data centers," *IEICE Transactions on Communications*, vol. E97-B, no. 1, pp. 10–18, 2014.
- [48] T. Adeshiyan, C. R. Attanasio, E. M. Farr et al., "Using virtualization for high availability and disaster recovery," *IBM Journal* of Research and Development, vol. 53, no. 4, pp. 8:1–8:11, 2009.
- [49] E. Bin, O. Biran, O. Boni et al., "Guaranteeing high availability goals for virtual machine placement," in *Proceedings of the* 31st International Conference on Distributed Computing Systems (ICDCS '11), pp. 700–709, July 2011.
- [50] D. S. Kim, F. Machida, and K. S. Trivedi, "Availability modeling and analysis of a virtualized system," in *Proceedings of the 15th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC '09)*, pp. 365–371, November 2009.
- [51] W. E. Smith, K. S. Trivedi, L. A. Tomek, and J. Ackaret, "Availability analysis of blade server systems," *IBM Systems Journal*, vol. 47, no. 4, pp. 621–640, 2008.
- [52] J. K. Muppala and C. Lin, "Dependability analysis of large-scale distributed systems using stochastic Petri nets," in *Proceedings of* the IEEE International Conference on Systems, Man, and Cybernetics, vol. 4, pp. 3033–3038, IEEE, Beijing, China, October 1996.
- [53] F. Machida, D. S. Kim, and K. S. Trivedi, "Modeling and analysis of software rejuvenation in a server virtualized system with live VM migration," *Performance Evaluation*, vol. 70, no. 3, pp. 212–230, 2013.
- [54] Y. Cao, H. Sun, K. S. Trivedi, and J. J. Han, "System availability with non-exponentially distributed outages," *IEEE Transactions* on *Reliability*, vol. 51, no. 2, pp. 193–198, 2002.
- [55] G. Ciardo, J. K. Muppala, K. S. Trivedi, J. K. Muppala, G. Ciardo, and K. S. Trivedi, "Stochastic reward nets for reliability prediction," *Journal of Parallel and Distributed Computing*, vol. 15, no. 3, pp. 255–269, 1992.
- [56] C. Constazltinescu and K. Trivedi, "A stochastic reward net model for dependability analysis of real-time computing systems," in *Proceedings of the 2nd IEEE Workshop on Real-Time* Applications, pp. 142–146, Washington, DC, USA, 1994.

- [57] G. Ciardo, J. Muppala, and K. S. Trivedi, "SPNP: stochastic Petri net package," in *Proceedings of the 3rd International Workshop* on Petri Nets and Performance Models (PNPM '89), pp. 142–151, IEEE, Kyoto, Japan, December 1989.
- [58] B. Silva, P. Maciel, E. Tavares, and A. Zimmermann, "Dependability models for designing disaster tolerant cloud computing systems," in *Proceedings of the 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN '13)*, pp. 1–6, June 2013.
- [59] F. Machida, D. S. Kim, and K. S. Trivedi, "Modeling and analysis of software rejuvenation in a server virtualized system," in *Proceedings of the IEEE 2nd International Workshop on Software Aging and Rejuvenation*, pp. 1–6, IEEE, San Jose, Calif, USA, November 2010.
- [60] M. Stansberry, 2013 Data Center Industry Survey, Uptime Institute LLC, New York, NY, USA, 2013.
- [61] K. Goševa-Popstojanova and K. S. Trivedi, "Stochastic modeling formalisms for dependability, performance and performability," in *Performance Evaluation: Origins and Directions*, vol. 1769 of *Lecture Notes in Computer Science*, pp. 403–422, Springer, Berlin, Germany, 2000.
- [62] P. Buchholz, "An adaptive aggregation/disaggregation algorithm for hierarchical Markovian models," *European Journal of Operational Research*, vol. 116, no. 3, pp. 545–564, 1999.
- [63] S. M. Koriem, "Accurate approximate analysis of cell-based switch architectures," *Journal of Systems and Software*, vol. 45, no. 2, pp. 155–171, 1999.
- [64] M. A. Marsan, R. Gaeta, and M. Meo, "Accurate approximate analysis of cell-based switch architectures," *Performance Evaluation*, vol. 45, no. 1, pp. 33–56, 2001.
- [65] I. Davies, W. J. Knottenbelt, and P. S. Kritzinger, "Symbolic methods for the state space exploration of GSPN models," in Proceedings of the 12th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation (TOOLS '02), pp. 188–199, 2002.
- [66] A. S. Miner, "Efficient state space generation of GSPNs using decision diagrams," in Proceedings of the International Conference on Dependable Systems and Networks (DNS '02), pp. 637– 646, June 2002.
- [67] P. Ballarini, S. Donatelli, and G. Franceschinis, "Parametric stochastic well-formed nets and compositional modelling," in Application and Theory of Petri Nets 2000: 21st International Conference, ICATPN 2000 Aarhus, Denmark, June 26–30, 2000 Proceedings, vol. 1825 of Lecture Notes in Computer Science, pp. 43–62, Springer, Berlin, Germany, 2000.
- [68] W. J. Knottenbelt, P. G. Harrison, M. A. Mestern, and P. S. Kritzinger, "A probabilistic dynamic technique for the distributed generation of very large state spaces," *Performance Evaluation*, vol. 39, no. 1–4, pp. 127–148, 2000.
- [69] M. Lanus, L. Yin, and K. S. Trivedi, "Hierarchical composition and aggregation of state-based availability and performability models," *IEEE Transactions on Reliability*, vol. 52, no. 1, pp. 44– 52, 2003.
- [70] O. C. Ibe, R. C. Howe, and K. S. Trivedi, "Approximate availability analysis of VAXcluster systems," *IEEE Transactions* on *Reliability*, vol. 38, no. 1, pp. 146–152, 1989.
- [71] K. S. Trivedi, R. Vasireddy, D. Trindade, S. Nathan, and R. Castro, "Modeling high availability," in *Proceedings of the 12th Pacific Rim International Symposium on Dependable Computing (PRDC '06)*, pp. 154–164, December 2006.

- [72] J. T. Blake and K. S. Trivedi, "Reliability analysis of interconnection networks using hierarchical composition," *IEEE Transactions on Reliability*, vol. 38, no. 1, pp. 111–120, 1989.
- [73] R. Ghosh, F. Longo, V. K. Naik, and K. S. Trivedi, "Modeling and performance analysis of large scale IaaS clouds," *Future Generation Computer Systems*, vol. 29, no. 5, pp. 1216–1234, 2013.
- [74] R. Ghosh, K. S. Trivedi, V. K. Naik, and D. S. Kim, "End-to-end performability analysis for infrastructure-as-a-service cloud: an interacting stochastic models approach," in *Proceedings of the 16th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC '10)*, pp. 125–132, December 2010.
- [75] V. Mainkar and K. S. Trivedi, "Fixed point iteration using stochastic reward nets," in *Proceedings of the 6th International* Workshop on Petri Nets and Performance Models, pp. 21–30, October 1995.
- [76] T. Angskun, G. Bosilca, G. Fagg, J. Pješivac-Grbović, and J. J. Dongarra, "Reliability analysis of self-healing network using discrete-event simulation," in *Proceedings of the 7th IEEE International Symposium on Cluster Computing and the Grid* (CCGrid '07), pp. 437–444, May 2007.
- [77] E. Topuz, "Reliability and availability basics," *IEEE Antennas and Propagation Magazine*, vol. 51, no. 5, pp. 231–236, 2009.
- [78] J. F. Castet and J. Saleh, "Survivability and resiliency of space-craft and space-based networks: a framework for characterization and analysis," in *Proceedings of the AIAA SPACE Conference & Exposition*, San Diego, Calif, USA, September 2008.
- [79] R. Entezari-Maleki, K. S. Trivedi, and A. Movaghar, "Performability evaluation of grid environments using stochastic reward nets," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 2, pp. 204–216, 2015.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 835460, 10 pages http://dx.doi.org/10.1155/2015/835460

Research Article

In-Plane Vibration Response of the Periodic Viaduct on Saturated Soil under Rayleigh Surface Wave

Hai-yan $Ju^{1,2}$ and Ming-fu $Fu^{1,2}$

¹School of Civil Engineering and Architecture, Nanchang University, Nanchang, Jiangxi 330031, China

Correspondence should be addressed to Hai-yan Ju; 2005992511@nit.edu.cn

Received 10 April 2015; Accepted 1 July 2015

Academic Editor: Xiaobo Qu

Copyright © 2015 H.-y. Ju and M.-f. Fu. 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 this study, the in-plane vibration response of the periodic viaduct on saturated soil under Rayleigh surface wave is studied. The Floquet transform method is used to decompose Rayleigh surface wave into a set of spatial harmonic waves. Considering the periodic condition of the viaduct, the wave number domain dynamic response of the periodic viaduct on saturated soil subjected to Rayleigh surface wave excitation is obtained by the transfer matrix method. Then the space domain dynamic response is retrieved by means of the inverse Floquet transform. Numerical results show that when the periodic viaduct is undergoing in-plane vibration, there exist three kinds of characteristic waves corresponding to axial compression, transverse shear, and bending vibration. Furthermore, when the frequency of Rayleigh wave is within the pass band of the periodic viaduct, the disturbance propagates over a very long distance and the attenuation of the wave motion far from the source is determined by the characteristic wave with the smallest attenuation, while the vibration attenuates rapidly and propagates in a short distance when the frequency of excitation source is in the range of band gap of periodic structure.

1. Introduction

As we know, viaduct structure is widely used in engineering. Usually, the viaduct structure generally has equal spans, which means the distance between adjacent piers of a multispan viaduct structure is constant; it can be considered as periodic structure, with the basic element consisting of three parts: a pier, two longitudinal beams, and three linking springs. The period is the distance between two neighboring piers.

The periodic structure has a significant vibration characteristic where energy band exists in periodic structure [1, 2]. When elastic wave propagates in a periodic structure, the vibration within a certain frequency range cannot be passed, which is called band gap, and the vibration within a certain frequency range can be passed, which is called pass band. It provides a new idea for the seismic design and vibration control by using vibration characteristics of periodic structures. With the rational design of the periodic viaduct in geometry and material parameters, it can ensure the frequency of main seismic waves is in the band gap of

the viaduct structure, which effectively reduces the structure vibration and damage caused by earthquake wave. Otherwise, if great energy seismic waves are difficult to pass the viaduct structure, sharp increase of energy may be caused in the structural. Therefore, considering the energy band principle of periodic structures, the seismic design and vibration control measures can be achieved by adjusting the structure itself without additional structures.

At present, there are many seismic design methods about viaduct, such as Response Spectrum Method [3–5], Time-History Analysis Method [6–8], and Random Vibration Method [9–11]. In the above methods, the viaduct is generally simplified as single degree of freedom system or multidegree of freedom system, and seismic wave is simulated by standing wave. Clearly, the analysis using standing wave method cannot reflect the propagation characteristics of vibration wave in the periodic viaduct. As for the structural vibration induced by seismic wave, the viaduct piers attached to soft foundation were firstly excited. So the seismic energy passed on the viaduct structure varies with the distance from the viaduct pier to vibration source location, usually

²Nanchang Institute of Technology, Nanchang, Jiangxi 330099, China

being nonlinear spatial distribution [12, 13]. Therefore, it is necessary to establish mathematical model which can reflect the propagation characteristics of nonuniform seismic wave in the periodic viaduct so as to provide theoretical basis for seismic design of the viaduct structure.

The periodic viaduct model, consisting of one pier, two longitudinal beams, and three springs, is established to describe the multispan viaduct structure [14, 15]. In this study, as for the Rayleigh surface wave at the bottoms of the piers on the saturated soil, the Floquet transform method is introduced to decompose it into a set of spatial harmonic waves. Considering the periodic condition of the viaduct, the eigenequation for the in-plane vibration of the viaduct in the wave number domain is obtained through transfer matrix method. Then the response in spatial domain of the periodic viaduct on saturated soil under Rayleigh waves can be retrieved by means of the inverse Floquet transform. The influence caused by the characteristic wave propagating in periodic structures and the different Rayleigh waves is discussed.

2. Control Equations of Periodic Viaduct In-Plane Vibration

Figure 1 shows the periodic viaduct structure with infinite number of spans. The track, track plates, and beams of each span are simplified as left and right horizontal beams, which are connected by a pier supported on a semi-infinite saturated ground. The intermediate track between neighboring ballasts connecting the left and right horizontal beams is simulated by spring, which is assumed to be able to support axial force, shear force, and bending moment; and the jointing elements between piers and the left or right horizontal beam are also simulated by springs. Therefore, each unit of the periodic viaduct includes a pier, two horizontal beams, and three springs.

Considering the in-plane vibration of the periodic viaduct due to Rayleigh wave, according to Euler-Bernoulli beam theory [16], the motion equations in frequency wave number domain for the pier of the *n*th span element of viaduct can be achieved as follows:

$$\begin{split} \frac{E_d \partial^2 \widehat{\overline{u}}_d^{(n)} \left(\kappa, z\right)}{\partial z^2} + \rho_d \omega^2 \widehat{\overline{u}}_d^{(n)} \left(\kappa, z\right) &= 0, \\ \frac{E_d I_d \partial^4 \widehat{\overline{v}}_d^{(n)} \left(\kappa, z\right)}{\partial z^4} - \rho_d A_d \omega^2 \widehat{\overline{v}}_d^{(n)} \left(\kappa, z\right) &= 0, \end{split} \tag{1}$$

where E_d , ρ_d , A_d , and I_d are elastic modulus, density, cross-sectional area, and rotational inertia of piers, respectively. $\widehat{\overline{u}}_d^{(n)}$ and $\widehat{\overline{v}}_d^{(n)}$ are the axial and tangential displacements of piers in frequency wave number domain.

Figure 2 shows the sign conventions for the internal forces of pier internal forces. The cross-sectional axial force $\widehat{\overline{N}}_d^{(n)}(\kappa,z)$, shear force $\widehat{\overline{Q}}_d^{(n)}(\kappa,z)$, and bending moment

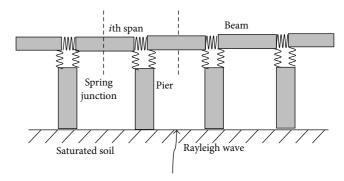


FIGURE 1: A schematic illustration of the periodic viaduct with spring junction subjected to Rayleigh wave.

 $\widehat{\overline{M}}_d^{(n)}(\kappa,z)$ in frequency wave number domain can be expressed as

$$\widehat{\overline{N}}_{d}^{(n)}(\kappa, z) = -\frac{E_{d}A_{d}\partial\widehat{\overline{u}}_{d}^{(n)}}{\partial z},$$

$$\widehat{\overline{M}}_{d}^{(n)}(\kappa, z) = -\frac{E_{d}I_{d}\partial^{2}\widehat{\overline{u}}_{d}^{(n)}}{\partial z^{2}},$$

$$\widehat{\overline{Q}}_{d}^{(n)}(\kappa, z) = -\frac{\partial\widehat{\overline{M}}_{d}^{(n)}}{\partial z}.$$
(2)

For the pier of the nth span element, the state vector $\widehat{\overline{\psi}}_d^{(n)}(\kappa,z)$ at arbitrary position Z, which is composed of displacement vector $\widehat{\overline{\mathbf{q}}}_d^{(n)}(\kappa,z)$ and internal force vector $\widehat{\overline{\mathbf{f}}}_d^{(n)}(\kappa,z)$, is expressed as

$$\widehat{\overline{\psi}}_{d}^{(n)}(\kappa, z) = \left\{ \widehat{\overline{\mathbf{q}}}_{d}^{(n)}(\kappa, z), \widehat{\overline{\mathbf{f}}}_{d}^{(n)}(\kappa, z) \right\}^{T},$$

$$\widehat{\overline{\mathbf{q}}}_{d}^{(n)}(\kappa, z) = \left\{ \widehat{\overline{u}}_{d}^{(n)}(\kappa, z), \widehat{\overline{v}}_{d}^{(n)}(\kappa, z), \widehat{\overline{\theta}}_{d}^{(n)}(\kappa, z) \right\}^{T},$$

$$\widehat{\overline{\mathbf{f}}}_{d}^{(n)}(\kappa, z) = \left\{ \widehat{\overline{N}}_{d}^{(n)}(\kappa, z), \widehat{\overline{Q}}_{d}^{(n)}(\kappa, z), \widehat{\overline{M}}_{d}^{(n)}(\kappa, z) \right\}^{T}.$$
(3)

Under the condition of in-plane vibration, using Euler-Bernoulli beam theory and the sign conventions shown in Figure 2(b), the motion equations for horizontal beam of the viaduct in frequency wave number domain are derived as follows:

$$\frac{E_b \partial^2 \widehat{\overline{u}}_b^{(n)}(\kappa, x)}{\partial x^2} + \rho_b \omega^2 \widehat{\overline{u}}_b^{(n)}(\kappa, x) = 0,$$

$$\frac{E_b I_b \partial^4 \widehat{\overline{v}}_b^{(n)}(\kappa, x)}{\partial x^4} - \rho_b A_b \omega^2 \widehat{\overline{v}}_b^{(n)}(\kappa, x) = 0,$$
(4)

where E_b , ρ_b , A_b , and I_b are elastic modulus, density, cross-sectional area, and rotational inertia of horizontal beams, respectively, $\widehat{\overline{u}}_b^{(n)}$ and $\widehat{\overline{v}}_b^{(n)}$ are axial and tangential displacements of horizontal beams in frequency wave number domain.

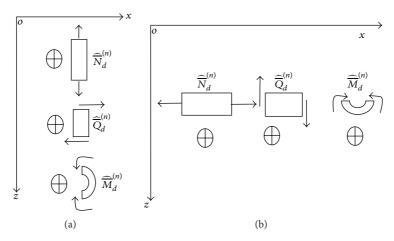


FIGURE 2: The sign conventions for the internal forces of a pier and beam. (a) The sign conventions for the internal forces of a pier. (b) The sign conventions for the internal forces of a beam.

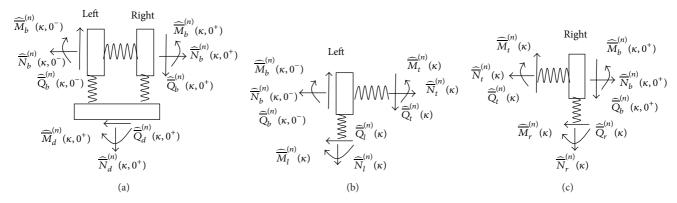


FIGURE 3: The illustration for the junction linking the pier with the left and right beams undergoing in-plane vibration. (a) The overall beambeam-pier junction. (b) The end of the left beam. (c) The end of the right beam.

According to Euler-Bernoulli beam theory [16], the transfer matrix of the pier and horizontal beam system can be deduced; detailed derivation of (5) can be found in Lu and Yuan [15]. Consider

$$\mathbf{T}^{(d)}(z) = \begin{bmatrix} T_{11} & 0 & 0 & T_{14} & 0 & 0 \\ 0 & T_{22} & T_{23} & 0 & T_{25} & T_{26} \\ 0 & T_{32} & T_{33} & 0 & T_{35} & T_{36} \\ T_{41} & 0 & 0 & T_{44} & 0 & 0 \\ 0 & T_{52} & T_{53} & 0 & T_{55} & T_{56} \\ 0 & T_{62} & T_{63} & 0 & T_{65} & T_{66} \end{bmatrix}.$$
(5)

As for beam-beam junction of the nth span in Figure 3, assuming that $k_t^{(t)}$, $k_t^{(s)}$, and $k_t^{(b)}$ are axial compression stiffness, shear stiffness, and rotational stiffness of beam-beam connection spring, respectively, the axial force $\widehat{\overline{N}}_t^{(n)}(\kappa)$, shear force $\widehat{\overline{Q}}_t^{(n)}(\kappa)$, and bending moment $\widehat{\overline{M}}_t^{(n)}(\kappa)$ of beam-beam

connection spring in frequency wave number domain are expressed as

$$\widehat{\overline{N}}_{t}^{(n)}(\kappa) = k_{t}^{(t)} \left[\widehat{\overline{u}}_{b}^{(n)}(\kappa, 0^{+}) - \widehat{\overline{u}}_{b}^{(n)}(\kappa, 0^{-}) \right],$$

$$\widehat{\overline{Q}}_{t}^{(n)}(\kappa) = k_{t}^{(s)} \left[\widehat{\overline{v}}_{b}^{(n)}(\kappa, 0^{+}) - \widehat{\overline{v}}_{b}^{(n)}(\kappa, 0^{-}) \right],$$

$$\widehat{\overline{M}}_{t}^{(n)}(\kappa) = -k_{t}^{(b)} \left[\widehat{\overline{\theta}}_{b}^{(n)}(\kappa, 0^{+}) - \widehat{\overline{\theta}}_{b}^{(n)}(\kappa, 0^{-}) \right],$$
(6)

where 0⁻ and 0⁺ are the left and the right of the connection spring, respectively.

Similarly, for the connection spring linking the pier with the left beam and the right beam, assuming that $k_l^{(t)}$, $k_l^{(s)}$, and $k_l^{(b)}$ are axial compression stiffness, shear stiffness, and rotational stiffness of left beam-pier connection spring, respectively, $k_r^{(t)}$, $k_r^{(s)}$, and $k_r^{(b)}$ for right beam-pier connection spring, as shown in Figure 3, the axial force $\widehat{\overline{(N)}}_l^{(n)}$ and $\widehat{\overline{N}}_r^{(n)}$, shear force $(\widehat{\overline{Q}}_l^{(n)}$ and $\widehat{\overline{Q}}_r^{(n)}$), and bending

moment $(\widehat{\overline{M}}_l^{(n)})$ and $(\widehat{\overline{M}}_r^{(n)})$ of the connection spring linking the pier with the left and right beams are expressed as

$$\widehat{\overline{N}}_{l}^{(n)}(\kappa) = k_{l}^{(t)} \left[\widehat{\overline{u}}_{d}^{(n)}(\kappa, 0^{+}) - \widehat{\overline{v}}_{b}^{(n)}(\kappa, 0^{-}) \right],
\widehat{\overline{Q}}_{l}^{(n)}(\kappa) = -k_{l}^{(s)} \left[\widehat{\overline{v}}_{d}^{(n)}(\kappa, 0^{+}) - \widehat{\overline{u}}_{b}^{(n)}(\kappa, 0^{-}) \right],$$

$$\widehat{\overline{M}}_{l}^{(n)}(\kappa) = -k_{l}^{(b)} \left[\widehat{\overline{\theta}}_{d}^{(n)}(\kappa, 0^{+}) + \widehat{\overline{\theta}}_{b}^{(n)}(\kappa, 0^{-}) \right],$$

$$\widehat{\overline{N}}_{r}^{(n)}(\kappa) = k_{r}^{(t)} \left[\widehat{\overline{u}}_{d}^{(n)}(\kappa, 0^{+}) - \widehat{\overline{v}}_{b}^{(n)}(\kappa, 0^{+}) \right],$$

$$\widehat{\overline{Q}}_{r}^{(n)}(\kappa) = -k_{r}^{(s)} \left[\widehat{\overline{v}}_{d}^{(n)}(\kappa, 0^{+}) - \widehat{\overline{u}}_{b}^{(n)}(\kappa, 0^{+}) \right],$$

$$\widehat{\overline{M}}_{r}^{(n)}(\kappa) = -k_{r}^{(b)} \left[\widehat{\overline{\theta}}_{d}^{(n)}(\kappa, 0^{+}) + \widehat{\overline{\theta}}_{b}^{(n)}(\kappa, 0^{+}) \right].$$
(8)

According to Figure 3, without considering the weight of the connection springs, the connection springs linking the pier with the left and right beams meet the following balance equations:

$$-\widehat{\overline{N}}_{b}^{(n)}(\kappa,0^{-}) + \widehat{\overline{N}}_{b}^{(n)}(\kappa,0^{+}) - \widehat{\overline{Q}}_{d}^{(n)}(\kappa,0^{+}) = 0,$$

$$-\widehat{\overline{Q}}_{b}^{(n)}(\kappa,0^{-}) + \widehat{\overline{Q}}_{b}^{(n)}(\kappa,0^{+}) + \widehat{\overline{N}}_{d}^{(n)}(\kappa,0^{+}) = 0,$$

$$\widehat{\overline{M}}_{b}^{(n)}(\kappa,0^{-}) - \widehat{\overline{M}}_{b}^{(n)}(\kappa,0^{+}) + \widehat{\overline{M}}_{d}^{(n)}(\kappa,0^{+}) = 0,$$

$$(9)$$

where $\widehat{\overline{N}}_b^{(n)}$, $\widehat{\overline{Q}}_b^{(n)}$, and $\widehat{\overline{M}}_b^{(n)}$ are the internal forces of horizontal beam; $\widehat{\overline{N}}_d^{(n)}$, $\widehat{\overline{Q}}_d^{(n)}$, and $\widehat{\overline{M}}_d^{(n)}$ are the internal forces of pier.

The connection spring of the left beam-pier and that of the right beam-pier meet the following balance equations:

$$\widehat{\overline{N}}_{b}^{(n)}(\kappa, 0^{-}) = \widehat{\overline{N}}_{t}^{(n)}(\kappa) - \widehat{\overline{Q}}_{l}^{(n)}(\kappa),$$

$$\widehat{\overline{Q}}_{b}^{(n)}(\kappa, 0^{-}) = \widehat{\overline{Q}}_{t}^{(n)}(\kappa) + \widehat{\overline{N}}_{l}^{(n)}(\kappa),$$

$$\widehat{\overline{M}}_{b}^{(n)}(\kappa, 0^{-}) = \widehat{\overline{M}}_{t}^{(n)}(\kappa) - \widehat{\overline{M}}_{l}^{(n)}(\kappa),$$

$$\widehat{\overline{N}}_{b}^{(n)}(\kappa, 0^{+}) = \widehat{\overline{N}}_{t}^{(n)}(\kappa) + \widehat{\overline{Q}}_{r}^{(n)}(\kappa),$$

$$\widehat{\overline{Q}}_{b}^{(n)}(\kappa, 0^{+}) = \widehat{\overline{Q}}_{t}^{(n)}(\kappa) - \widehat{\overline{N}}_{r}^{(n)}(\kappa),$$

$$\widehat{\overline{M}}_{b}^{(n)}(\kappa, 0^{+}) = \widehat{\overline{M}}_{t}^{(n)}(\kappa) + \widehat{\overline{M}}_{r}^{(n)}(\kappa).$$
(11)

3. Wave Field Solution of Saturated Soil under Rayleigh Surface Wave

3.1. Biot's Theory and Helmhotlz Vector Decomposition. According to the theory of saturated soil [12, 13], the constitutive equations of porous media have the following:

$$\sigma_{ij} = 2\mu\varepsilon_{ij} + \lambda\delta_{ij}e - \alpha\delta_{ij}p, \tag{12}$$

$$p_f = -\alpha Me + M\theta, \tag{13}$$

$$e = u_{i,i}$$

$$\theta = -w_{i,i},\tag{14}$$

$$w_i = \phi \left(U_i - u_i \right),\,$$

where σ_{ij} is the stress of soil, ε_{ij} is the strain tensor of soil, δ_{ij} is the Kronecker delta, u_i and U_i are, respectively, the average displacements of the soil and fluid, w_i is the penetration displacement of the fluid, p_f is the excess pore water pressure, λ and μ are lame constant, e is the volumetric strain of soil skeleton, ϑ is the fluid volume increment of unit porous media, α and M are Biot's parameters with regard to the saturated porous media compression, and ϕ is the porosity of porous media.

The motion equations of porous media can be expressed as

$$\mu u_{i,jj} + \left(\lambda + \alpha^2 M + \mu\right) u_{j,ji} + \alpha M w_{j,ji} = \rho \ddot{u}_i + \rho_f \ddot{w}_i, \quad (15)$$

$$\alpha M u_{i,ji} + M w_{i,ji} = \rho_f \ddot{u}_i + m \ddot{w}_i + b_p \dot{w}_i, \tag{16}$$

where ρ and ρ_f are, respectively, the density of porous media and the density of fluid, ρ_s is the density of soil skeleton, $m=a_\infty\rho_f/\phi$, a_∞ is the bending coefficient of porous media, $b_p=\eta/k$, b_p is the interaction force between soil skeleton and fluid, η is the viscosity coefficient of porous medium, and k is the dynamic permeability coefficient of porous medium.

Based on the method of Helmhotlz vector decomposition and Fourier transform, the soil displacement in frequency domain has the following form:

$$\overline{u}_i = \overline{\varphi}_{,i} + e_{ijk}\overline{\psi}_{k,j}, \tag{17}$$

where the superscript "–" expresses Fourier transform from t to ω , $\overline{\varphi}$ and $\overline{\psi}_k$ (k=1,2,3) are, respectively, scalar potential and vector potential of soil displacement in transform domain, e_{ijk} is the Ricci symbol, the vector potential $\overline{\psi}_k$ satisfies the regular condition, and

$$\overline{\varphi}_{i,i} = 0. \tag{18}$$

Because there are two kinds of P wave (P_1 wave and P_2 wave) in saturated soil, (17) can be expressed as

$$\overline{u}_{i} = \overline{\varphi}_{,i} + e_{ijk}\overline{\psi}_{k,j} = \overline{\varphi}_{f,i} + \overline{\varphi}_{s,i} + e_{ijk}\overline{\psi}_{k,j},$$
 (19)

where $\overline{\varphi}_f$ and $\overline{\varphi}_s$ are scalar potential of P₁ wave and P₂ wave. According to the analysis of Bonnet [16, 17], the pore pressure is as follows:

$$\overline{p}_f = A_f \overline{\varphi}_{f,ii} + A_s \overline{\varphi}_{s,ii}, \qquad (20)$$

where A_f and A_s are constants determined by Biot's control equation.

Equations (13), (16), (19), and (20) lead to the following equation:

$$\begin{split} & \left[\left(\lambda + 2\mu - \beta_2 A_f \right) \overline{\varphi}_{f,jj} + \beta_3 \overline{\varphi}_f \right]_{,i} \\ & \quad + \left[\left(\lambda + 2\mu - \beta_2 A_s \right) \overline{\varphi}_{s,jj} + \beta_3 \overline{\varphi}_s \right]_{,i} \\ & \quad + e_{iml} \left[\mu \overline{\psi}_{l,jj} + \beta_3 \overline{\psi}_l \right]_{,m} = 0. \end{split} \tag{21}$$

Equation (21) has the following expressions:

$$(\lambda + 2\mu - \beta_2 A_f) \overline{\varphi}_{f,jj} + \beta_3 \overline{\varphi}_f = 0,$$

$$(\lambda + 2\mu - \beta_2 A_s) \overline{\varphi}_{s,jj} + \beta_3 \overline{\varphi}_s = 0,$$

$$\mu \overline{\psi}_{i,jj} + \beta_3 \overline{\psi} = 0,$$
(22)

where $\beta_1 = m\omega^2 - ib_p\omega$, $\beta_2 = \alpha - \rho_f\omega^2/\beta_1$, and $\beta_3 = \rho\omega^2 - \rho_f^2\omega^4/\beta_1$.

Equations (16), (17), (18), and (19) lead to the following equation:

$$\overline{p}_{f,ii} + \frac{\beta_1}{M} \overline{p}_f + \left(\alpha \beta_1 - \rho_f \omega^2\right) \overline{u}_{i,i} = 0.$$
 (23)

Equation (24) can be reduced by substituting (19) and (20) into (22):

$$\begin{split} \left[A_{f} \overline{\varphi}_{f,ii} + \left(\beta_{5} A_{f} - \beta_{4} \right) \overline{\varphi}_{f} \right]_{,jj} \\ + \left[A_{s} \overline{\varphi}_{s,ii} + \left(\beta_{5} A_{f} - \beta_{4} \right) \overline{\varphi}_{s} \right]_{,jj} = 0. \end{split} \tag{24}$$

According to equation (24), the following equations should be satisfied.

$$A_{f}\overline{\varphi}_{f,ii} + (\beta_{5}A_{f} - \beta_{4})\overline{\varphi}_{f} = 0,$$

$$A_{s}\overline{\varphi}_{s,ii} + (\beta_{5}A_{s} - \beta_{4})\overline{\varphi}_{s} = 0,$$
(25)

where $\beta_4 = \rho_f \omega^2 - \alpha \beta_1$ and $\beta_5 = \beta_1 / M$.

Equation (26) can be obtained by the integration of (22) and (25):

$$A_{f,s}^{2} + \frac{\beta_{2} - (\lambda + 2\mu)\beta_{4} - \beta_{1}\beta_{3}}{\beta_{1}\beta_{4}} A_{f,s} + \frac{(\lambda + 2\mu)\beta_{3}}{\beta_{1}\beta_{4}}$$
(26)

Then, (22) leads to the following Helmholtz equations:

$$\nabla^2 \overline{\varphi}_f + k_f^2 \overline{\varphi}_f = 0,$$

$$\nabla^2 \overline{\varphi}_s + k_s^2 \overline{\varphi}_s = 0,$$

$$\nabla^2 \overline{\psi} + k_s^2 \overline{\psi} = 0,$$
(27)

where $k_f^2 = (\beta_5 A_f - \beta_4)/A_f$, $k_s^2 = (\beta_5 A_s - \beta_4)/A_s$, $k_t^2 = \beta_3/\mu$, k_f , k_s , and k_t are complex wave numbers of P_1 wave, P_2 wave, and S wave, and $\overline{\psi}$ is vector potential of shear wave. In order to guarantee the attenuation of the body waves, $\text{Im}(k_f)$, $\text{Im}(k_s)$, and $\text{Im}(k_t)$ should be nonpositive. Because the speed of P_1 wave is larger than that of P_2 wave, inequality $\text{Re}(k_f) \leq \text{Re}(k_s)$ should always hold. Detailed derivation of above poroelastic model can be found in Lu et al. [14].

3.2. The Wave Field Solution of Saturated Soil under Rayleigh Surface Wave. Assuming that the Rayleigh wave is two-dimensional inhomogeneous plane wave, the two-dimensional Helmholtz equations of saturated soil wave functions in the frequency domain can be obtained by the decoupling control equations for Biot's theory in Cartesian coordinate. Consider

$$\nabla_{\perp}^{2} \overline{\varphi}_{f} + k_{f}^{2} \overline{\varphi}_{f} = 0,$$

$$\nabla_{\perp}^{2} \overline{\varphi}_{s} + k_{s}^{2} \overline{\varphi}_{s} = 0,$$

$$\nabla_{\perp}^{2} \overline{\psi} + k_{t}^{2} \overline{\psi} = 0,$$
(28)

where φ_f , φ_s , and $\overline{\psi}$ are the potential functions for P_1 wave, P_2 wave, and S wave of the saturated half space and $\nabla_{\perp}^2 = \partial^2/\partial x^2 + \partial^2/\partial^2 z$ is the Laplacian operator in the two-dimensional Cartesian coordinate system. The displacements and the pore pressure of the pore fluid can be expressed as

$$u_{x} = \frac{\partial \overline{\varphi}_{f}}{\partial x} + \frac{\partial \overline{\varphi}_{s}}{\partial x} + \frac{\partial \overline{\psi}}{\partial z},$$

$$u_{z} = \frac{\partial \overline{\varphi}_{f}}{\partial z} + \frac{\partial \overline{\varphi}_{s}}{\partial z} + \frac{\partial \overline{\psi}}{\partial x},$$

$$p_{f} = -A_{f}k_{f}^{2}\overline{\varphi}_{f} - A_{s}k_{s}^{2}\overline{\varphi}_{s}.$$
(29)

Rayleigh surface wave is determined by the following potential function:

$$\overline{\varphi}_{f}(x,z)e^{i\omega t} = A_{f}^{R} \left[e^{-ik_{r}x - ik_{f}n_{f}z}\right]e^{i\omega t},$$

$$\overline{\varphi}_{s}(x,z)e^{i\omega t} = A_{s}^{R} \left[e^{-ik_{r}x - ik_{s}n_{s}z}\right]e^{i\omega t},$$

$$\overline{\psi}(x,z)e^{i\omega t} = A_{t}^{R} \left[e^{-ik_{r}x - ik_{t}n_{t}z}\right]e^{i\omega t},$$
(30)

where A_f^R , A_s^R , and A_t^R are, respectively, the wave function amplitude of P_1 wave, P_2 wave, and S wave in Rayleigh wave and k_f , k_s , and k_t are the complex wave number of the Rayleigh surface waves, respectively; n_f , n_s , and n_t are the complex direction cosines of P_1 wave, P_2 wave, and S wave.

The following can be reduced by substituting the potential function into the two-dimensional Helmholtz equations:

$$n_f^2 k_f^2 + k_r^2 = k_f^2,$$

$$n_s^2 k_s^2 + k_r^2 = k_s^2,$$

$$n_t^2 k_t^2 + k_r^2 = k_t^2.$$
(31)

The following boundary conditions for a fully permeable surface are as follows:

$$\overline{\sigma}_{zz}(x,0) = 0,$$

$$\overline{\sigma}_{zx}(x,0) = 0,$$

$$\overline{p}_f(x,0) = 0.$$
(32)

The displacements, stresses, and pore pressure can be obtained based on the equations (28)–(32). Three equations

about A_f^R , A_s^R and A_t^R can be reduced based on the boundary conditions. The complex Rayleigh equation of saturated soil can be calculated by the conditions of the equation coefficient ranks as the zero. In order to guarantee the attenuation of the body waves, $\operatorname{Im}(k_r)$ should be nonpositive; the roots of n_f , n_s , and n_t satisfy $\operatorname{Re}(k_i n_i) \geq 0$ and $\operatorname{Im}(k_i n_i) \leq 0$, i = f, s, t.

3.3. Floquet Transform Method. The displacement amplitude at the bottom of different piers of the periodic structure varies with spatial locations in the action of Rayleigh waves, and the displacive phase transition between adjacent piers is uncertain. That is to say, the dynamic response of periodic viaduct structure is caused by a series of nonuniform waves; it is impossible to employ the periodic condition of the viaduct to simplify its dynamic analysis directly. Therefore, the Floquet transform method is introduced to convert the spatial nonuniform waves to harmonic waves in wave number domain. For periodic viaduct in-plane vibration, considering the lattice vector cycle of adjacent lattice points in the one-dimensional direction to be L, the lattice vector R can be expressed as $\mathbf{R} = nLe$ according to literature [18], where *e* is basis vector of the one-dimensional direction. If $f(\mathbf{R}) = f(nL)$ is the discrete functions in spatial domain of one-dimensional vector, the Floquet transform and inverse transform can be defined as follows [19]:

$$F(f(nL)) = \hat{f}(\kappa) = \sum_{n=-\infty}^{+\infty} f(nL) e^{i\kappa nL},$$

$$f(nL) = \frac{L}{2\pi} \int_{-\pi/L}^{\pi/L} \hat{f}(\kappa) e^{-i\kappa nL} d\kappa,$$
(33)

where κ is the wave number of lattice waves and the superscript "~" means wave number domain.

Thus, the Floquet transform of discrete spatial sequence function f[(m+n)L] can be expressed as

$$F(f[(m+n)L]) = \widehat{f}^{(n)}(\kappa)$$

$$= \sum_{m=-\infty}^{+\infty} f[(m+n)L] e^{i\kappa mL}$$

$$= e^{-i\kappa nL} \widehat{f}(\kappa).$$
(34)

Considering the solution of the viaduct in wave field due to Rayleigh wave, the Fourier transform between time t and frequency ω is defined as follows:

$$\overline{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt,$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \overline{f}(\omega) e^{i\omega t} d\omega,$$
(35)

where the superscript "-" indicates the function is to be in frequency domain.

With (34) and (35), the dynamic response in time domain can be transformed into harmonic wave with amplitude $\hat{\overline{f}}(k)$ in frequency wave number domain.

4. Dynamic Response of In-Plane Vibration of the Periodic Viaduct

4.1. Dynamic Response in Frequency Wave Number Domain. The displacement vector $\widehat{\overline{\mathbf{q}}}_d^{(n)}(\kappa,L_d)$ at the bottom of viaduct pier due to Rayleigh waves can be deduced based on the Floquet transform of discrete spatial sequence function. Through the transfer matrix of pier, the displacement vector $\widehat{\overline{\mathbf{q}}}_d^{(n)}(\kappa,0^+)$ and the internal force vector at the top of pier can be derived as follows:

$$\widehat{\overline{\mathbf{q}}}_{d}^{(n)}\left(\kappa,0^{+}\right) = \mathbf{C}_{d}\widehat{\overline{\mathbf{f}}}_{d}^{(n)}\left(\kappa,0^{+}\right) + \mathbf{T}_{qq}^{(d)-1}\left(L_{d}\right)\widehat{\overline{\mathbf{q}}}_{d}^{(n)}\left(\kappa,L_{d}\right),$$
(36)

where
$$\mathbf{C}_d = -\mathbf{T}_{qq}^{(d)-1}(L_d)\mathbf{T}_{qf}^{(d)}(L_d)$$
, $\mathbf{T}^{(d)}(L_d) = \begin{bmatrix} \mathbf{T}_{qq}^{(d)}(L_d) & \mathbf{T}_{qf}^{(d)}(L_d) \\ \mathbf{T}_{fq}^{(d)}(L_d) & \mathbf{T}_{ff}^{(d)}(L_d) \end{bmatrix}$, and $\mathbf{T}_{qq}^{(d)}(L_d)$ and $\mathbf{T}_{qf}^{(d)}(L_d)$ are the submatrix of the pier transfer matrix.

Equation (10) is expressed in matrix form as

$$\widehat{\overline{\mathbf{f}}}_{d}^{(n)}\left(\kappa,0^{+}\right) = \mathbf{E}_{l}^{(a)}\widehat{\overline{\mathbf{f}}}_{b}^{(n)}\left(\kappa,0^{-}\right) + \mathbf{E}_{r}^{(a)}\widehat{\overline{\mathbf{f}}}_{b}^{(n)}\left(\kappa,0^{+}\right),\tag{37}$$

where
$$\mathbf{E}_{l}^{(a)} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$
 and $\mathbf{E}_{r}^{(a)} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

The following equation can be obtained by substituting

The following equation can be obtained by substituting (37) into (36):

$$\widehat{\overline{\mathbf{q}}}_{d}^{(n)}\left(\kappa,0^{+}\right) = E_{l}^{(b)}\widehat{\overline{\mathbf{f}}}_{b}^{(n)}\left(\kappa,0^{-}\right) + E_{r}^{(b)}\widehat{\overline{\mathbf{f}}}_{b}^{(n)}\left(\kappa,0^{+}\right) + \mathbf{T}_{qq}^{(d)-1}\left(L_{d}\right)\widehat{\overline{\mathbf{q}}}_{d}^{(n)}\left(\kappa,L_{d}\right),$$
(38)

where $\mathbf{E}_l^{(b)} = \mathbf{C}_d \mathbf{E}_l^{(a)}$ and $\mathbf{E}_r^{(b)} = \mathbf{C}_d \mathbf{E}_r^{(a)}$.

The internal force vectors of the left and the right beams in the *n*th span unit are deduced and are obtained as follows based on (6), (7), and (10) and (6), (8), and (11):

$$\widehat{\overline{\mathbf{f}}}_{b}^{(n)}(\kappa,0^{-}) = \mathbf{J}_{ll}\widehat{\overline{\mathbf{q}}}_{b}^{(n)}(\kappa,0^{-}) + \mathbf{J}_{lr}\widehat{\overline{\mathbf{q}}}_{b}^{(n)}(\kappa,0^{+})
+ \mathbf{J}_{ld}\widehat{\overline{\mathbf{q}}}_{d}^{(n)}(\kappa,0^{+}),$$

$$\widehat{\overline{\mathbf{f}}}_{b}^{(n)}(\kappa,0^{+}) = \mathbf{J}_{rl}\widehat{\overline{\mathbf{q}}}_{b}^{(n)}(\kappa,0^{-}) + \mathbf{J}_{rr}\widehat{\overline{\mathbf{q}}}_{b}^{(n)}(\kappa,0^{+})
+ \mathbf{J}_{rd}\widehat{\overline{\mathbf{q}}}_{d}^{(n)}(\kappa,0^{+}),$$
(39)

where
$$\mathbf{J}_{ll} = \begin{bmatrix} -(k_t^{(t)} + k_l^{(s)}) & 0 & 0 \\ 0 & -(k_t^{(s)} + k_l^{(t)}) & 0 \\ 0 & 0 & k_t^{(b)} + k_l^{(b)} \end{bmatrix}, \mathbf{J}_{lr} = \begin{bmatrix} k_t^{(t)} & 0 & 0 \\ 0 & k_t^{(s)} & 0 \\ 0 & 0 & -k_t^{(b)} \end{bmatrix},$$

$$\mathbf{J}_{rr} = \begin{bmatrix} k_t^{(t)} + k_r^{(s)} & 0 & 0 \\ 0 & k_t^{(s)} + k_r^{(t)} & 0 \\ 0 & 0 & -(k_t^{(b)} + k_r^{(b)}) \end{bmatrix}, \mathbf{J}_{ld} = \begin{bmatrix} 0 & k_l^{(s)} & 0 \\ k_l^{(t)} & 0 & 0 \\ 0 & 0 & k_l^{(b)} \end{bmatrix}, \mathbf{J}_{rl} = \begin{bmatrix} -k_t^{(t)} & 0 & 0 \\ 0 & -k_t^{(s)} & 0 \\ 0 & 0 & k_t^{(b)} \end{bmatrix}, \text{ and } \mathbf{J}_{rd} = \begin{bmatrix} 0 & -k_r^{(s)} & 0 \\ -k_r^{(t)} & 0 & 0 \\ 0 & 0 & -k_r^{(b)} \end{bmatrix}.$$

Substituting (38) into (39), the state vector of the left side and right side of horizontal beam-beam spring can be expressed as

$$\left\{ \frac{\widehat{\mathbf{q}}_{b}^{(n)}(\kappa, 0^{+})}{\widehat{\mathbf{f}}_{b}^{(n)}(\kappa, 0^{+})} \right\} = \mathbf{S}_{J} \left\{ \frac{\widehat{\mathbf{q}}_{b}^{(n)}(\kappa, 0^{-})}{\widehat{\mathbf{f}}_{b}^{(n)}(\kappa, 0^{-})} \right\} + \mathbf{A}^{-1} \widehat{\overline{\mathbf{R}}}^{(n)}(\kappa), \quad (40)$$

where
$$\mathbf{A} = \begin{bmatrix} \mathbf{J}_{lr} & \mathbf{J}_{ld} \mathbf{E}_{r}^{(b)} \\ \mathbf{J}_{rr} & \mathbf{J}_{rd} \mathbf{E}_{r}^{(b)-1} \end{bmatrix}$$
, $\mathbf{B} = \begin{bmatrix} -\mathbf{J}_{ll} & \mathbf{I} - \mathbf{J}_{ld} \mathbf{E}_{l}^{(b)} \\ -\mathbf{J}_{rl} & -\mathbf{J}_{rd} \mathbf{E}_{l}^{(b)} \end{bmatrix}$, $\widehat{\overline{\mathbf{R}}}^{(n)}(\kappa) = \begin{bmatrix} -\mathbf{J}_{ld} \mathbf{T}_{qq}^{(d)-1}(L_{d}) \widehat{\overline{\mathbf{q}}}_{d}^{(n)}(\kappa, L_{d}) \\ -\mathbf{J}_{rd} \mathbf{T}_{qq}^{(d)-1}(L_{d}) \widehat{\overline{\mathbf{q}}}_{d}^{(n)}(\kappa, L_{d}) \end{bmatrix}$, and $\mathbf{S}_{J} = \mathbf{A}^{-1} B$.

The state vector of the left beam and right beam can be obtained by the solution of (40) and horizontal beam transfer matrix $T^{(b)}(L_b)$:

$$\widehat{\overline{\psi}}_{b}^{(n)}\left(\kappa, \frac{L_{b}}{2}\right)
= \mathbf{T}^{(b)}\left(\frac{L_{b}}{2}\right)\mathbf{S}_{J}\mathbf{T}^{(b)}\left(\frac{L_{b}}{2}\right)\widehat{\overline{\psi}}_{b}^{(n)}\left(\kappa, -\frac{L_{b}}{2}\right)
+ \mathbf{T}^{(b)}\left(\frac{L_{b}}{2}\right)\mathbf{A}^{-1}\widehat{\overline{\mathbf{R}}}^{(n)}\left(\kappa\right).$$
(41)

According to Bloch's theory [19], the state vector of periodic structure satisfies the following equation in wave number domain:

$$\widehat{\overline{\psi}}_{b}^{(n)}\left(\kappa, \frac{L_{b}}{2}\right) = e^{-i\kappa L} \widehat{\overline{\psi}}_{b}^{(n)}\left(\kappa, -\frac{L_{b}}{2}\right). \tag{42}$$

Using (41) and (42), the state vector of the left beam in the periodic element can be obtained:

$$\left[\mathbf{T}^{(b)}\left(\frac{L_b}{2}\right)\mathbf{S}_{J}\mathbf{T}^{(b)}\left(\frac{L_b}{2}\right) - e^{-i\kappa L}\mathbf{I}_{6\times 6}\right]\widehat{\overline{\psi}}_{b}^{(n)}\left(\kappa, -\frac{L_b}{2}\right)
= \mathbf{T}^{(b)}\left(\frac{L_b}{2}\right)\mathbf{A}^{-1}\widehat{\overline{\mathbf{R}}}^{(n)}(\kappa).$$
(43)

Excluding the vibration of piers caused by the Rayleigh waves, $\widehat{\overline{R}}^{(n)}(\kappa)$ is zero in (43); thus, the characteristic equation of in-plane vibration for the periodic viaduct can be obtained:

$$\left[\mathbf{T}^{(b)}\left(\frac{L_b}{2}\right)\mathbf{S}_J\mathbf{T}^{(b)}\left(\frac{L_b}{2}\right) - e^{-i\kappa_{\alpha}L}\mathbf{I}_{6\times6}\right]\widehat{\overline{\psi}}_b^{(n)}\left(\kappa, -\frac{L_b}{2}\right)$$

$$= 0,$$
(44)

where κ_{α} is complex wave number of wave propagation in periodic viaduct of which the real part means the phase transition of lattice wave and the imaginary part indicates wave decay in periodic viaduct.

4.2. Dynamic Response in Spatial Domain. The state vector of the left beam can be gotten in the frequency wave number domain based on (43). The solution $\overline{\psi}_h^{(n)}(nL_h, -L_h/2)$ in the

TABLE 1: The parameters of the piers and beams.

Young's modulus of pier E_d (Pa)	2.8×10^{10}
Poisson's ratio v_d	0.3
Pier density ρ_d (kg/m ³)	3.0×10^{3}
Pier height L_d (m)	7.5
Pier radius R_d (m)	0.3
Young's modulus of beam E_b (Pa)	2.8×10^{10}
Beam density ρ_b (kg/m ³)	3.5×10^{3}
Span L_b (m)	10.0
Cross-sectional height of beam h_b (m)	0.3
Cross-sectional width of beam w_b (m)	2.0

TABLE 2: The spring stiffness of beam-beam and beam-pier.

	k_{tt} (N/m)	5.0×10^{8}
Spring stiffness of beam-beam	k_{ts} (N/m)	5.0×10^{8}
	k_{tb} (N·m/rad)	1.0×10^{8}
	k_{lt} (N/m)	2.0×10^{8}
Spring stiffness of left beam-pier	k_{ls} (N/m)	2.0×10^{8}
	k_{lb} (N·m/rad)	0.5×10^{8}
	k_{rt} (N/m)	2.0×10^{8}
Spring stiffness of right beam-pier	k_{rs} (N/m)	2.0×10^{8}
	k_{rb} (N·m/rad)	0.5×10^{8}

frequency spatial domain can be obtained by means of the inverse Floquet transform. Consider

$$\overline{\psi}_{b}^{(n)}\left(nL_{b}, -\frac{L_{b}}{2}\right) = \frac{L_{b}}{2\pi} \int_{-\pi/L_{b}}^{\pi/L_{b}} \widehat{\overline{\psi}}_{b}^{(n)}\left(\kappa, -\frac{L_{b}}{2}\right) d\kappa$$

$$= \frac{L_{b}}{2\pi} \int_{-\pi/L_{b}}^{\pi/L_{b}} \widehat{\overline{\psi}}_{b}\left(\kappa, -\frac{L_{b}}{2}\right) e^{-in\kappa L_{b}} d\kappa.$$
(45)

5. Numerical Results

As for the periodic viaduct shown in Figure 1, considering the cross section of horizontal beam to be rectangular and that of pier to be circular, the parameters of the pier and beams given in Table 1 and the spring stiffness of beam-beam and beam-pier given in Table 2 are employed to study the dynamic response under Rayleigh surface wave.

In the paper, Rayleigh wave is adopted as the vibration source in saturated soil to simulate seismic wave field. The center coordinate (x_s , y_s , and z_s) of the unit circle load is (0.0 m, 20.0 m, and 10.0 m), the radius R is 0.5 m, and the load amplitude F_z is 1.0 N. The parameters of saturated soil are as follows: $\mu = 2.0 \times 10^7 \,\text{N/m}^2$, $a_\infty = 2.0$, $\lambda = 4.0 \times 10^7 \,\text{N/m}^2$, $\alpha = 0.97$, $b_p = 1.94 \times 10^6 \,\text{kg/m}^3$ ·s, $M = 2.4 \times 10^8 \,\text{N/m}^2$, $\phi = 0.4$, $\rho_s = 2.0 \times 10^3 \,\text{kg/m}^3$, and $\rho_f = 1.0 \times 10^3 \,\text{kg/m}^3$.

5.1. Energy Band Analysis of the Periodic Structure. The transfer matrix for each span of the periodic viaduct is a 6×6 matrix in (44). Therefore, there exist three kinds of characteristic waves corresponding to axial compression, transverse shear, and bending vibration, respectively, in the periodic

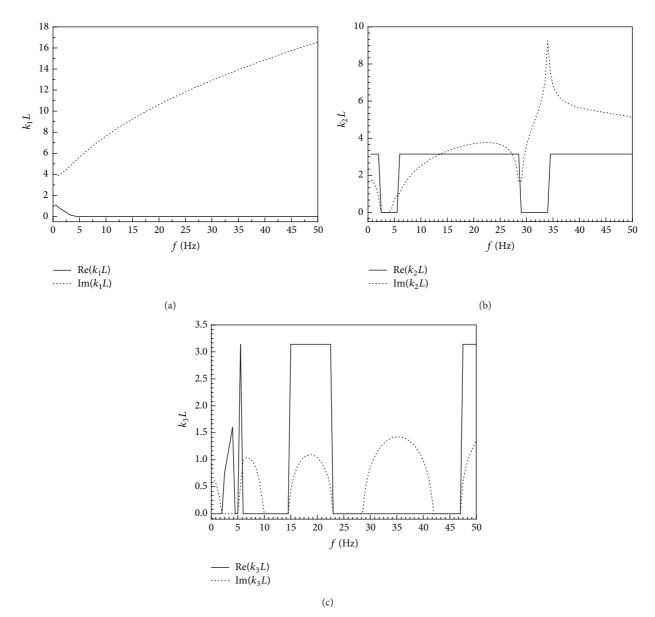


FIGURE 4: The energy bands for the characteristic waves of the periodic viaduct undergoing in-plane vibration. (a) The wave number for the first kind of wave. (b) The wave number for the second kind of wave. (c) The wave number for the third kind of wave.

viaduct for the in-plane vibration [19]; the real part represents the phase transformation of the wave propagation in the structure, and the imaginary part represents the attenuation of the propagating wave.

Figure 4 shows the energy bands of the three kinds of characteristic waves propagating in the periodic viaduct for in-plane vibration. The distribution of their energy bands varies with different frequencies. The imaginary part of the first kind and the second kind of waves is larger, which indicates the first kind and second kind of waves cannot propagate far in the periodic viaduct, and vibration waves decay rapidly. The imaginary part of the third kind of wave shows the alternating pass band and band gap at different frequencies and is smaller than that of above two kinds of waves, which indicates the wave mainly propagating in the viaduct vibration is the third kind.

5.2. Dynamic Response of the Periodic Structure under Nonuniform Load. Using Rayleigh surface wave as vibration source, according to the solution of literature [14], the displacement vector at the bottom of pier for the periodic viaduct can be obtained in the frequency domain. By means of the Floquet transform of (1), the displacement vector at the bottom of pier can be obtained in the frequency wave number domain, wherein the infinite sequence summation of (1) is expressed in finite terms with the sum number 101 in the paper. The inverse Floquet transform of (2) from the wave number domain to spatial domain can be realized by numerical integration method.

Figures 5 and 6 show dimensionless axial amplitude $(u_x\mu R/F_z)$ and tangential displacement amplitude $(u_z\mu R/F_z)$ of the horizontal beam end of the periodical viaduct when

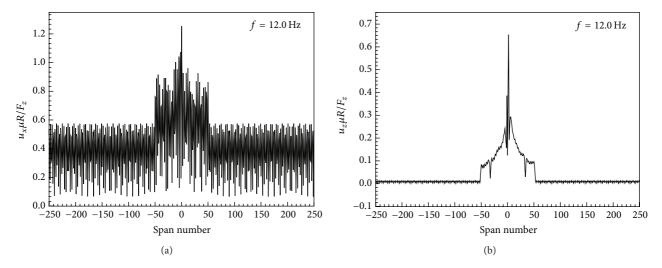


Figure 5: Dynamic response of the periodic viaduct when the frequency of the excitation source is equal to 12.0 Hz (pass band). (a) The amplitude of the longitudinal displacement. (b) The amplitude of the transverse displacement.

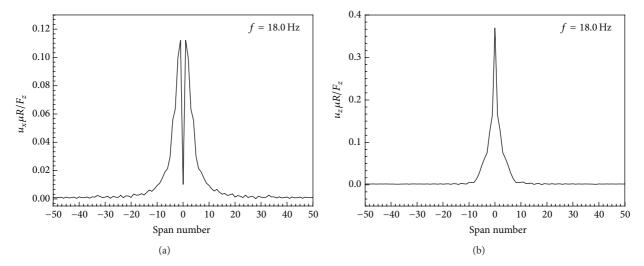


FIGURE 6: Dynamic response of the periodic viaduct when the frequency of the excitation source is equal to 12.0 Hz (stop band). (a) The amplitude of the longitudinal displacement. (b) The amplitude of the transverse displacement.

the frequency of vibration source is equal to 12.0 Hz and 18.0 Hz, respectively.

As shown in Figure 5(a), when the frequency of vibration source is equal to 12.0 Hz, the axial displacement of the beam at a distance of 100 spans (–50, 50) from the vibration source decays rapidly, while it decays slowly when the distance is over 100 spans. The main reason is the fact that the vibration of the viaduct is mainly caused by external vibration source within the range of 100 spans, but, when the distance to the vibration source is more than a certain range (outside of 100 spans), due to the load frequency of vibration source within the pass band of periodic viaduct energy band, the characteristic wave can propagate in periodic structure and cannot decay, which causes vibrations of structure far from the vibration source. However, as shown in Figure 5(b), the transverse displacement far away from the vibration source (outside of 100 spans) attenuates rapidly. Therefore, it can be

seen that the wave propagation with vibration source within the pass band of viaduct structure is mainly the third kind of wave.

As shown in Figure 6, when the frequency of vibration source is equal to 18.0 Hz, the dynamic response at the range of 20 spans (-10, 10) away from the excitation source is larger, while the longitudinal displacement and transverse displacement decay rapidly when the structure is far away from the excitation source. The main reason is the fact that the vibration attenuates rapidly and propagates in a short distance when the frequency of excitation source is in the range of band gap of periodic structure energy band.

6. Conclusion

In the paper, combining the Floquet transform and transfer matrix method, using the periodicity of viaduct structure, the in-plane vibration response of periodic viaduct on saturated soil subjected to Rayleigh surface wave excitation is analyzed. Firstly, the periodical structure model of beambeam and beam-pier connected by springs is established in the frequency wave number domain, and the transfer matrix of spring junctions, horizontal beams, and pier are deduced. Then the characteristic equation and the vibration control equations of viaduct structure are established. Finally, the solution of spatial domain is obtained by the inverse Floquet transform. Using numerical examples, the propagation characteristics of waves and the vibration characteristics of periodical viaduct subjected to Rayleigh wave are discussed. And the numerical results show the following:

- (1) There are three kinds of wave which existed in periodic viaduct. The first kind of wave decays rapidly, the second kind of wave propagates only in a limited frequency domain, and the third kind of wave propagates in the pass band and decays slowly.
- (2) The vibration characteristic of periodic structure far from the vibration source of Rayleigh wave is determined by the characteristic wave propagating in the structure. When the frequency of the vibration source is in the range of pass band, the characteristic wave in the structure will decay slowly and propagate farther. Therefore, there exists structure vibration far away from the vibration source.
- (3) When the frequency of the incident Rayleigh wave is in the range of band gap of periodic viaduct, the dynamic response of viaduct structure only exists in the vicinity of vibration source, and vibration waves decay rapidly, which cannot spread along the structure.

Conflict of Interests

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

Acknowledgments

This paper was supported by the National Natural Science Foundation (Project 51269021), the Jiangxi Province Science and Technology Project (20133ACB20006, 20114BAB216010, and 20122BBG70174), and the Jiangxi Provincial Education Department of Science and Technology Project (GJJ14755 and GJJ14756) in China.

References

- [1] M. M. Sigalas and E. N. Economou, "Elastic and acoustic wave band structure," *Journal of Sound and Vibration*, vol. 158, no. 2, pp. 377–382, 1992.
- [2] M. S. Kushwaha, P. Halevi, L. Dobrzynski, and B. Djafari-Rouhani, "Acoustic band structure of periodic elastic composites," *Physical Review Letters*, vol. 71, no. 13, pp. 2022–2025, 1993.
- [3] A. D. Kiureghian and A. Neuenhofer, "Response spectrum method for multi-support seismic excitations," *Earthquake*

- Engineering & Structural Dynamics, vol. 21, no. 8, pp. 713–740, 1992.
- [4] M. P. Singh and G. O. Maldonado, "An improved response spectrum method for calculating seismic design response: classically damped structures," *Earthquake Engineering & Structural Dynamics*, vol. 20, no. 7, pp. 621–635, 1991.
- [5] L. Chen, Seismic response analysis on a long span hybrid girder continuous rigid-frame bridge [M.S. thesis], Central South University, Changsha, China, 2013.
- [6] A. S. Nazmy and A. M. Abdel-Ghaffar, "Non-linear earthquakeresponse analysis of long-span cable-stayed bridges: theory," *Earthquake Engineering & Structural Dynamics*, vol. 19, no. 1, pp. 45–62, 1990.
- [7] J. T. Du, Seismic analysis of long-span continuous rigid frame bridge [M.S. thesis], Chang'an University, Xi'an, China, 2013.
- [8] T. J. Ingham, S. Rodriguez, and M. Nader, "Nonlinear analysis of the Vincent Thomas Bridge for seismic retrofit," *Computers and Structures*, vol. 64, no. 5-6, pp. 1221–1238, 1997.
- [9] J. H. Lin, Y. H. Zhang, Q. S. Li, and F. W. Williams, "Seismic spatial effects for long-span bridges, using the pseudo excitation method," *Engineering Structures*, vol. 26, no. 9, pp. 1207–1216, 2004.
- [10] J. J. Wang, Study on the dynamic response of curved continuous bridges under spatial seismic excitation [M.S. thesis], Beijing University of Technology, Beijing, China, 2013.
- [11] K. Soyluk, "Comparison of random vibration methods for multi-support seismic excitation analysis of long-span bridges," *Engineering Structures*, vol. 26, no. 11, pp. 1573–1583, 2004.
- [12] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range," *Journal of the Acoustical Society of America*, vol. 28, no. 2, pp. 179–191, 1956.
- [13] M. A. Biot, "Mechanics of deformation and acoustic propagation in porous media," *Journal of Applied Physics*, vol. 33, no. 4, pp. 1482–1498, 1962.
- [14] J.-F. Lu, D.-S. Jeng, and W.-D. Nie, "Dynamic response of a pile embedded in a porous medium subjected to plane SH waves," *Computers and Geotechnics*, vol. 33, no. 8, pp. 404–418, 2006.
- [15] J.-F. Lu and H.-Y. Yuan, "The sequence Fourier transform method for the analysis of a periodic viaduct subjected to nonuniform seismic waves," *Acta Mechanica*, vol. 224, no. 9, pp. 2143–2168, 2013.
- [16] G. Bonnet, "Basic singular solutions for a poroelastic medium in the dynamic range," *Journal of the Acoustical Society of America*, vol. 82, no. 5, pp. 1758–1762, 1987.
- [17] B. Xu, Dynamic Responses of the Saturated Poro-Elastic Half-Space Due to a Moving Loads and Analysis of Vibration Isolation Effectiveness with Pile Rows, Shanghai Jiao Tong University, 2009.
- [18] J. Wen, D. Yu, G. Wang, H. Zhao, and Y. Liu, "Elastic wave band gaps in flexural vibrations of straight beams," *Chinese Journal of Mechanical Engineering*, vol. 41, no. 4, pp. 1–6, 2005.
- [19] C. Kittel, Introduction to Solid State Physics, Wiley, New York, NY, USA, 1996.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 537362, 10 pages http://dx.doi.org/10.1155/2015/537362

Research Article

Quantitative Risk Analysis of Offshore Fire and Explosion Based on the Analysis of Human and Organizational Factors

Yan Fu Wang, 1 Yu Lian Li, 1 Biao Zhang, 1 Pei Na Yan, 1 and Li Zhang 2

¹Department of Safety Science and Engineering, China University of Petroleum, Qingdao 266580, China ²Jiangsu Academy of Safety Science and Technology, Nanjing 210042, China

Correspondence should be addressed to Yan Fu Wang; 66923203@qq.com

Received 12 May 2015; Accepted 30 June 2015

Academic Editor: Xiaobo Qu

Copyright © 2015 Yan Fu Wang 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.

A dynamic risk analysis model of offshore fire and explosion is proposed in this paper. It considers the effect of human and organizational factors in a more explicit way than current traditional risk analysis methods. This paper begins with exploring the recent advances on offshore fire and explosion risk analysis theories, followed by briefly introducing the research techniques employed in the proposed hybrid causal logic model which consists of event tree, fault tree, Bayesian network, and system dynamics. Thereafter, it proposes a quantitative risk analysis framework. At last, the applicability of this model to the offshore platform is also discussed. It aims to provide guideline for risk analysis of offshore fire and explosion.

1. Introduction

Over the past two decades, a number of serious accidents including the Piper Alpha accident have attracted public concerns over offshore safety and reliability. Offshore platform is in a harsh environment due to highly concentrated equipment, a large amount of explosive substances, oil/gas pipelines, and flange leak sources. Hydrocarbon fires and explosions are extremely hazardous on offshore platforms. The fire and explosion will not only result in significant casualties and economic losses, but also cause serious pollution and damage to surrounding environment and coastal marine ecosystems [1]. The high cost of offshore platforms, fire/explosion severity, and complexity of marine environment determine the necessity and difficulty of fire/explosion risk analysis [2]. How to propose efficient safety measures to prevent the escalation of such accidents requires comprehensive knowledge of accident-related phenomena, as well as the tools of adequate risk analysis [3].

Historical statistics show that the majority of offshore fire and explosion accidents are caused by human and organizational error (HOE) [4]. As a result, Norway and UK offshore legislation and guidelines require that HOE analysis should be included in quantitative risk analysis. However, the mainstream offshore fire/explosion risk analysis methods focus more on probability of equipment and structures failure. Recently, some attempt to build human reliability analysis models used in quantitative risk analysis chose to leverage expert information due to lack of HOE data. There have been repeated calls to expand the technical basis of human reliability analysis by systematically integrating information from different domains. Thus refining and improving analytic methods of human reliability analysis to more accurately simulate and quantify the impact of HOE have become a matter of urgency.

This paper is organized as follows. Section 2 analyzes the recent advances on offshore fire and explosion risk assessment. Section 3 briefly introduces event tree, fault tree, Bayesian network (BN), and system dynamics. Then, in Section 4, a dynamic quantitative risk analysis framework is proposed and its applicability for the offshore platform is also discussed, followed by conclusions in Section 5.

2. Recent Advances of Offshore Fire and Explosion Risk Analysis

Quantitative risk analysis involves four main steps: hazard identification, consequence assessment, probability calculation, and finally risk quantification [5–8]. Three kinds of

researches on offshore fire/explosion risk analysis are carried out worldwide: the first one is using statistical methods to predict fire/explosion risk based on historical data [9]; the second one is carrying out risk analysis using commercial software [10]; the third one is integrating new theory with traditional risk assessment methods [11].

In this section, a literature review is carried out to identify existing methods for risk assessment of offshore fire and explosion.

2.1. Consequence Analysis of Offshore Fire and Explosion. Some offshore fire/explosion risk assessment studies focus on consequences and impacts analysis: Krueger and Smith demonstrate how a scenario-based approach of fire risk assessment can be effectively applied to calculate potential impacts of credible fire scenarios on the platform process equipment, structural members, and safety systems [12]. Pula et al. considered fire consequence modeling as a suite of submodels such as individual fire models, radiation model, overpressure model, smoke and toxicity models, and human impact models. This comprehensive suite of models was then revised and its performance is compared with the ones used in a commercial software package for offshore risk assessment [13, 14]. Suardin et al. adopted a grid-based approach for fire and explosion assessment to enable better consequence/impact modeling and enhanced on-site ignition model. This approach features built-in calculations for jet and pool fire size estimation for gas/liquid releases and the ability to perform quantitative risk analysis to specify the personnel and equipment risk [15]. The Joint Industry Project carries out a series of experimental tests to evaluate the load characteristics of steel and concrete tubular members under jet fire to investigate the jet fire load characteristics [16] and to examine the effect of wind on the thermal diffusion characteristics of floating production storage and offloading [17]. Moreover, a modeling technique for computational fluid dynamics simulations of hydrocarbon explosions and fire is developed [18]. These insights developed will be very useful for the fire engineering of floating production storage and offloading topsides.

2.2. Frequency Analysis of Offshore Fire and Explosion. There are also few offshore fire/explosion risk assessment studies focused on frequency and probability analysis: Ronza et al. [19] analyzed the correlations of ignition probability and the mass flow rate by analyzing tens of thousands of records of hydrocarbon spills by HMIRS (a database about hazardous materials incident reporting system). Moosemiller [20] develops the algorithms based on a variety of design, operating, and environmental conditions to calculate probabilities of immediate ignition and delayed ignition resulting in explosion. On the other hand, Vinnem proposes failure models for hydrocarbon leaks during maintenance work in process plants on offshore petroleum installations based on seventy hydrocarbon leaks accidents [21]. Because major accidents are always of low frequency, data collected from them are not sufficient. As a result, it is very difficult to use conventional statistical methods to analyze them. In order

to improve the situation, a methodology has been proposed based on hierarchical Bayesian analysis and accident precursor data to do risk analysis of major accidents [22].

Based on the aforementioned works on leak frequency and ignition modeling, many researchers have proposed new methodologies for frequency analysis of offshore fire and explosion. The Joint Industry Project introduces a number of procedures [9] and reviews the state-of-the-art technologies [23] for quantitative assessment and management of fire and gas explosion risks in offshore installations focusing on defining the frequency of fires and explosions in offshore installations. Aside from the application of the fault tree, event tree, and BN [24-27], a methodology [28] based on Bow-tie and real time predictive models is proposed to conduct dynamic risk assessment of the drilling operations. The application of the Bow-tie model is used to analyze the potential accident scenarios, their causes, and the associated consequences. Real time predictive models for the failure probabilities of key barriers are developed to conduct dynamic risk assessment of the drilling operations. Vinnem et al. [29] proposed the Risk_OMT model combining the traditional quantitative risk analysis models and HOE analysis through Risk Influence Factor model, for quantitative risk analysis of platform specific hydrocarbon release frequency. The risk of fire and explosion in floating production storage and offloading is quantitatively assessed by Dan et al., in which the consequence is simulated by PHAST and the frequency is analyzed based on the analysis of historical data using event tree model [30].

The current generation of quantitative risk analysis does not include the quantitative impacts of HOE on the safety performance of equipment and personnel. There are a number of technical challenges in developing an offshore fire and explosion risk analysis model based on the effect analysis of HOE.

2.3. Effect Analysis of Human and Organizational Error to Offshore Fire and Explosion Risk. With the extension of human reliability research field from human-machine systems to human inherent factors (psychology, emotion, and behavior), increasing attention has been paid to the relationship between HOE and offshore fire/explosion risk. The relevant references are shown as follows: an extensive survey [31, 32] is followed to identify the relationship between HOE and hydrocarbon leaks on Norwegian oil and gas producing platforms. The results from regression analysis on survey data show that the psychosocial risk indicator significantly impacts frequency of hydrocarbon leaks.

Although it is generally acknowledged that HOE has a significant impact on the fire and explosion risk, it is difficult to quantify how the HOE affects the fire and explosion risk due to scarcity of HOE data and uncertainty of HOE. In order to deal with the problem, Musharraf et al. [33] addressed the issue of handling uncertainty associated with expert judgments with evidence theory and described the dependency among the human factors and associated actions using BN. They also present a virtual experimental technique

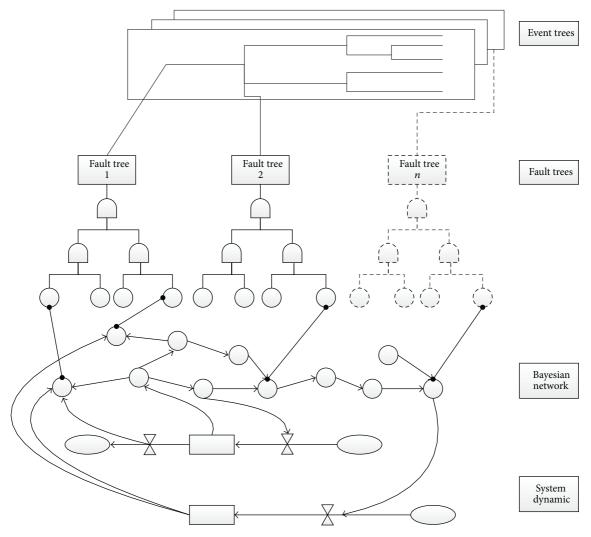


FIGURE 1: The proposed hybrid framework.

for data collection for BN to human reliability analysis. Groth et al. [34] proposed the use of BN to formally incorporate simulator data into estimation of human error probabilities.

3. Quantitative Risk Analysis of Offshore Fire and Explosion Based on Human and Organizational Errors

In this section, a dynamic quantitative risk analysis model of offshore fire and explosion is built by incorporating the effect of HOE.

3.1. Incorporating Human and Organizational Factors into Quantitative Risk Analysis. A method of applying BN in risk analyses has been suggested in the hybrid causal logic framework and fully developed by letting BN be logically and probabilistically integrated into fault tree. Thereafter, some parts of the risk analysis can be addressed in fault tree, while others are addressed in BN. Event tree/fault tree are often

considered as the best option for technical aspects, while HOEs in many cases fit better into a BN. By taking the advantages of the three techniques, the result of combining event tree/fault tree and BN is normally a more detailed risk model with higher resolution.

The above hybrid causal logic model can express the static relationships between logical variables. However, it cannot deal with dynamic characteristics of process variables and human behavior. In order to quantify the dynamic influence of HOEs on the total risk, system dynamics is combined with the above hybrid causal logic model. The new framework is illustrated in Figure 1, which shows the link between system dynamics, BN, event tree, and fault tree. The first interface in this hybrid framework is the one between BN and event tree/fault tree. System dynamics model describes dynamic deterministic relations and provides a dynamic integration among the other two models. In the proposed framework, different models will have both inputs and outputs to the system dynamics model to allow the entire hybrid framework to capture feedback and delays.

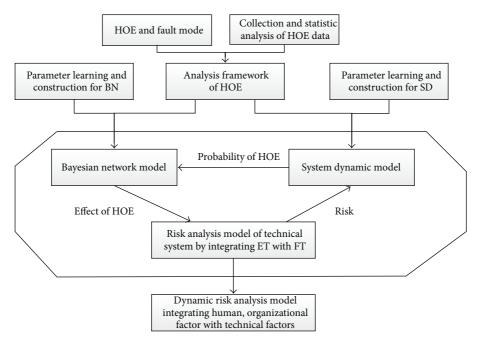


FIGURE 2: Dynamic risk analysis model integrating human and organizational factors with technical factor.

Based on Figure 1, a dynamic risk analysis model is proposed by integrating human and organizational factors with technical factor as shown in Figure 2.

Firstly, historical HOE data are collected and statistically analyzed to build a HOE framework using the proposed hybrid causal logic. Parameter learning and construction of BN and system dynamics model are researched separately to build a more refined HOE model. The interface of system dynamics with BN, event tree, and fault tree can be captured by importing and exporting the data from system dynamics model. The target node calculated from BN is imported to system dynamics and processed inside system dynamics, and the estimated values from system dynamics can be exported to BN. The interaction effects of various factors could be researched in this cyclic process. At last, a dynamic quantitative risk analysis model is built by integrating human and organizational factors with technical factors.

- 3.2. Quantitative Risk Analysis Model of Offshore Fire and Explosion Based on the Effect Analysis of HOEs. As shown in Section 3.1, the dynamic effects of human, organizational, and technical factors to system risk are quantitatively simulated by integrating system dynamics and BN with event tree and fault tree. Based on this, a quantitative risk analysis model for offshore fire and explosion is discussed from both consequence and probability perspective in this section and is illustrated in Figure 3.
- 3.2.1. Consequence Analysis of Offshore Fire and Explosion. Firstly, a set of hydrocarbon release scenarios, including the position, the size of latent leak source, and ignition source, are defined by hazard identification study using FMEA method. The purpose is to identify and describe the scenarios

that may lead to fire/explosion of offshore platform. The most credible fire/explosion scenarios are fixed via Latin Hypercube sampling methods.

Every scenario is simulated using FLACS to characterize the fire/explosion load profiles according to temperature and heat amount. The structure consequences of fire/explosion are simulated via nonlinear structural analysis. The personnel consequences of fire/explosion are analyzed considering the exposed individuals and the fire/explosion load of every accident scenario. For example, potential loss of life (PLL) can be calculated by the following equation:

$$PLL = \sum_{n=1}^{N} \sum_{j=1}^{J} F_{nj} \cdot C_{nj}, \tag{1}$$

where F_{nj} = annual frequency of scenario n with personnel consequence j, C_{nj} = the expected number of fatalities for scenario n with personnel consequence j, N = the total number of scenarios, and J = the total number of the consequence types, including immediate fatalities, escape fatalities, and rescue fatalities.

3.2.2. Probability Analysis of Offshore Fire and Explosion. Historical data about leak, fire, and explosion of offshore platform are collected and statistically analyzed. Based on this, the fire and explosion frequency model is constructed by integrating event tree/fault tree and BN with system dynamics: the failures of safety barriers preventing offshore fire and explosion are analyzed using event tree/fault tree to develop a risk analysis model for technical systems; the dynamic effects of HOE on leak frequency and ignition probability are quantitatively simulated by integrating system dynamics with BN. The system dynamics module depicts

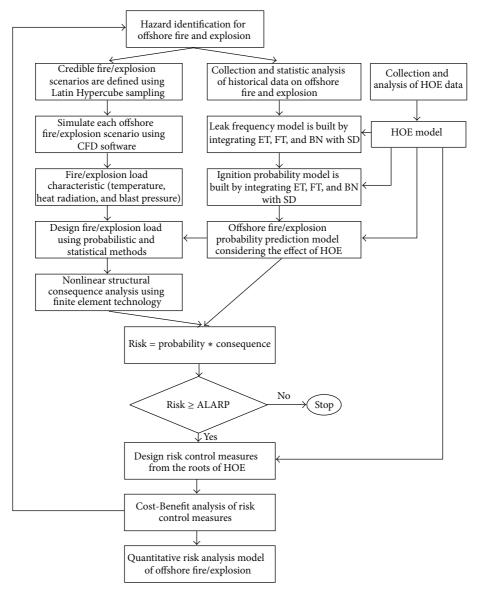


FIGURE 3: Quantitative risk analysis model of offshore fire and explosion.

dynamic deterministic relations in the proposed hybrid causal logic model and provides a dynamic integration among event tree/fault tree and BN model.

Review and analysis of historical documents are performed to get the offshore industry average probabilities/ frequencies of human, organizational, and technical factor, which are assigned to the initiating events and the basic events in the hybrid causal logic model and carry out a quantitative analysis of the offshore fire and explosion frequency by using these data. The results of this calculation may to some degree reflect offshore platform specific conditions since specific data should be applied when possible. Offshore platform data may be found in, for example, incident databases, log data, and maintenance databases. In practice, extensive use of industry average data is necessary to be able to carry out the quantitative analysis. Several databases are available presenting offshore industry average data like OREDA for

equipment reliability data and THERP and CORE-DATA for HOE data. In some cases, neither offshore platform specific data nor generic data may be found, and it may be necessary to use expert judgment to assign probabilities.

3.2.3. Risk Analysis of Offshore Fire and Explosion. As shown in Figure 3, the fire and explosion risk of offshore platform can be calculated as the product of frequency and consequences. If the calculated risk level is greater than the acceptable risk level, then the risk control options should be adopted from the root of human and organizational factors. The adopted risk control options should be evaluated using Cost-Benefit analysis method to get the most beneficial measures to reduce the fire/explosion risk of offshore platform.

"As low as reasonably practicable" principle is defined in terms of exceeding damage probability for main safety functions or probability of accident escalation. The "as low

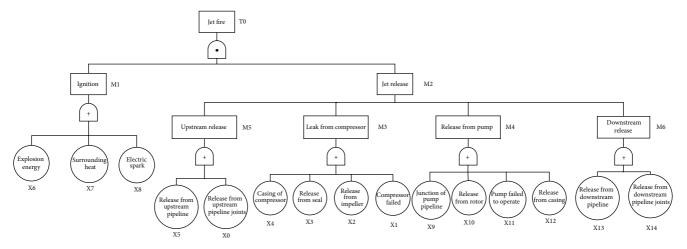


FIGURE 4: Fault tree of jet fire.

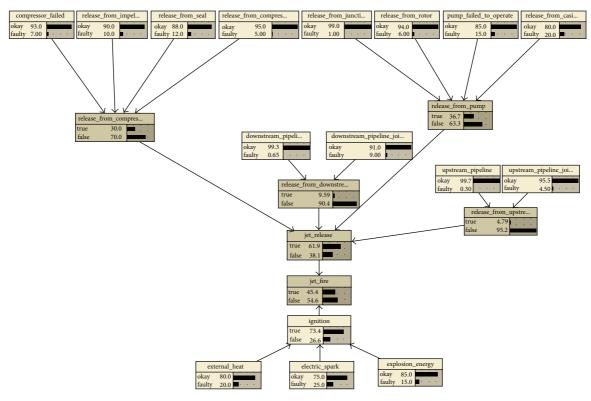


FIGURE 5: Bayesian network of jet fire.

as reasonably practicable" principle in conjunction with the results of quantitative risk analysis can be used to manage and reduce the risks to the lowest level that is reasonably practicable.

4. Case Study for Fire Probability Analysis on Offshore Platform

In this section, the proposed model in Section 3.1 is demonstrated on a simple case study of jet fire on offshore platform. The detailed scenarios are analyzed using fault tree as shown

in Figure 4. The probability data of basic events were collected from Offshore Reliability Data Handbook [35], World-Wide Offshore Accident Databases, and reports by HSE [2].

The basic events, their important degree of probability, and failure probability are shown in Table 1.

The probability of jet fire is 0.505938 according to the calculation results of fault tree. From the calculation for importance degree of probability, it can be seen that the effects of X6, X7, and X8 on the probability of jet fire are larger than other basic events. The fault tree shown in Figure 4 can be converted into BN, which is drawn in Figure 5.

TABLE 1: The failure probability and basic events of fault tree.

	Basic event	Failure probability	Importance degree of probability
X0	Release from joints of upstream pipeline	0.045	0.00865
X1	Compressor completely failed causing release	0.07	0.00865
X2	Release from impeller	0.1	0.00865
X3	Release from seal	0.12	0.00865
X4	Release from casing of compressor	0.05	0.00865
X5	Release from upstream pipeline	0.048	0.00865
X6	Ignition due to explosion energy	0.15	0.01231
X7	Ignition due to external heat from surrounding	0.2	0.0124
X8	Ignition due to electric spark	0.25	0.01314
X9	Release from junction of pump and pipeline	0.01	0.00865
X10	Release from rotor	0.06	0.00865
X11	Pump failed to operate causing release	0.15	0.00865
X12	Release from casing	0.2	0.00865
X13	Release from downstream pipeline	0.0065	0.00865
X14	Release from joints of downstream pipeline	0.09	0.00865

TABLE 2: Conditional probability and mutual information of basic event.

Basic event	Prior probability (%)	Posterior probability (%)	Mutual information
Release from joints of upstream pipeline	4.5	7.4	0.001226
Compressor completely failed	7	11.4	0.00186
Release from impeller	10	16.3	0.002763
Release from seal	12	19.5	0.003406
Release from compressor	5	8.13	0.001296
Release from upstream pipeline	0.3	0.49	7.772e - 005
Ignition due to explosion energy	36.9	75.5	0.04534
Ignition due to external heat	30.3	72	0.05432
Ignition due to electric spark	25.8	47.3	0.002085
Release from junction of pump	1	1.62	0.0002445
Release from rotor	6	9.73	0.001559
Pump failed to operate	15	24.3	0.004398
Release from casing	20	32.4	0.006315
Release from downstream pipeline	0.65	1.06	2.175e - 005
Leak from joints of downstream pipeline	9	14.7	0.002478

From BN inference shown in Figure 5, it can be seen that the occurrence probability of jet fire is computed as 0.0917 per year. The prior probability, posterior probability, and mutual information (entropy reduction) of each basic event are compared, which is shown in Table 2.

From Table 2, it can be concluded that "external heat," "explosion energy," and "electric spark" have the larger contribution to the occurrence of jet fire. The BN shown in Figure 5 is extended by incorporating the effect of human and organizational factors as shown in Figure 6.

The prior probability, posterior probability, and mutual information (entropy reduction) of each human and organization factor are compared as shown in Table 3.

From Table 3, it can be seen that human factor "not_comply_with_instruction" and organizational factor "inefficient_emergency_plan" have the largest contribution to the occurrence of the eventual fire accident. This analysis shows that particular attention should be paid to "comply with instruction" and "efficient emergency plan."

Dynamic quantitative risk analysis is carried out by integrating human and organizational factors with technical factors (see Figure 7).

The simulation results of Figure 7 are shown in Figure 8. Figure 8 displays the probability trend that could be traced in jet fire, ignition, and jet release in a period of 10 years. The advantage of the dynamic model over those of

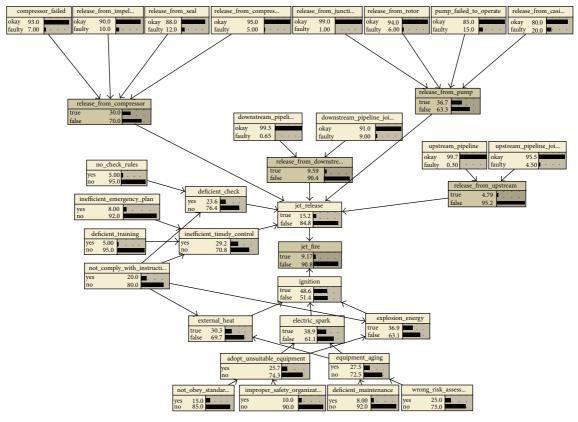


FIGURE 6: Extended Bayesian network considering human and organizational factors.

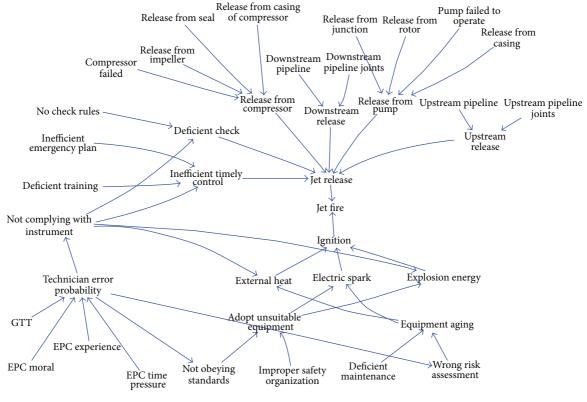


FIGURE 7: Dynamic risk analysis of jet fire using system dynamics.

Basic event	Prior probability (%)	Posterior probability (%)	Mutual information	
no_check_rules	5	5.6	0.0000537	
not_comply_with_instruction	20	75.2	0.107	
inefficient_emergency_plan	8	13.4	0.002483	
deficient_training	5	8.85	0.001914	
improper_safety_organization	10	11.3	0.0001325	
not_obey_standards	15	17.3	0.0002871	
deficient_maintenance	8	8.85	0.0000702	
wrong_risk_assessment	25	26.7	0.0001054	

TABLE 3: Conditional probability and mutual information of human and organization factor.

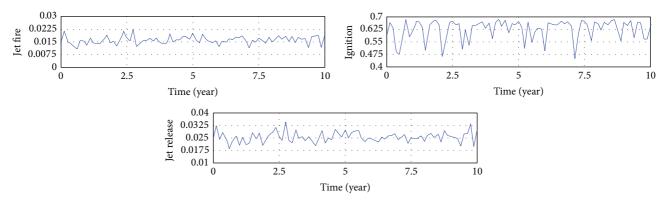


FIGURE 8: Jet fire probability varies with time during a period of ten years.

static models is that it can show the dynamic probability variation with time.

5. Conclusion

The exploration of accidents in light of human error linked to underlying factors related to the human and organization work has been established as a major priority. In order to improve traditional risk assessment methods without considering HOE, new risk assessment methods need to be researched further based on analyzing HOE. The purpose of this paper is to simulate the dynamic effect of HOE on offshore fire/explosion risk, and its contributions can be summarized as follows:

- (1) A hybrid framework, including an integration of system dynamics, BN, event tree, and fault tree, is built to analyze the dynamic offshore fire/explosion risk based on the effect analysis of HOE. The proposed model is demonstrated on a simple case study of jet fire on offshore platform.
- (2) The hybrid framework integrates deterministic and probabilistic modeling techniques, which can be used to analyze the dynamic effects of HOEs on the risk of complex social-technical system.
- (3) The quantitative risk analysis framework for offshore fire and explosion is discussed from both consequence and probability perspective.

(4) The effective prevention measures to reduce the risk of offshore fire/explosion can be designed from the root of HOE. This will provide guideline for risk management of offshore platform.

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 no. 51409260), the Fundamental Research Funds for the Central Universities (Grant no. 14CX05035A), and Shandong Province Natural Science Fund Project (Grant no. ZR2012EEM023).

References

- [1] HSE, "Offshore accident," Incident Statistics Reports OTO96.954, HSE, London, UK, 1996.
- [2] HSE, Accident Statistics for Floating Offshore Units on the UK Continental Shelf (1980–2003), HSE, 2005.
- [3] J. K. Paik and J. Czujko, "Explosion and fire engineering and gas explosion of FPSOs (phase I): hydrocarbon releases on FPSOs —review of HSE's accident database," Final Report No. EFEF JIP-02, Research Institute of Ship and Offshore Structural

- Design Innovation, Pusan National University, Busan, Korea, 2009.
- [4] R. G. Bea, "The role of human error in design. Construction and reliability of the marine structure," U. S. A. Ship Structure Committee Report SSC2378, 1994.
- [5] X. Qu, Y. Yang, Z. Liu, S. Jin, and J. Weng, "Potential crash risks of expressway on-ramps and off-ramps: a case study in Beijing, China," *Safety Science*, vol. 70, pp. 58–62, 2014.
- [6] Q. Meng, J. Weng, and X. Qu, "A probabilistic quantitative risk assessment model for the long-term work zone crashes," *Acci*dent Analysis and Prevention, vol. 42, no. 6, pp. 1866–1877, 2010.
- [7] Q. Meng, X. Qu, K. T. Yong, and Y. H. Wong, "QRA model based risk impact analysis of traffic flow in urban road tunnels," *Risk Analysis*, vol. 31, no. 12, pp. 1872–1882, 2011.
- [8] Q. Meng, X. Qu, X. Wang, V. Yuanita, and S. C. Wong, "Quantitative risk assessment modelling for non-homogeneous urban road tunnels," *Risk Analysis*, vol. 31, no. 3, pp. 382–403, 2011.
- [9] J. K. Paik, J. Czujko, B. J. Kim et al., "Quantitative assessment of hydrocarbon explosion and fire risks in offshore installations," *Marine Structures*, vol. 24, no. 2, pp. 73–96, 2011.
- [10] L.-B. Li, Z.-L. Wang, Q. Yin, and Y.-J. Sun, "Study on influencing factors of risk of offshore platforms under fire and gas explosion disaster," *The Ocean Engineering*, vol. 29, no. 2, pp. 92–98, 2011.
- [11] W. Røed, A. Mosleh, J. E. Vinnem, and T. Aven, "On the use of the hybrid causal logic method in offshore risk analysis," *Reliability Engineering and System Safety*, vol. 94, no. 2, pp. 445– 455, 2009.
- [12] J. Krueger and D. Smith, "A practical approach to fire hazard analysis for offshore structures," *Journal of Hazardous Materials*, vol. 104, no. 1–3, pp. 107–122, 2003.
- [13] R. Pula, F. I. Khan, B. Veitch, and P. R. Amyotte, "Revised fire consequence models for offshore quantitative risk assessment," *Journal of Loss Prevention in the Process Industries*, vol. 18, no. 4–6, pp. 443–454, 2005.
- [14] R. Pula, F. I. Khan, B. Veitch, and P. R. Amyotte, "A grid based approach for fire and explosion consequence analysis," *Process Safety and Environmental Protection*, vol. 84, no. 2, pp. 79–91, 2006.
- [15] J. A. Suardin, A. Jeff McPhate Jr., A. Sipkema, M. Childs, and M. S. Mannan, "Fire and explosion assessment on oil and gas floating production storage offloading (FPSO): an effective screening and comparison tool," *Process Safety and Environmental Protection*, vol. 87, no. 3, pp. 147–160, 2009.
- [16] B. J. Kim, J. K. Seo, J. H. Park et al., "Load characteristics of steel and concrete tubular members under jet fire: an experimental and numerical study," *Ocean Engineering*, vol. 37, no. 13, pp. 1159–1168, 2010.
- [17] J. Y. Yoon, S. H. Kim, G. C. Yu et al., "Effects of wind on the heat flow of fpso topsides subject to fire: an experimental and numerical study," in *Proceedings of the ASME 29th International Con*ference on Ocean, Offshore and Arctic Engineering, pp. 949–958, June 2010.
- [18] J. K. Paik, B. J. Kim, J. S. Jeong et al., "CFD simulations of gas explosion and fire actions," *Ships and Offshore Structures*, vol. 5, no. 1, pp. 3–12, 2010.
- [19] A. Ronza, J. A. Vílchez, and J. Casal, "Using transportation accident databases to investigate ignition and explosion probabilities of flammable spills," *Journal of Hazardous Materials*, vol. 146, no. 1-2, pp. 106–123, 2007.
- [20] M. Moosemiller, "Development of algorithms for predicting ignition probabilities and explosion frequencies," *Journal of Loss*

- Prevention in the Process Industries, vol. 24, no. 3, pp. 259–265, 2011.
- [21] J. E. Vinnem, "On the development of failure models for hydrocarbon leaks during maintenance work in process plants on offshore petroleum installations," *Reliability Engineering and System Safety*, vol. 113, no. 1, pp. 112–121, 2013.
- [22] N. Khakzad, F. Khan, and N. Paltrinieri, "On the application of near accident data to risk analysis of major accidents," *Reliability Engineering and System Safety*, vol. 126, pp. 116–125, 2014.
- [23] J. K. Paik and J. Czujko, "Assessment of hydrocarbon explosion and fire risks in offshore installations: recent advances and future trends," *IES Journal—Part A: Civil and Structural Engineering*, vol. 4, no. 3, pp. 167–179, 2011.
- [24] Y.-F. Wang, M. Xie, M. S. Habibullah, and K.-M. Ng, "Quantitative risk assessment through hybrid causal logic approach," Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, vol. 225, no. 3, pp. 323–332, 2011.
- [25] Y. F. Wang, S. Faghih Roohi, X. M. Hu, and M. Xie, "Investigations of human and organizational factors in hazardous vapor accidents," *Journal of Hazardous Materials*, vol. 191, no. 1–3, pp. 69–82, 2011.
- [26] Y. F. Wang, M. Xie, K.-S. Chin, and X. J. Fu, "Accident analysis model based on Bayesian Network and Evidential Reasoning approach," *Journal of Loss Prevention in the Process Industries*, vol. 26, no. 1, pp. 10–21, 2013.
- [27] Y. F. Wang, M. Xie, K. M. Ng, and M. S. Habibullah, "Probability analysis of offshore fire by incorporating human and organizational factor," *Ocean Engineering*, vol. 38, no. 17-18, pp. 2042– 2055, 2011.
- [28] M. Abimbola, F. Khan, and N. Khakzad, "Dynamic safety risk analysis of offshore drilling," *Journal of Loss Prevention in the Process Industries*, vol. 30, no. 1, pp. 74–85, 2014.
- [29] J. E. Vinnem, R. Bye, B. A. Gran et al., "Risk modelling of maintenance work on major process equipment on offshore petroleum installations," *Journal of Loss Prevention in the Process Industries*, vol. 25, no. 2, pp. 274–292, 2012.
- [30] S. Dan, C. J. Lee, J. Park, D. Shin, and E. S. Yoon, "Quantitative risk analysis of fire and explosion on the top-side LNG-liquefaction process of LNG-FPSO," *Process Safety and Environmental Protection*, vol. 92, pp. 430–441, 2014.
- [31] J. E. Vinnem, J. A. Hestad, J. T. Kvaløy, and J. E. Skogdalen, "Analysis of root causes of major hazard precursors (hydrocarbon leaks) in the Norwegian offshore petroleum industry," *Reliability Engineering and System Safety*, vol. 95, no. 11, pp. 1142–1153, 2010.
- [32] L. I. V. Bergh, A. J. Ringstad, S. Leka, and G. I. J. M. Zwetsloot, "Psychosocial risks and hydrocarbon leaks: an exploration of their relationship in the Norwegian oil and gas industry," *Journal of Cleaner Production*, vol. 84, no. 1, pp. 824–830, 2014.
- [33] M. Musharraf, D. Bradbury-Squires, F. Khan, B. Veitch, S. Mackinnon, and S. Imtiaz, "A virtual experimental technique for data collection for a Bayesian network approach to human reliability analysis," *Reliability Engineering and System Safety*, vol. 132, pp. 1–8, 2014.
- [34] K. M. Groth, C. L. Smith, and L. P. Swiler, "A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods," *Reliability Engineering and System Safety*, vol. 128, pp. 32–40, 2014.
- [35] Det Norske Veritas, OREDA-Offshore Reliability Data Handbook, Det Norske Veritas, Høvik, Norway, 2009.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 194568, 10 pages http://dx.doi.org/10.1155/2015/194568

Research Article

Improvement on Load-Induced Cascading Failure in Asymmetrical Interdependent Networks: Modeling and Analysis

Haiyan Han and Rennong Yang

School of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an 710038, China

Correspondence should be addressed to Haiyan Han; hanhaiyan03@126.com

Received 2 June 2015; Revised 1 August 2015; Accepted 4 August 2015

Academic Editor: Xiaobo Qu

Copyright © 2015 H. Han and R. Yang. 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.

Many real-world systems can be depicted as interdependent networks and they usually show an obvious property of asymmetry. Furthermore, node or edge failure can trigger load redistribution which leads to a cascade of failure in the whole network. In order to deeply investigate the load-induced cascading failure, firstly, an asymmetrical model of interdependent network consisting of a hierarchical weighted network and a WS small-world network is constructed. Secondly, an improved "load-capacity" model is applied for node failure and edge failure, respectively, followed by a series of simulations of cascading failure over networks in both interdependent and isolated statuses. The simulation results prove that the robustness in isolated network changes more promptly than that in the interdependent one. Network robustness is positively related to "capacity," but negatively related to "load." The hierarchical weight structure in the subnetwork leads to a "plateau" phenomenon in the progress of cascading failure.

1. Introduction

Complex network theory has been a successful tool in modeling and analysis of modern systems [1]. A variety of network models have been proposed to approximate realistic systems such as power grids [2], transportation systems [3], communication networks [4], and other systems. Most of them focus on single and isolated systems. Recently, interdependent network [5] provides a new insight in understanding the structure [6], percolation [7], spreading processes [8], evolution games [9], and robustness [10, 11] of complex systems. This kind of structure usually consists of two or more subnetworks, with the subnetworks working dependently on each other [12, 13]. For example, power grids and computer systems must depend on each other because computers require power grids to supply electricity. Power grids, in turn, rely on computer systems to control power transmission process.

In 2010, Buldyrev et al. [14] found that interdependent networks have become significantly more vulnerable than their noninteracting counterparts under random attack. Gao et al. [5] reviewed the connectivity properties of "networks of networks" formed by interdependent random networks.

Hu et al. [15] constructed a partially coupled network with both interdependent and interconnecting links. The interconnections satisfy "one to one" condition. They found that the change of interconnecting links leads to the change of the phase transition from second order to first order through hybrid phase transition. Huang et al. [16] studied the robustness of interdependent networks under targeted attack on high or low degree nodes. It provided a routine method to study the degree-based targeted attack problems in both single networks with dependency links [17, 18] and other general randomly connected and uncorrelated interdependent networks. Moreover, Parshani et al. [19] described the dynamic process of cascading failures on two partially interdependent networks.

Currently, researchers focus on the problem of modelling an interdependent network model for analyzing the progress of cascading failure. Based on classical network models such as ER [20], WS [21], and BA [22], researchers studied well some symmetrical and asymmetrical networks [19, 23, 24], such as ER-ER, WS-WS, BA-BA, ER-WS, BA-ER, and BA-WS; but these models differ greatly from real-world systems [25, 26]. Taking interdependent networks like ground transportation network and airline network as an example, ground

transportation network displays a hierarchical property. That is, stations at provincial level have higher capacity and more importance than municipal stations. However, the airline network can be seen as a single level network with small world property [27]. So, some other researchers [28–30] made their trials to construct framework of asymmetrical interdependent networks which are more authentic to reality.

Moreover, load such as cargoes transported in the transportation network and electric stream in power grids can trigger cascading failures [31]. The load carried by the failed nodes or edges will not disappear but will flow to the remaining part of the network, which, in a possible way, will cause further failures. As far as we are concerned, not enough attention has been paid to the cascading failure induced by load redistribution in interdependent networks.

In order to investigate the load-induced cascading failure in interdependent networks, we propose asymmetrical interdependent networks model in Section 2 and the load-induced cascading failure model in Section 3. Section 4 simulates the cascading failure in the proposed model and compares it with a WS-WS symmetrical network when nodes and edges suffer from intentional attacks, respectively. Section 5 summarizes the contribution of this paper and identifies future research needs.

2. Network Model

Figure 1 shows an asymmetrical interdependent network model. Network has a hierarchical and weighted structure, where nodes are assigned with weights and distributed into different levels, where Network is a single level network. The numbers of nodes in Network and Network are equal. The coupling proportion is set to p_1 . Nodes in different subnetworks randomly construct "one to one" connections; for example, node i_A in Network merely couples with j_B in Network p_B .

The model can be described mathematically by G(N, E, W), where $N = N_A + N_B$ is the set of nodes from Network_A and Network_B. $E = E_A + E_B + E_{AB}$ denotes the set of edges, where E_A (E_B) denotes the internal edges of Network_A (Network_B) and E_{AB} represents the coupling edges between the two subnetworks. Matrix $W = \begin{bmatrix} W_A & W_{AB} \\ W_{AB}^T & W_B \end{bmatrix}$ represents the adjacency matrix, where W_A and W_B represent the connections inside Network_A and Network_B, respectively, while W_{AB} is the coupling adjacency matrix. The element ω_{ij} of W represents the weight of connection between two neighboring nodes i and j. Node intensity s_i is defined to represent the total weights of all edges connected with node i:

$$s_i = \sum_{j \in \Gamma_i} \omega_{ij},\tag{1}$$

where Γ_i is the neighborhood of node i, including neighboring nodes in both Network_A and Network_B.

As mentioned before, in Network_A, nodes are divided into several levels. There are H levels in Network_A and each node from a certain level has M subordinate nodes from the next level. There are no connections between subordinate

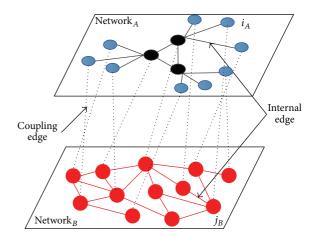


FIGURE 1: Model of asymmetrical interdependent networks, in which two subnetworks are partly coupled together with proportion p_1 . The solid lines in Network_A and Network_B represent the internal edges, while the dashed lines connecting the two subnetworks represent the coupling edges.

nodes in the same level except for the top level, where all nodes are connected with each other. Internal edges connecting different levels have diverse weights representing the difference in importance of interlevel connections.

The hierarchical structure is established as follows. Firstly, we establish a fully connected top level, the level marker is set to h=1, the number of nodes is n_1 , and the edge weight is ω_1 . Then, each node in top level establishes relations with M subordinate nodes from the next level, where the level marker is set to h=2. Nodes in this level do not connect with each other. We repeat the former step until h=H; then, Network $_A$ with hierarchical weight is completed. The weight of edges connecting level h and level h+1 is set to

$$\omega_{h,h+1} = \omega_1 - C \cdot (h-1), \qquad (2)$$

where constant *C* controls the weightiness of internal edges which connect neighboring levels.

Network $_B$ is evolved by the rule of WS small-world network because small-world network yields the shortest and most effective paths [32]. The weight of each internal edge is equivalent. It turns from regular network to random network by adjusting the reconnecting probability $p_2 \in [0,1]$, where $p_2 = 0$ corresponds to the case of nearest-neighbor coupled network; the nodes are placed on a ring lattice with periodic boundary conditions and each node is initially connected to K nearest neighbors. When $p_2 \in (0,1)$, the WS small-world network model is created by rewiring a small fraction of the links with probability p_2 to nodes chosen at random. The average degree of node is still K. When $p_2 = 1$, it corresponds to a completely random network.

The coupling proportion is set to $p_1 = 0.7$ between Network_A and Network_B. The sizes of the subnetworks are set to $N_A = N_B = 340$. In Network_A, M = 4, H = 4, and $H_1 = 4$, so the numbers of subordinate nodes in the next three levels are 16, 64, and 256. Because C = 2 and $\omega_1 = 10$, the weights

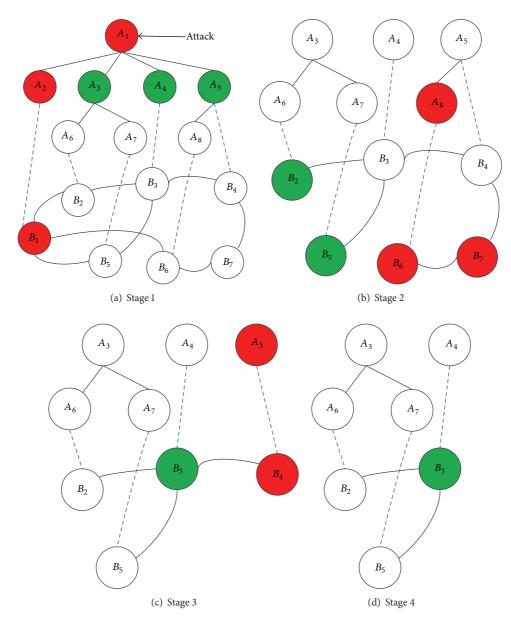


FIGURE 2: The cascading failure under node attack on asymmetrical interdependent network. The initial topology of the asymmetrical network, shown in (a), has a set of nodes in Network_A and Network_B, labeled as $\{A_1, A_2, \ldots, A_8\}$ and $\{B_1, B_2, \ldots, B_7\}$, respectively. The internal edges are represented as solid black lines and the coupling edges are represented as dashed blue lines. Red circles represent the failed nodes and green circles represent the nodes that suffered from the load redistribution but are still active.

of internal edges that connect different levels are $\omega_{1,2}=8$, $\omega_{2,3}=6$, and $\omega_{3,4}=4$ according to (2). Network $_B$ is a WS small-world network with $p_2=0.1$ and K=4. Due to the randomly "one to one" relationship, we set the weight of coupling edges to be equal to the maximum weight of internal edges connected to the coupled node in Network $_A$.

3. Cascading Failure Model

The coupling property makes the interdependent networks fragile when suffering from intentional attack [16, 18]. Here we focus on the cascading failure of interdependent network

induced by load. We define the initial loads $L_i(0)$ of node i as a function of the node intensity s_i :

$$L_i(0) = \lambda s_i^{\alpha},\tag{3}$$

where load parameters $\lambda > 0$ and $\alpha > 0$ are adjustable to control the distribution of initial load. We can see that each node bears more or less some initial loads according to the node intensity and the loads on each node are a nonlinear function unless the parameter α is equal to 1. ML model [33] conjectures that the capacity of a node is proportional to its

initial loads. There is a linear relationship between capacity C_i of node i and its initial load $L_i(0)$. Consider

$$C_i = (1 + \beta) \cdot L_i(0), \tag{4}$$

where $\beta > 0$ is the tolerance parameter. It is well studied that big β leads to excellent robustness. However, considering other perspectives such as cost, the capacity is impossible to be infinite, so it tends to a finite β . In addition, the linear relationship does not fit for many real-world networks, so a nonlinear "load-capacity" model is proposed as follows:

$$C_i = L_i(0) + \beta \cdot L_i(0)^{\theta}, \qquad (5)$$

where two tunable parameters $\beta > 0$ and $\theta > 0$ are introduced. If $\theta = 1$, it decays to the linear "load-capacity" model. According to a local nearest redistribution strategy, the proportions of load distribution Π_j and the new added load $\Delta L_{i \to j}$ where the failed node i passes to its neighbor j are as follows[34]:

$$\Pi_{j} = \frac{C_{j}}{\sum_{n \in \Gamma_{i}} C_{n}},$$

$$\Delta L_{i \to j} = \Pi_{j} L_{i} = \frac{C_{j}}{\sum_{n \in \Gamma_{i}} C_{n}} L_{i},$$
(6)

where n is the neighbor of the failed node i and Γ_i is the set of neighbors which node i connects. If the initial load of node j plus the load that node i transfers to exceeds its capacity $(C_j < L_j + \Delta L_{i \to j})$, node j fails and leads to a new round of load redistribution. The process repeats until there is no overloaded node or the entire network is paralyzed. The evolving procedure is illustrated in Figure 2.

At stage 1 in Figure 2(a), node A_1 is attacked and fails; the load distributes to its neighbors $\{A_2, A_3, A_4, A_5\}$. Node A_2 fails because of overloading and it triggers the failure of B_1 because the load of B_1 exceeds its capacity after receiving some amount of load from A_2 . Meanwhile, nodes A_3 , A_4 , and A_5 are still active because they are not overloaded. At stage 2 in Figure 2(b), due to the load redistribution of node B_1 , node B_6 is overloaded, but nodes B_2 and B_5 are still active. B_6 distributes the load to its neighboring nodes which leads to the failure of nodes B_7 and A_8 . In this way, the cascading failure propagates in both Network $_A$ and Network $_B$. At stage 3 in Figure 2(c), the failures of nodes B_7 and A_8 trigger the failures of B_4 and A_5 . There is no new failure in the interdependent networks in Figure 2(d) and the cascading progress ends at stage 4.

We use R_s to express the relative scale damage caused by node failures:

$$R_s = \frac{S'}{S},\tag{7}$$

where S(S') is the sum of initial node intensity before (after) node failures.

Similar to the case of node failure, the initial load of edge $L_{ij}(0)$ is related to the node intensities of i and j, as defined in (8). The "load-capacity" model of edges can be expressed

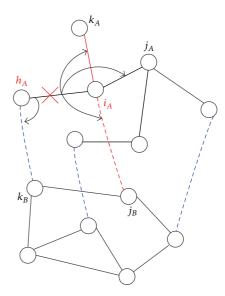


FIGURE 3: The load redistribution triggered by an edge failure. The internal edges are represented as solid black lines, while the coupling edges are represented as dashed blue lines. The black arrowed lines represent the directions of load redistribution from the failed edge and the red solid line represents the edge which is overloaded.

in (9). Figure 3 shows the load redistribution rule for edge failure. Consider

$$L_{ij}(0) = \left(s_i \cdot s_j\right)^{\gamma},\tag{8}$$

where $\gamma > 0$ is a tunable parameter to control initial load of edges. Consider

$$C_{ij} = L_{ij}(0) + \beta \cdot L_{ij}(0)^{\theta}$$
. (9)

If a single internal edge $e_{i_Ah_A}$ in Network $_A$ fails, the load on the broken edge will be redistributed bidirectionally to its neighboring internal and coupling edges. If the load on neighboring edges plus the extra load exceeds their capacity, this may trigger a recursive process of cascading failures, such as internal edge $e_{i_Ak_A}$ and internal edge $e_{i_Aj_B}$ shown in Figure 3. After coupling edge $e_{i_Aj_B}$ fails, its load proportionally is reassigned to neighboring edges connected to node i_A in Network $_A$ and node j_B in Network $_B$. Consider

$$\Delta L_{i_A \to j_A} = \frac{C_{i_A j_A}}{\sum_{a \in \Gamma_{i_A}} C_{i_A a} + \sum_{b \in \Gamma_{j_B}} C_{j_B b} - 2C_{i_A j_B}} L_{i_A j_B},$$

$$\Delta L_{j_B \to k_B} = \frac{C_{j_B k_B}}{\sum_{a \in \Gamma_{i_A}} C_{i_A a} + \sum_{b \in \Gamma_{i_D}} C_{j_B b} - 2C_{i_A j_B}} L_{i_A j_B},$$
(10)

where Γ_{i_A} (Γ_{j_B}) is the set of neighbors of node i_A (j_B) and $L_{i_Aj_B}$ is the initial load of edge $e_{i_Aj_B}$. $\Delta L_{i_A \to j_A}$ is the additional load distributed to edge $e_{i_Aj_A}$. $\Delta L_{j_B \to k_B}$ is the additional load distributed to edge $e_{j_Bk_B}$. Only if $C_{mn} > L_{mn} + \Delta L_{m \to n}$ for any edge will there be no new edge failure in the system.

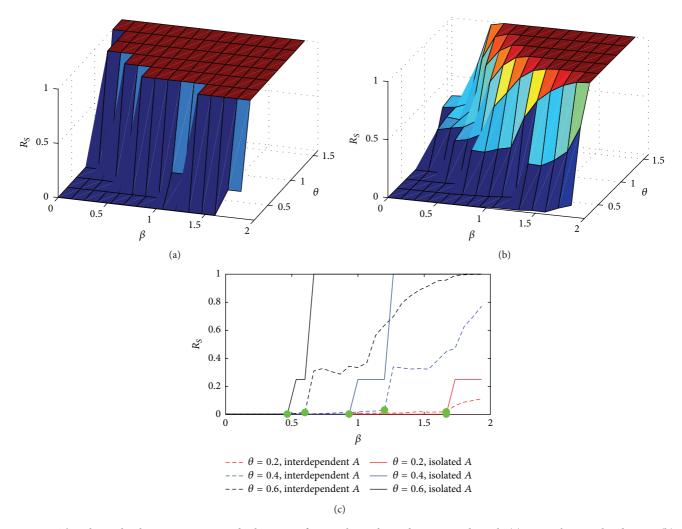


FIGURE 4: The relationship between capacity and robustness of Network_A under node intentional attack. (a) Network_A in isolated status. (b) Network_A in interdependent status. (c) A comparison of Network_A in both isolated status and interdependent status. Capacity parameters $\beta \in [0,2)$ and $\theta \in [0,2,1.6]$. Load parameter $\alpha = 0.8$. In (c), the critical value β_C is labeled with green dot. The solid lines denote the isolated status, while the dashed lines denote the interdependent status. Simulation results are averaged over 1000 independent trials.

We finally adopt the relative size of removed edge as *R* to describe the network robustness under edge attack:

$$R = \sum_{ij} \frac{E_{ij}}{E},\tag{11}$$

where E_{ij} is the number of failed edges due to cascading failure after e_{ij} is cut off, while E is the total number of edges in the whole system before cascading failures.

4. Results and Discussion

4.1. Analysis of Network Robustness under Node Attack. A similar phenomenon of the isolated status and interdependent status is that R_S has positive relations with the capacity parameters β and θ . This means that high capacity benefits the robustness of Network $_A$. Additionally, it is notable that the curve of R_S abruptly emerges when β approaches a critical value in Figure 4(a), and then it rapidly grows to 1; but,

in Figure 4(b), after its abrupt emergence, the robustness indicator then shows a slow increase to 1. In order to find the detailed difference between the two statuses, we fix θ at 0.2, 0.4, and 0.6 separately to compare the curves of R_S in isolated status and interdependent status in Figure 4(c). It can be seen that the solid lines are above the corresponding dashed ones because the interdependent relationship leads to a weaker robustness of Network $_A$. Furthermore, the robustness indicator R_S in isolated status undergoes a short "plateau" state before it grows to 1. In comparison, in the interdependent status, the robustness undergoes a hybrid phase transition.

The robustness of Network $_A$ has a positive relation with the capacity parameter β in Figure 5. However, it has a negative relationship with load parameter α as shown in Figures 5(a) and 5(b). We fix α at 1.2, 1.4, and 1.6 separately to compare the changes of robustness in both statuses. We also find that the robustness of Network $_A$ in isolated status is better than that in interdependent status. In addition, the trend of each curve in Figure 5(c) is similar to that in Figure 4(c).

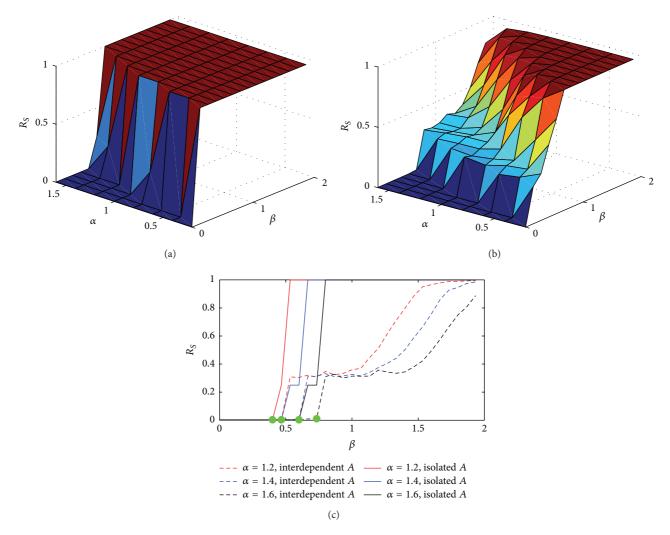


FIGURE 5: The relationship between load and robustness of Network_A under node intentional attack. (a) Network_A in isolated status. (b) Network_A in interdependent status. (c) A comparison of Network_A in both isolated status and interdependent status. Capacity parameters $\beta \in [0,2)$ and $\theta = 0.8$. The critical value β_C is labeled with green dot. The solid lines denote the robustness of Network_A in isolated status, while the dashed lines denote the robustness of Network_A in interdependent status. Simulation results are averaged over 1000 independent trials

Because of the hierarchical weight structure in Network_A, nodes from the same level have equal initial load and capacity according to (1) and (3). If a node fails, the proportion of load redistribution is divided by the level to which the neighboring nodes belong. The failure probabilities of neighboring nodes from the same level are equivalent, while those of the nodes from different levels are not. That finally leads to the sharp change of R_S when most neighbors fail or the short plateau without large scale of failures in the neighborhood. The hierarchical weight structure in Network_A also leads to a cascade of failures in the isolated status unlike a regular first-order phase transition. Meanwhile, the interdependent relationship leads to the failures from a first-order phase transition to a second-order phase transition.

4.2. Analysis of Network Robustness under Edge Attack. When the edge with maximum initial load $\max(L_{ij}(0))$ in Network_A

suffers from intentional attack, the relationship among load, capacity, and robustness is analyzed based on the results in Figures 6 and 7. Under edge attack, the robustness indicator *R* also shows positive correlations with the capacity parameters β and θ but a negative correlation with load parameter γ . The results shown in Figures 6 and 7 also prove that the robustness can be enhanced by high capacity as well as low initial load. As shown in Figures 6(a) and 6(b) and Figures 7(a) and 7(b), the enhancement of robustness is more prompt in isolated status than in interdependent status. In Figures 6(c) and 7(c), the solid line denoting each value of θ and γ is above the corresponding dashed ones which also indicates a weaker robustness of $\mathsf{Network}_A$ in interdependent status. The hierarchical weight structure in Network_A also leads to a plateau on the curves of R in Figure 7(c). Because the initial load and capacity of internal edge L_{ii} which connect the neighboring levels are equal, according to (1), (2), and

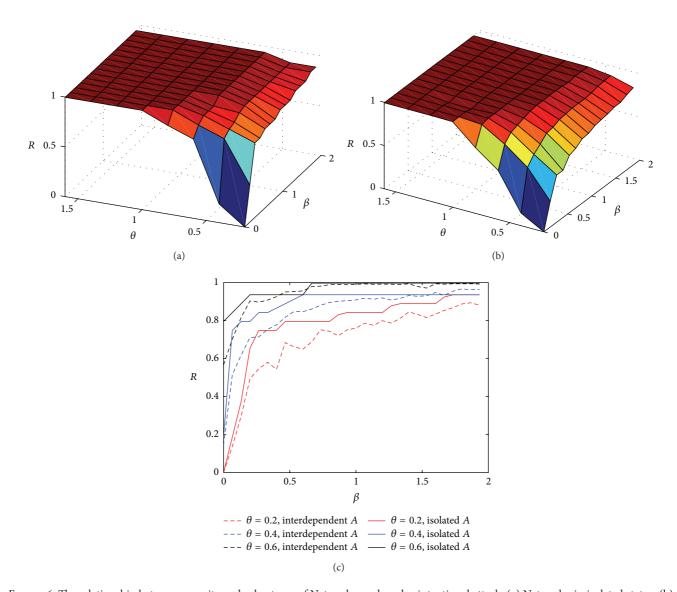


FIGURE 6: The relationship between capacity and robustness of Network_A under edge intentional attack. (a) Network_A in isolated status. (b) Network_A in interdependent status. (c) A comparison of Network_A in both isolated status and interdependent status. Capacity parameters $\beta \in [0,2)$ and $\theta \in [0.2,1.6]$. Load parameter $\gamma = 0.8$. The solid lines denote the isolated status, while the dashed lines denote the interdependent status. Simulation results are averaged over 1000 independent trials.

(8). Edges connecting the neighboring levels have equal initial load and capacity. If one internal edge fails, the failure probabilities of edges connecting the same neighboring levels are equivalent, while those of the edges connecting different levels are not. So it leads to the sharp change of *R* when most edges fail or the plateau without large scale of edge failures.

4.3. Comparison of Asymmetrical and Symmetrical Models. We compare between robustness of the asymmetrical network model we propose and that of a symmetrical network model (WS-WS). The configuration of WS subnets in the symmetrical network model is the same as that of the WS subnet in the asymmetrical network model. The robustness of network under different initial conditions is shown in Figures 8 and 9.

We test intentional node attacks on the studied models under two initial conditions in Figure 8. It can be seen that the robustness of the symmetrical model is stronger than that of the asymmetric model. Additionally, no matter in the symmetrical or the asymmetrical case, the robust levels of Network $_A$ and Network $_B$ under the same conditions are very close to one another.

Additionally, we test intentional edge attacks on the two models under different initial conditions in Figure 9. Similar to the cases on node attack, the robust levels in the symmetrical model are better than those in the asymmetric model. However, what is notable is that $R-\beta$ curves of each subnetwork in the asymmetrical model deviate from each other, while those of the symmetrical model remain resemblant. Network_A is less vulnerable than Network_B.

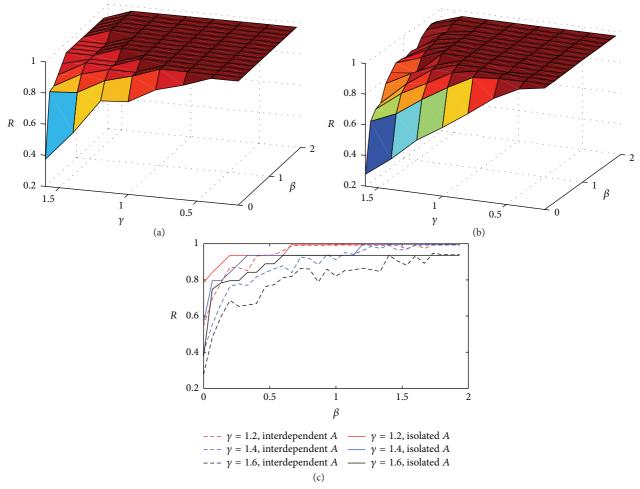


FIGURE 7: The relationship between load and robustness under edge intentional attack in isolated status (a) and in interdependent status (b). (c) A comparison of Network_A in both isolated status and interdependent status. Capacity parameters $\beta \in [0,2)$ and $\gamma \in [0.2,1.6]$. Capacity parameter $\theta = 0.8$. The solid lines denote the robustness of Network_A in isolated status, while the dashed lines denote the robustness of Network_A in interdependent status. Simulation results are averaged over 1000 independent trials.

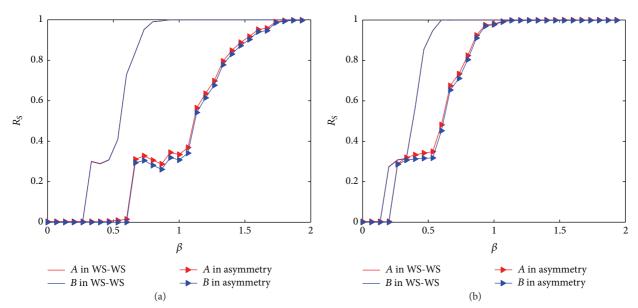


FIGURE 8: Comparison of node attack in symmetrical and asymmetrical network models under different initial conditions. (a) $\alpha=0.8$ and $\theta=0.6$. (b) $\alpha=0.6$ and $\theta=0.8$. The red lines without triangles denote the robustness in WS-WS network model, while solid lines with triangles denote the robustness in Network_B. Simulation results are averaged over 1000 independent trials.

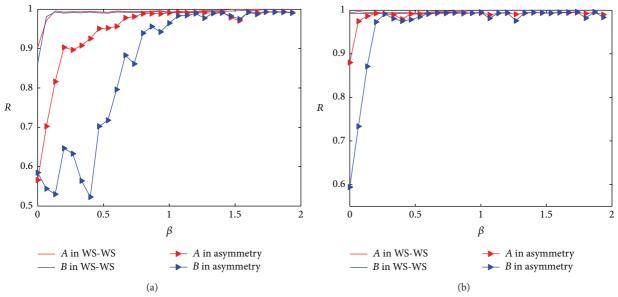


FIGURE 9: Comparison of edge attack in symmetrical and asymmetrical network models under different initial conditions. (a) $\gamma=0.8$ and $\theta=0.6$. (b) $\gamma=0.6$ and $\theta=0.8$. The red lines without triangles denote the robustness in WS-WS network model, while solid lines with triangles denote the robustness in Network_B. Simulation results are averaged over 1000 independent trials.

5. Conclusion

Interdependent network is a representative abstraction of realistic systems, where failures in one system may trigger a cascade of failures in the other system through interactions among systems. For instance, load redistribution caused by node or edge failures in one subnetwork may trigger an iterative progress of cascading failure in interdependent networks. In this paper, firstly, we propose an interdependent network model considering asymmetrical property of the two subnetworks. Secondly, a nonlinear "load-capacity" model is also proposed for nodes and edges, respectively, to model the cascading failures. Thirdly, we conduct a systematic series of experiments to analyze the robustness of interdependent network under attack and make comparisons with the situation of well-studied symmetrical WS-WS network model. (1) Interdependent networks are more fragile compared with isolated ones. (2) The hierarchical weight structure leads to a short "plateau" phenomenon in isolated status unlike the regular first-order phase transition in nonhierarchical isolated networks. Under node attack, only if the capacity reaches a critical value β_C can the robustness abruptly emerge and rapidly approach 1. Furthermore, the interconnections lead to the robustness curve from a first-order phase transition to a second-order phase transition in interdependent networks. (3) Under edge attack, the hierarchical weight structure also causes a plateau in an iterative process of cascading failures in isolated status. The change of robustness in isolated network is more prompt than that in interdependent network. (4) Compared with symmetrical WS-WS network model, the robustness of the asymmetrical model is weaker. Under node attack, no matter in the symmetrical or the asymmetrical case, the two subnetworks have similar robustness. However, under edge attack, subnetworks in the asymmetrical model perform differently, while those of the symmetrical model remain resemblant.

Conflict of Interests

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

Authors' Contribution

Haiyan Han planned the work, implemented the experiment, and drafted the main part of the paper. Rennong Yang contributed to the error analysis.

Acknowledgments

The work described in this paper is supported by the Twelfth Five-Year Plan Pre-Research Project (no. 402040401) and Science and Technology Research and Development Project of Shaanxi Province (no. 2013kjxx-82).

References

- [1] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.
- [2] C. M. Schneider, A. A. Moreira, J. S. Andrade Jr., S. Havlin, and H. J. Herrmann, "Mitigation of malicious attacks on networks," Proceedings of the National Academy of Sciences of the United States of America, vol. 108, no. 10, pp. 3838–3841, 2011.
- [3] G. Li, S. D. S. Reis, A. A. Moreira et al., "Towards design principles for optimal transport networks," *Physical Review Letters*, vol. 104, no. 1, Article ID 018701, 4 pages, 2010.
- [4] J. Zhao, H. Zhou, B. Chen, and P. Li, "Research on the structural characteristics of transmission grid based on complex network theory," *Journal of Applied Mathematics*, vol. 2014, Article ID 261798, 12 pages, 2014.
- [5] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, "Networks formed from interdependent networks," *Nature Physics*, vol. 8, no. 1, pp. 40–48, 2012.

- [6] L. Zhang, J. Cao, and J. Li, "Complex networks: statistical properties, community structure, and evolution," *Mathematical Problems in Engineering*, vol. 2015, Article ID 590794, 7 pages, 2015.
- [7] A. Bashan, R. Parshani, and S. Havlin, "Percolation in networks composed of connectivity and dependency links," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 5, Article ID 051127, 8 pages, 2011.
- [8] S. Funk and V. A. A. Jansen, "Interacting epidemics on overlay networks," *Physical Review E*, vol. 81, no. 3, Article ID 036118, 10 pages, 2010.
- [9] M. D. Santos, S. N. Dorogovtsev, and J. F. Mendes, "Biased imitation in coupled evolutionary games in interdependent networks," *Scientific Reports*, vol. 4, pp. 4436–4439, 2014.
- [10] C. D. Brummitt, R. M. D'Souza, and E. A. Leicht, "Suppressing cascades of load in interdependent networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 12, pp. E680–E689, 2012.
- [11] C. M. Schneider, N. Yazdani, N. A. M. Araújo, S. Havlin, and H. J. Herrmann, "Towards designing robust coupled networks," *Scientific Reports*, vol. 3, pp. 1969–1972, 2013.
- [12] R. Parshani, S. V. Buldyrev, and S. Havlin, "Critical effect of dependency groups on the function of networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 3, pp. 1007–1010, 2011.
- [13] J. Shao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, "Cascade of failures in coupled network systems with multiple support-dependence relations," *Physical Review E*, vol. 83, no. 3, Article ID 036116, 9 pages, 2011.
- [14] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.
- [15] Y. Hu, B. Ksherim, R. Cohen, and S. Havlin, "Percolation in interdependent and interconnected networks: abrupt change from second- to first-order transitions," *Physical Review E*, vol. 84, Article ID 066116, 6 pages, 2011.
- [16] X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, "Robustness of interdependent networks under targeted attack," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 6, Article ID 065101, 4 pages, 2011.
- [17] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, "Robustness of a network formed by n interdependent networks with a one-to-one correspondence of dependent nodes," *Physical Review E*, vol. 85, no. 6, Article ID 066134, 13 pages, 2012.
- [18] G. Dong, J. Gao, R. Du et al., "Robustness of network of networks under targeted attack," *Physical Review E*, vol. 87, Article ID 052804, 7 pages, 2013.
- [19] R. Parshani, S. V. Buldyrev, and S. Havlin, "Interdependent networks: reducing the coupling strength leads to a change from a first to second order percolation transition," *Physical Review Letters*, vol. 105, Article ID 048701, 4 pages, 2010.
- [20] P. Erdös and A. Rényi, "On the evolution of random graphs," Publications of the Mathematical Institute of the Hungarian Academy of Sciences Series A, vol. 5, pp. 17–61, 1960.
- [21] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [22] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [23] S. M. Chen, X. Q. Zou, H. Lü, and Q. G. Xu, "Research on robustness of interdependent network for suppressing cascading failure," *Acta Physica Sinica*, vol. 63, no. 2, Article ID 028902, 10 pages, 2014.

- [24] M. Gong and L. Ma, "Enhancing robustness of coupled networks under targeted recoveries," *Science Report*, vol. 5, Article ID 8439, 7 pages, 2015.
- [25] Z. D. Zhao, Y. Liu, and M. Tang, "Epidemic variability in hierarchical geographical networks with human activity patterns," *Chaos*, vol. 22, no. 2, Article ID 023150, 7 pages, 2012.
- [26] J. McNerney, B. D. Fath, and G. Silverberg, "Network structure of inter-industry flows," *Physica A*, vol. 392, no. 24, pp. 6427– 6441, 2013.
- [27] L. P. Chi, R. Wang, H. Su et al., "Structural properties of US flight network," *Chinese Physics Letter*, vol. 20, no. 8, pp. 1393–1396, 2003
- [28] I. B. Utne, P. Hokstad, and J. Vatn, "A method for risk modeling of interdependencies in critical infrastructures," *Reliability Engineering & System Safety*, vol. 96, no. 6, pp. 671–678, 2011.
- [29] G. Fu, R. Dawson, M. Khoury, and S. Bullock, "Interdependent networks: vulnerability analysis and strategies to limit cascading failure," *The European Physical Journal B*, vol. 87, article 148, 10 pages, 2014.
- [30] M. Khoury, S. Bullock, G. Fu, and R. Dawson, "Improving measures of topological robustness in networks of networks and suggestion of a novel way to counter both failure propagation and isolation," *Infrastructure Complexity*, vol. 2, article 1, 20 pages, 2015.
- [31] S. D. Li, L. Li, Y. Jia, X. Liu, and Y. Yang, "Identifying vulnerable nodes of complex networks in cascading failures induced by node-based attacks," *Mathematical Problems in Engineering*, vol. 2013, Article ID 938398, 10 pages, 2013.
- [32] L. Lü, D.-B. Chen, and T. Zhou, "The small world yields the most effective information spreading," *New Journal of Physics*, vol. 13, Article ID 123005, 6 pages, 2011.
- [33] A. E. Motter, "Cascade control and defense in complex networks," *Physical Review Letters*, vol. 93, no. 9, Article ID ID098701, 4 pages, 2004.
- [34] X. Z. Peng, H. Yao, J. Du, Z. Wang, and C. Ding, "Invulnerability of scale-free network against critical node failures based on a renewed cascading failure model," *Physica A: Statistical Mechanics and its Applications*, vol. 421, pp. 69–77, 2015.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 329626, 8 pages http://dx.doi.org/10.1155/2015/329626

Research Article

A Secure Ciphertext Self-Destruction Scheme with Attribute-Based Encryption

Tonghao Yang, Junquan Li, and Bin Yu

Zhengzhou Institute of Information Science and Technology, Zhengzhou 450000, China

Correspondence should be addressed to Tonghao Yang; youngtonghao@163.com

Received 17 June 2015; Revised 29 September 2015; Accepted 5 October 2015

Academic Editor: Mark Leeson

Copyright © 2015 Tonghao Yang 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 secure destruction of expired data is one of the important contents in the research of cloud storage security. Applying the attribute-based encryption (ABE) and the distributed hash table (DHT) technology to the process of data destruction, we propose a secure ciphertext self-destruction scheme with attribute-based encryption called SCSD. In SCSD scheme, the sensitive data is first encrypted under an access key and then the ciphertext shares are stored in the DHT network along with the attribute shares. Meanwhile, the rest of the sensitive data ciphertext and the shares of access key ciphertext constitute the encapsulated self-destruction object (EDO), which is stored in the cloud. When the sensitive data is expired, the nodes in DHT networks can automatically discard the ciphertext shares and the attribute shares, which can make the ciphertext and the access key unrecoverable. Thus, we realize secure ciphertext self-destruction. Compared with the current schemes, our SCSD scheme not only can support efficient data encryption and fine-grained access control in lifetime and secure self-destruction after expiry, but also can resist the traditional cryptanalysis attack as well as the Sybil attack in the DHT network.

1. Introduction

Cloud storage has attracted much attention from both industry and academia for its low cost, flexible deployment, and strong extensibility in recent years. The cloud storage system is composed of massive storage resource on the Internet as well as the resource management and access control mechanism for the resource accessing transparency of users [1]. With friendly user interface and strong extensibility, the cloud storage system can provide users with unlimited storing space; thus, it can form a new delivery model called storage as a service [2]. Cloud storage brings new opportunities for efficiency increasing, cost saving, and green computing in the area of information technology; however, it is also faced with some security challenges.

In the service model of cloud storage, data is outsourced to the storage server which performs as the third party. So, data is out of the control of data owner and the security of data highly depends on the server. Due to the dishonesty of cloud storage server, the data owner will first encrypt the original

sensitive data and then outsource the ciphertext to the cloud in order to keep the confidentiality of data. The encryption key is kept by the data owner privately. However, even if the data is stored by cloud in the form of ciphertext, there are some security risks. For example, in order to improve the service reliability, the cloud may make several backups for the user's data and distribute them to different storage servers [3]. On this condition, when the data has expired and the owner needs to delete the data from the storage servers, the cloud server may not destruct all the backups of data. Once adversaries get the encryption key and the backups of the ciphertext from cloud, the sensitive data can be recovered and the confidentiality is destroyed. Therefore, the assured destruction of expired data, namely, the thorough deletion and the permanent elimination of ciphertext, is one of the important contents in the research of cloud storage security [4].

In this paper, applying the attribute-based encryption and the distributed hash table (DHT) technology to the process of data destruction in the cloud storage environment, we propose a secure ciphertext self-destruction scheme with attribute-based encryption called SCSD. In SCSD scheme, the sensitive data is first encrypted under an access key, and then the access key is encrypted using an attributebased encryption method. The ciphertext of sensitive data is extracted and transformed in order to get the ciphertext shares, which are stored in the DHT network along with the attribute shares. Meanwhile, the rest of the sensitive data ciphertext and the shares of access key ciphertext constitute the encapsulated self-destruction object (EDO), which is stored in the cloud. When the sensitive data is expired, the nodes in DHT networks can automatically discard the ciphertext shares and the attribute shares, which can make the ciphertext of sensitive data and the access key unrecoverable. Thus, we realize secure ciphertext self-destruction. Compared with the current schemes, our SCSD scheme can resist the traditional cryptanalysis attack as well as the Sybil attack in the DHT network.

The rest of the paper is organized as follows. In Section 2, we introduce some related works of the secure data destruction. Then, in Section 3, we review some preliminaries. Next, we introduce the system and security model and the detailed construction of our SCSD scheme in Section 4. In Section 5, we make an evaluation for the scheme in security analysis and scheme performance. Finally, concluding remarks and future work are given in Section 6.

2. Related Works

In cloud storage system, some data is stored in the servers for a long time, which can be compromised by adversaries, because the data may be backed up by the cloud servers and these backups may still exist after the delete command of users. It is difficult to destruct all the backups in the cloud, and the following works are some attempts to achieve the secure destruction of data.

Perlman is the first to focus on the secure deletion of documents [7]. Perlman designed an unrecoverable system for documents. The encryption key is deleted when it is expired; thus, the document encrypted under this key can not be recovered. However, this system considers only the lifetime of encryption key. Besides, this is a local-centered system and is unfit for the cloud environment. Then, following this idea, FADE [8], one secure overlap cloud storage system built under the existing cloud infrastructure, is developed. This system can assure the deletion of documents and can support different document access policies. Another feasible system is Ephemerizer [9], which needs a trusted server to store and manage the decryption key. In Ephemerizer, the data owner sets the expired time for the decryption key. The trusted server deletes the decryption key once the key is expired. Thus, the ciphertext is unreadable.

The above methods follow the idea of centralized solution, which has some limitations as follows. (1) The key management depends too much on the server. (2) When there is an investigation from government, the administrator needs to give up the right of key management. This condition makes the server no longer trusted. (3) There is a need for additional commands and operations to achieve the assured deletion of data.

In order to solve the problem brought by the centralized destruction scheme, Geambasu et al. propose an interesting data self-destruction system called Vanish [5]. The private data is encrypted under a symmetric key, which is divided into several key shares using threshold secret sharing scheme and then distributed to a large scale DHT P2P network. The nodes in the DHT network will automatically delete the key shares periodically, which will result in the unreadable ciphertext. Thus, it realizes the self-destruction of data and this needs trusted servers or additional operations. Wang et al. improve the Vanish system by extracting and distributing parts of ciphertext to the DHT network [6]. This improvement will resist the traditional cryptanalysis attack and brute-force attack more efficiently.

However, [10] points out that there are Sybil attacks against the Vuze DHT network adopted by Vanish system. Adversaries can get enough key shares to reconstruct the key before the ciphertext is expired. Thus, there are security problems in the schemes of [5, 6]. Besides, these decentralized solutions adopt the symmetric encryption algorithms, which will bring complex key management and distribution problems. To solve these problems, an improved system called SafeVanish is proposed [11]. RSA algorithm is adopted to firstly encrypt the symmetric key in order to resist the Sybil attack. But this system can not support fine-grained access control mechanism. Applying attribute-based encryption algorithm, Xiong et al. [12] firstly propose a secure selfdestruction scheme, which can support fine-grained access control on documents. However, the direct adoption of attribute-based encryption algorithm on documents is not efficient.

Therefore, a secure sensitive data self-destruction scheme, which supports efficient data encryption and key management, fine-grained access control in lifetime and secure self-destruction after expiry, and traditional cryptanalysis attack and Sybil attack resistance, is needed in the cloud storage environment.

3. Preliminaries

3.1. Distributed Hash Table. Distributed hash table (DHT) [13] supports a distributed database storage model. And DHT network is comprised of large-scaled distributed infrastructures in the P2P networks which support the query, storage, retrieval, and management of data without servers. Every node in the DHT network is responsible for a small-scaled routing and can store parts of data. Thus, the whole DHT network realizes an addressing and storing of data. There are many DHT networks in Internet, such as Vuze, Chord, OpenDHT, and Pastry.

The index of every document stored in the DHT network can be expressed as a pair of (K, V). K is denoted as the hash value of name or other descriptive pieces of information of the document; V can be denoted as the IP address or other descriptive pieces of information of the node that stored the document in DHT network. All of the index items compose a large document index hash table. When K is specified, the location of document can be assured through the corresponding relationship.

Every DHT network has the following three important characteristics, which is suitable for constructing data selfdestruction scheme in cloud storage environment:

- (1) Data availability: DHT network can provide reliable distributed storage capacity, which assures the availability of the data stored in the nodes of DHT network in the lifetime. This is the foundation of constructing data self-destruction scheme.
- (2) Automatic data deletion in the nodes in DHT network: nodes in DHT network can automatically remove the old data in order to store the new data periodically. Thus, the data stored in the nodes will be destroyed automatically after expiry, which provides a mechanism for ciphertext self-destruction.
- (3) Large-scaled and global distribution: for example, there are more than one million of active nodes in Vuze network simultaneously, and these nodes are distributed to more than 190 countries all over the world. These completely distributed nodes in DHT network can provide attack resistance capability for self-destruction scheme.

3.2. Attribute-Based Encryption. Attribute-based encryption (ABE), a typical public key cryptography, was firstly proposed by Sahai and Waters in 2005 [14]. In an ABE scheme, the identifier for a user is a set of descriptive attributes rather than a string of characters in identity-based encryption (IBE). Every attribute can be mapped to an element in \mathbb{Z}_p^* using a hash function. The ciphertext and user's key are both associated with the attributes. ABE can support threshold policy of attributes. Namely, if and only if the number of same attributes in both sets of attributes ω and ω^* is greater than or equal to a certain threshold value, a user with a set of attributes ω can decrypt the ciphertext successfully which is encrypted under a set of attributes ω^* .

Specifically, an authority firstly defines a threshold value k and generates the system public key, the length of which is related to the number of attributes in ω^* . Then, the authority generates the private key for user with a set of attributes ω . ω is associated with a random k-1 order polynomial q(x). In a decryption process, if $\omega \cap \omega^* \geq k$, then the user chooses random k attributes in the set $\omega \cap \omega^*$ and reconstructs the encryption key through Lagrange's interpolation on the associated polynomial q(x). Thus, the user can decrypt the ciphertext and get the plaintext.

3.3. Threshold Secret Sharing. Threshold secret sharing scheme was first proposed by Shamir [15]. The main idea is to divide the secret data into n shares and then distribute these shares to n users. If there is k or more than k shares are extracted from these users, then the secret data can be generated. Otherwise, the secret data can not be generated. This method is called (k, n) threshold secret sharing.

Generally, threshold secret sharing scheme can be achieved by using Lagrange's interpolation polynomial. If there is an interpolation polynomial

$$Q(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_2x^2 + a_1x^1 + a_0$$
 (1)

and there are n different points $(x_0, y_0), \ldots, (x_i, y_i), \ldots, (x_{n-1}, y_{n-1})$ that satisfy the equation Q(x) = y, then Q(x) is called Lagrange's polynomial, which is composed of the following basic polynomial $Q(x) = \sum_{j=0}^{n-1} y_j q_j(x)$, where $q_j(x) = \prod_{0 \le i \le n-1, i \ne j} ((x - x_i)/(x_j - x_i))$.

Namely, given n different points satisfying Q(x) = y, we can reconstruct a unique n - 1 order polynomial Q(x).

4. SCSD Scheme Construction

In this section, we first describe the system model of the secure ciphertext self-destruction (SCSD) scheme. Then, the detailed algorithm descriptions and the outline of scheme are introduced as follows.

4.1. System Model. The SCSD system comprises six different entities: authority, cloud storage servers, DHT network, data owners, data consumers, and adversaries, as shown in Figure 1.

Authority. Authority provides the system with security parameters setup and key generation processes. Besides, it also assigns attributes for each user.

Cloud Storage Servers. Cloud storage servers are responsible for storing the data sent by the users and assuring that only authenticated users can get access to the data.

DHT Network. Nodes in the DHT network are responsible for storing the ciphertext shares and the attribute shares and can automatically discard the stored data.

Data Owners. A data owner generates sensitive data and then encrypts it under a random access key. Ciphertext shares are sent by data owner to the DHT network along with the attribute shares. Besides, EDO is sent to cloud by data owner.

Data Consumers. The data consumer downloads ciphertext shares and attribute shares from the DHT network and EDO from the cloud. Then, he can decrypt the EDO if his attributes satisfy the ABE threshold policy.

Adversaries. Adversaries may try to capture the data in the cloud or in DHT network.

This paper is aiming at preventing the leakage of sensitive data stored in the cloud after expiry. For example, sensitive information in user's historic archive may leak out in the condition of an investigation from government. We assume that the data owner and other authenticated users trust each other. Thus, adversaries may try to compromise the EDO in the cloud after the lifetime of EDO. Or the adversaries may capture the ciphertext shares and the attribute shares stored in DHT network within the lifetime of EDO. So, in the security model of our scheme, we divide the behavior of adversaries into the following two kinds. (1) Adversaries compromise the EDO in the cloud after the lifetime of EDO. The adversary tries to analyze the sensitive data from the EDO. (2) Adversaries compromise the ciphertext shares and the attribute shares stored in DHT network within the

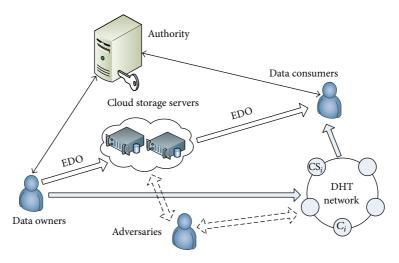


FIGURE 1: The system model of SCSD scheme.

lifetime of EDO. The adversary tries to decrypt the ciphertext and get the sensitive information according to the shares.

4.2. Algorithm Descriptions. Algorithms of our SCSD scheme are described as follows.

(1) Setup(λ) \rightarrow (MSK, PK, USK): given a security parameter λ , the authority firstly generates the master secret parameters MSK = $(r_1, ..., r_n, y)$, which are all chosen randomly from \mathbb{Z}_p^* . Then, the authority generates the public parameters PK = (Attri, g, G_1 , G_2 , e, k, n, t, s, a, b, H, E, Dec, R_1, \ldots, R_n, Y), where Attri is the set of total *n* attributes of users and each attribute in Attri is associated with one unique element in \mathbb{Z}_p^* . G_1 is a multiplicative cyclic group with the generator $g. e: G_1 \times G_1 \rightarrow G_2$ is a bilinear map. k is the threshold value for the total n attributes of users. t is the threshold value for the total s ciphertext shares. a is the number of bits in each associated ciphertext extraction. b is the times of extraction, $H: \{0,1\}^{\frac{1}{n}} \rightarrow \{0,1\}^{m}$ is a hash function. $E:(K,M)\to C$ is a symmetric encryption algorithm and Dec : $(K,C) \rightarrow M$ is the corresponding decryption algorithm. $R_1 = g^{r_1}, \dots, R_n = g^{r_n}, Y = e(g, g)^y$. Besides, the authority also generates secret key for user with attribute set Attri_u. The authority chooses a polynomial q(x)with k-1 degree and sets q(0) = y. Then, the user's secret key is generated as USK = $(S_i)_{i \in Attri_u}$, where $S_i = g^{q(i)/r_i}$.

(2) $\operatorname{Enc}(M) \to (C_M, C_K)$: given sensitive data M, a data owner with an attribute set Attri_o firstly chooses a random access key $K \in \mathbb{Z}_p^*$ and generates the ciphertext of M as $C_M = E(K, M)$. Then, the data owner chooses a random value $\gamma \in \mathbb{Z}_p^*$ and generates the ciphertext of K as $C_K = (\operatorname{Attri}_o, C^* = KY^\gamma, \{C_i = R_i^\gamma\}_{i \in \operatorname{Attri}_o})$, where $\{C_i = R_i^\gamma\}_{i \in \operatorname{Attri}_o}$ are the attribute shares.

(3) Associpher(C_M) \rightarrow (A): given a ciphertext C_M , the data owner firstly divides the ciphertext into blocks of m bits. If the last block is less than m bits, then several bits of "0" are added to the end until the length of the last block is m bits.

Suppose the ciphertext is divided as $A_1 \parallel A_2 \parallel \cdots \parallel A_d$; the data owner associates the blocks as follows:

$$A'_{1} = A_{1} \oplus H (A_{2} \parallel \cdots \parallel A_{d})$$

$$A'_{2} = A_{2} \oplus H (A'_{1} \parallel A_{3} \parallel \cdots \parallel A_{d})$$

$$\vdots$$

$$A'_{i} = A_{i} \oplus H (A'_{1} \parallel \cdots \parallel A'_{i-1} \parallel A_{i+1} \cdots \parallel A_{d})$$

$$\vdots$$

$$A'_{d} = A_{d} \oplus H (A'_{1} \parallel \cdots \parallel A'_{i} \parallel \cdots \parallel A'_{d-1}).$$
(2)

Then, the associated ciphertext is $A = A_1' \parallel \cdots \parallel A_i' \parallel \cdots \parallel A_d'$.

(4) DeAssocipher(A) \rightarrow (C_M): this is the inverse algorithm of Associpher(C_M) \rightarrow (A). Given an associated ciphertext $A = A'_1 \parallel \cdots \parallel A'_i \parallel \cdots \parallel A'_d$, a data consumer performs as follows:

$$A_{d} = A'_{d} \oplus H \left(A'_{1} \parallel \cdots \parallel A'_{i} \parallel \cdots \parallel A'_{d-1} \right)$$

$$A_{d-1} = A'_{d-1} \oplus H \left(A'_{1} \parallel \cdots \parallel A'_{i} \parallel \cdots \parallel A'_{d-2} \parallel A_{d} \right)$$

$$\vdots$$

$$A_{i} = A'_{i} \oplus H \left(A'_{1} \parallel \cdots \parallel A'_{i-1} \parallel A_{i+1} \cdots \parallel A_{d} \right)$$

$$\vdots$$

$$A_{1} = A'_{1} \oplus H \left(A_{2} \parallel \cdots \parallel A_{d} \right).$$
(3)

Then, the data consumer gets the ciphertext C_M from the association $A_1 \parallel A_2 \parallel \cdots \parallel A_d$.

(5) CipherShareGen(A) \rightarrow (CS, C'): given the associated ciphertext A, for $i=1,2,\ldots,b$, the data owner firstly extracts the bits located in $[1,a\cdot t]$ in $A^{(i)}$, where $A^{(i)}$ is the remaining associated ciphertext after the (i-1)th extraction from A. Note that $A^{(1)}=A$. All of the extracted ciphertext is denoted by EC = (m_1,m_2,\ldots,m_b) , where $m_i=m_{[i][0]}\parallel m_{[i][1]}\parallel\cdots\parallel m_{[i][t-1]}$ is the ith extracted ciphertext from A. The remaining associated ciphertext after the bth extraction from A is denoted by C'. Then, the data owner generates b polynomials as follows:

$$f_{1}(x) = m_{[1][t-1]}x^{t-1} + m_{[1][t-2]}x^{t-2} + \dots + m_{[1][1]}x^{1} + m_{[1][0]}$$

$$\vdots$$

$$f_{i}(x) = m_{[i][t-1]}x^{t-1} + m_{[i][t-2]}x^{t-2} + \dots + m_{[i][1]}x^{1} + m_{[i][0]}$$

$$\vdots$$

$$f_{b}(x) = m_{[b][t-1]}x^{t-1} + m_{[b][t-2]}x^{t-2} + \dots + m_{[b][1]}x^{1}$$

The data owner chooses s different integers x_1, x_2, \ldots, x_s and then computes the value of $f_1(x_i), f_2(x_i), \ldots, f_b(x_i)$ for $i = 1, 2, \ldots, s$. Finally, the data owner gets the ciphertext shares $CS = (CS_1, CS_2, \ldots, CS_s)$, where $CS_i = (x_i, f_1(x_i), f_2(x_i), \ldots, f_b(x_i))$ for $i = 1, 2, \ldots, s$.

- (6) ShareDistribute(CS, C_K) \rightarrow (CI, AJ): given the ciphertext shares CS and the attribute shares $\{C_i = R_i^{\gamma}\}_{i \in \text{Attri}_o}$ from C_K , the data owner firstly chooses a random index CI for CS as a seed to a pseudorandom number generator. Then, the data owner runs the generator to generate s indices I_1, I_2, \ldots, I_s . For $i = 1, 2, \ldots, s$, each ciphertext share CS_i is stored in the node indexed by I_i in the DHT network. Similarly, for the attribute shares $\{C_i = R_i^{\gamma}\}_{i \in \text{Attri}_o}$ from C_K , the data owner firstly chooses a random index AJ as a seed to a pseudorandom number generator. Then, the data owner runs the generator to generate n indices I_1, I_2, \ldots, I_n . For $i = 1, 2, \ldots, n$, each attribute share C_i is stored in the node indexed by I_i in the DHT network.
- (7) EDOGen(Attri_o, C^* , C', CI, AJ) \rightarrow (EDO): given the attribute set of the data owner Attri_o, C^* from C_K , C', CI, and AJ, the data owner generates the encapsulated self-destruction object EDO = (Attri_o, C^* , C', CI, AJ) and then sends the EDO to the cloud.
- (8) KeyRecover(EDO, USK) \rightarrow (K): before the expiration timestamp of EDO, a data consumer, with a secret key USK and an attributes set Attri_c , firstly gets the EDO from the cloud. Then, the data consumer runs the pseudorandom number generator to generate n indices J_1, J_2, \ldots, J_n of attribute shares $\{C_i = R_i^{\gamma}\}_{i \in \operatorname{Attri}_o}$ under the seed AJ. Then, the data consumer gets as many $C_i = R_i^{\gamma}$, $i \in \operatorname{Attri}_o$, as possible from the DHT network according to the indices J_1, J_2, \ldots, J_n .

In order to recover the access key K, the data consumer chooses a set of k attribute shares $\mathrm{Att} \in \mathrm{Attri}_o \cap \mathrm{Attri}_c$. Note that if there are no more than k attribute shares in the set of $\mathrm{Attri}_o \cap \mathrm{Attri}_c$, the data consumer can not recover the access key K since he can not satisfy the ABE threshold policy. If there is a set of attribute shares Att , the data consumer firstly gets Lagrange's coefficient $\Delta_{i,\mathrm{Att}}(x) = \prod_{i,j \in \mathrm{Att}, i \neq j} ((x-j)/(i-j))$ and then recovers the access key as follows:

$$\frac{C^*}{\prod_{i \in Att} \left(e\left(S_i, C_i\right)\right)^{\Delta_{i,Att}(0)}}$$

$$= \frac{KY^{\gamma}}{\prod_{i \in Att} \left(e\left(g^{q(i)/r_i}, g^{r_i \gamma}\right)\right)^{\Delta_{i,Att}(0)}}$$

$$= \frac{Ke\left(g, g\right)^{\gamma \gamma}}{\prod_{i \in Att} \left(e\left(g, g\right)^{\gamma \gamma i}\right)^{\Delta_{i,Att}(0)}} = \frac{Ke\left(g, g\right)^{\gamma \gamma}}{e\left(g, g\right)^{\gamma q(0)}} = K.$$
(5)

- (9) PlainRecover(EDO, K) \rightarrow (M): given the EDO from the cloud, the data consumer runs the pseudorandom number generator to generate s indices I_1, I_2, \ldots, I_s of ciphertext shares $CS = (CS_1, CS_2, \ldots, CS_s)$ under the seed CI. Then, the data consumer gets more than t-1 CS_i , $i=1,2,\ldots,s$, from the DHT network. From these CS_i , $i=1,2,\ldots,s$, the data consumer can reconstruct the polynomials $f_1(x), f_2(x), \ldots, f_b(x)$ using Lagrange's interpolation. Then, the data consumer gets $EC = (m_1, m_2, \ldots, m_b)$ from these polynomials and generates the associated ciphertext A. Finally, the original ciphertext C_M is generated by running DeAssocipher(A) \rightarrow (C_M) algorithm. The plaintext is recovered from $M = Dec(K, C_M)$.
- 4.3. Outline of SCSD Scheme. There are two main phases of SCSD scheme, namely, the data encapsulation phase and the data reconstruction phase. The outline of SCSD scheme is illustrated in Figure 2.

In data encapsulation phase (Phase I), the data owner firstly runs the algorithm $\mathrm{Enc}(M) \to (C_M, C_K)$ to generate the ciphertext of sensitive data under ABE. Then, the data owner runs the algorithms $\mathrm{Associpher}(C_M) \to (A)$, $\mathrm{CipherShareGen}(A) \to (\mathrm{CS}, C')$, and $\mathrm{ShareDistribute}(\mathrm{CS}, C_K) \to (\mathrm{CI}, \mathrm{AJ})$ in turn to get the ciphertext shares and attribute shares and then distributes the shares to the DHT network. Besides, the data owner runs the algorithm $\mathrm{EDOGen}(\mathrm{Attri}_o, C^*, C', \mathrm{CI}, \mathrm{AJ}) \to (\mathrm{EDO})$ to get the EDO and then sends the EDO to the cloud.

In data reconstruction phase (Phase II), the data consumer firstly runs the algorithm KeyRecover(EDO, USK) \rightarrow (K) to generate the access key of ciphertext before the EDO expires. Note that if the data consumer does not satisfy the ABE threshold policy defined by the data owner, he can not recover the access key successfully. Then, the data consumer runs the algorithm PlainRecover(EDO, K) \rightarrow (M) to get the ciphertext and finally recovers the sensitive data.

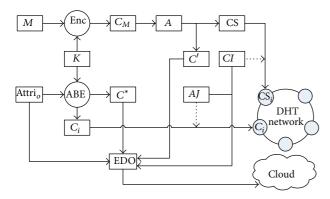


FIGURE 2: The outline of SCSD scheme.

5. Analysis and Performance

In this section, we evaluate our SCSD scheme by modularizing it into two parts, namely, security analysis and scheme performance.

5.1. Security Analysis. In the applications of our scheme, because adversaries can not specify the particular object of attack before the expiration timestamp, we assume that the copies of EDO stored in the cloud are secure during this time. Besides, because the attribute shares and ciphertext shares stored in the DHT network will be discarded after the expiry of EDO, once the DHT network is updated periodically, the contents of EDO copies will be unreadable.

There are mainly two kinds of attack aiming at our scheme. The first one is cracking the expired EDO copies stored in the cloud through cryptanalysis attack and brute-force attack. Despite the fact that the attribute shares and ciphertext shares are discarded, there are still EDO copies stored in the cloud. The other kind of attack is aiming at collecting the attribute shares and ciphertext shares in the DHT network before the expiration timestamp of EDO, and these shares will be used in the tracing attack against the EDO copies stored in the cloud.

Therefore, the security of our scheme is mainly affected by two aspects. One is the security of encryption algorithm used in the sensitive data encryption under the access key, which depends on the capability of resisting the cryptanalysis attack and brute-force attack. The other is the security of DHT network that stored the attributes shares and ciphertext shares, which depends on the capability of resisting sniffing attack, hopping attack, and other DHT Sybil attacks. So, we make the security analysis of our scheme based on these two aspects as follows. The brief comparisons of security properties of our SCSD scheme [5, 6] are summarized in Table 1.

5.1.1. The Security of Encryption Algorithm. The brute-force attack is implemented by trying any possible decryption keys on the ciphertext to recover the plaintext. This kind of attack is based on the integrity of ciphertext. So, adversaries should first get the integrated ciphertext before implementing the

Table 1: Comparisons of security properties.

Characteristics	Scheme			
Cital acteristics	SCSD	Reference [5]	Reference [6]	
Key destruction			√	
Ciphertext destruction		×	\checkmark	
Cryptanalysis attack resistance	e √	×	\checkmark	
Brute-force attack resistance		×	\checkmark	
Sybil attack resistance		×	×	
Asymmetric encryption		×	×	
Fine-grained access control	$\sqrt{}$	×	×	
Simple key management	$\sqrt{}$	×	×	

brute-force attack. In our scheme, however, the sensitive data is first encrypted under the random access key and then the ciphertext is associated and extracted. Because every block of the associated ciphertext is correlated with each other, once some of the blocks are extracted, the remaining blocks will be no more integrated. Therefore, without the integrated ciphertext, adversaries can not recover the sensitive data by the brute-force attack.

Besides, implementing the traditional cryptanalysis attack is also based on an integrated ciphertext. Because the remaining ciphertext blocks stored in the cloud are incomplete, the traditional cryptanalysis attack had no effect on our scheme.

5.1.2. The Security of DHT Network. In the following, we will discuss whether adversaries can crack the EDO copies by attacking the DHT network before the expiration timestamp of EDO. Because adversaries can not specify the particular object of attack before the expiration timestamp, the adversaries may try to get as many attribute shares and ciphertext shares as possible during this time. For example, the adversaries may keep on attacking the DHT network in order to get enough shares. However, this kind of attack will bring expensive cost to the adversaries.

Due to the characteristic of DHT network, the method of attacking the DHT network to get the attribute shares and ciphertext shares is very difficult. Reference [5] has made detailed analysis aiming at various kinds of DHT attacks by performing simulations in the Vuze DHT network. The result shows that it is impossible for the adversaries to get enough shares from DHT network by implementing sniffing attack, hopping attack, and other DHT attacks. Therefore, in the same way, the adversaries in our scheme also can not get enough attribute shares or ciphertext shares by attacking the DHT network in order to crack the EDO copies stored in the cloud.

5.2. Performance and Optimization. In this section, we first make a performance evaluation of SCSD on the time cost in both the data encapsulation phase and the data reconstruction phase, respectively. Then, we implement the parameter optimization by analyzing the tradeoff between security and availability of our scheme.

5.2.1. Performance Evaluation. In Phase I, the communication overhead is mainly caused by the distribution of ciphertext shares and attribute shares to the DHT network. The computation overhead is mainly caused by the ABE algorithm on the access key, the symmetric encryption algorithm on sensitive data, and the association and the shares generation algorithm on ciphertext. In Phase II, the communication overhead is also mainly caused by the collection of ciphertext shares and attribute shares from the DHT network. The computation overhead is mainly caused by the reconstruction of the access key and the ciphertext.

Based on the above analysis, we execute our SCSD scheme and measure the times spent in the two main phases. For the sake of simplicity, we set the total shares n=s and the threshold k=t for the ciphertext shares and attribute shares, respectively. The evaluation uses an Intel G2130 3.2 GHz with 4 GB of RAM, Java 1.6, and a broadband network. The times of the two main phases are shown in Figure 3.

Figure 3 shows that the data collection and reconstruction phase is relatively fast. The time cost of data encapsulation and distribution, however, is quite large. Fortunately, a simple pretreatment, pregenerating the access key and prepushing shares into the DHT network, can be implemented. As shown in Figure 3, this pretreatment can lead the time of data encapsulation phase to a fixed 1.6 s. Thus, the performance of SCSD scheme is relatively effective and efficient.

5.2.2. Parameter Optimization. Next, we assume that the adversaries have comprised 5% of the nodes in a thousand-node DHT network. We will show how the security and the availability of our scheme are affected by the parameters n and the threshold k. The probability that an adversary captures sufficient shares to reconstruct the ciphertext shares is shown in Figure 4. It is clear that increasing the number of shares can decrease the adversary's success probability. Furthermore, the security can also be enhanced as the threshold increases.

As shown in Figure 5, the availability is also affected by the parameters. The maximum timeout gets longer as the number of shares increases. And longer timeout can also be supported by smaller threshold since the scheme can tolerate more share loss. So, the choice of threshold can represent a tradeoff between security and availability. High threshold can provide more security and low threshold can provide longer lifetime. Therefore, by choosing the proper share number and threshold, we can get a tradeoff of high security and good availability.

Besides the parameters, there are other kinds of optimizations for our scheme. Because of the adoption of ABE algorithm, our SCSD scheme can implement one-to-many authorization and access control flexibly. Moreover, the access key can be used repeatedly in the condition of timely processing huge volume of data while the security requirement is lower. And if the requirement of security is higher, the ciphertext shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute of the ciphertext shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute of the ciphertext shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute of the ciphertext shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and the attribute shares $CS = (CS_1, CS_2, ..., CS_s)$ and $CS = (CS_1, CS_2, ..., CS_s)$

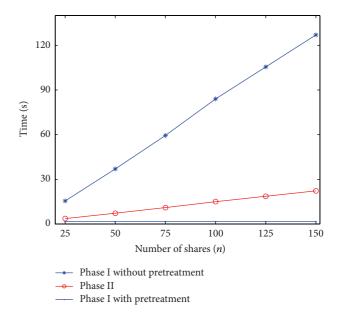


FIGURE 3: Performance of SCSD scheme.

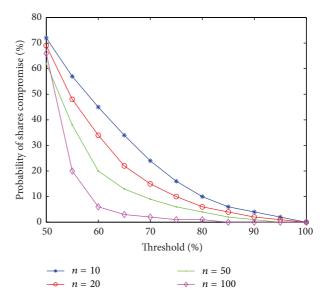


FIGURE 4: Parameters and security.

6. Conclusion

In cloud storage system, secure data destruction is one of the problems that need to be addressed in data security. Many data destruction schemes have been proposed in recent years. However, there are still some limitations. In this paper, we mainly focus on the ciphertext destruction and propose a secure ciphertext self-destruction scheme with attribute-based encryption called SCSD, which applies the attribute-based encryption and the distributed hash table technology to the process of data destruction in the cloud

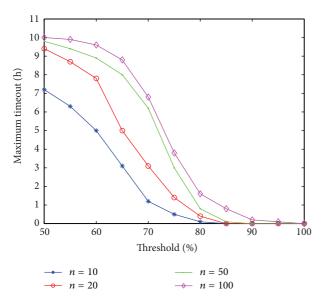


FIGURE 5: Parameters and availability.

storage environment. Compared with the current schemes, our scheme can resist the traditional cryptanalysis attack as well as the Sybil attack in the DHT network. Besides, the performance of SCSD scheme is relatively effective and efficient.

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the School Innovation Foundation and the Doctorial Foundation under Grant 2014JY170. The authors thank the anonymous reviewers for their useful comments and suggestions.

References

- [1] M. Armbrust, A. Fox, R. Griffith et al., "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.
- [2] D. P. Shah and P. A. Ganatra, "Comparative study of data possession techniques for data storage as a service (DSaaS)," *International Journal of Computer Applications*, vol. 80, no. 4, pp. 38–42, 2013.
- [3] D. Catteddu, "Cloud computing: benefits, risks and recommendations for information security," in *Web Application Security*, p. 17, Springer, 2010.
- [4] L. M. Kaufman, "Data security in the world of cloud computing," *IEEE Security and Privacy*, vol. 7, no. 4, pp. 61–64, 2009.
- [5] R. Geambasu, T. Kohno, A. Levy et al., "Vanish: increasing data privacy with self-destructing data," in *Proceedings of the 18th USENIX Security Symposium*, pp. 299–315, Montreal, Canada, August 2009.

- [6] G. Wang, F. Yue, and Q. Liu, "A secure self-destructing scheme for electronic data," *Journal of Computer and System Sciences*, vol. 79, no. 2, pp. 279–290, 2013.
- [7] R. Perlman, "File system design with assured delete," in Proceedings of the 3rd IEEE International Security in Storage Workshop (SISW '05), pp. 83–88, IEEE, San Francisco, Calif, USA, December 2005.
- [8] Y. Tang, P. P. C. Lee, J. C. S. Lui, and R. Perlman, "FADE: secure overlay cloud storage with file assured deletion," in Security and Privacy in Communication Networks: 6th Iternational ICST Conference, SecureComm 2010, Singapore, September 7–9, 2010. Proceedings, vol. 50 of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pp. 380–397, Springer, Berlin, Germany, 2010.
- [9] R. Perlman, "The Ephemerizer: making data disappear," *Journal of Information System Security*, vol. 1, no. 1, pp. 21–32, 2005.
- [10] S. Wolchok, S. O. Hofmann, N. Heninger et al., "Defeating vanish with low-cost Sybil attacks against large DHT," in *Proceedings of the 17th Annual Network & Distributed System Security Conference (NDSS '10)*, pp. 1–15, San Diego, Calif, USA, February 2010.
- [11] L. Zeng, Z. Shi, S. Xu, and D. Feng, "SafeVanish: an improved data self-destruction for protecting data privacy," in *Proceedings* of the IEEE 2nd International Conference on Cloud Computing Technology and Science, pp. 521–528, IEEE, Indianapolis, Ind, USA, December 2010.
- [12] J. Xiong, Z. Yao, J. Ma et al., "A secure document self-destruction scheme: an ABE approach," in *Proceedings of the 15th IEEE International Conference on High Performance Computing and Communications (HPCC '13)*, pp. 59–64, Zhangjiajie, China, November 2013.
- [13] D. Frank, A Distributed Hash Table, Massachusetts Institute of Technology, 2005.
- [14] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in Advances in Cryptology—EUROCRYPT 2005: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22–26, 2005. Proceedings, vol. 3494 of Lecture Notes in Computer Science, pp. 457–473, Springer, Berlin, Germany, 2005.
- [15] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.
- [16] S. Rhea, B. Godfrey, B. Karp et al., "OpenDHT: a public DHT service and its uses," *Computer Communication Review*, vol. 35, no. 4, pp. 73–84, 2005.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 212050, 11 pages http://dx.doi.org/10.1155/2015/212050

Research Article

Predicting Free Flow Speed and Crash Risk of Bicycle Traffic Flow Using Artificial Neural Network Models

Cheng Xu,^{1,2} Qiangwei Li,² Zhaowei Qu,¹ and Sheng Jin³

¹College of Transportation, Jilin University, Changchun 130022, China

Correspondence should be addressed to Sheng Jin; jinsheng@zju.edu.cn

Received 12 June 2015; Accepted 29 September 2015

Academic Editor: Mark Leeson

Copyright © 2015 Cheng Xu 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.

Free flow speed is a fundamental measure of traffic performance and has been found to affect the severity of crash risk. However, the previous studies lack analysis and modelling of impact factors on bicycles' free flow speed. The main focus of this study is to develop multilayer back propagation artificial neural network (BPANN) models for the prediction of free flow speed and crash risk on the separated bicycle path. Four different models with considering different combinations of input variables (e.g., path width, traffic condition, bicycle type, and cyclists' characteristics) were developed. 459 field data samples were collected from eleven bicycle paths in Hangzhou, China, and 70% of total samples were used for training, 15% for validation, and 15% for testing. The results show that considering the input variables of bicycle types and characteristics of cyclists will effectively improve the accuracy of the prediction models. Meanwhile, the parameters of bicycle types have more significant effect on predicting free flow speed of bicycle compared to those of cyclists' characteristics. The findings could contribute for evaluation, planning, and management of bicycle safety.

1. Introduction

Traffic safety and crash risk of both motorized vehicles and bicycles are the high-priority issues to traffic engineers and researchers [1-4]. Recently, with the rapid growth of bicycles (including classic bicycles and electric bicycles) in developing countries such as Vietnam, Malaysia, Indonesia, and China, there have been many efficiency and safety problems for bicycle traffic flow. Although there are many significant environmental, climate, congestion, and public health benefits for cycling, bicycle crash is still a serious issue [5]. According to the statistical data from the Ministry of Public Security in China [6], the percentages of deaths and injuries of cyclists in all travel modes have been increasing, up to around 15% and 17%, respectively. In 2012, there were nearly 9,000 people who died in bicycle traffic crashes in China. Therefore, the improvement of bicycle safety is very important and urgent for both traffic engineers and researchers.

Speed is a fundamental measure of traffic performance of a highway system and can be widely used to describe the condition of the traffic flow and as an input for travel time, delay, and level of service determination [7]. Meanwhile, speed is also an important factor in road safety. There have been lots of studies having found that speed not only affects the severity of a crash but is also related to the risk of being involved in a crash [8]. There is a strong relationship between crash risk of motorized vehicles and speed under free flow conditions [9]. Similar conclusions can also be found for bicycle traffic flow [10]. Therefore, modelling and analysis of impact factors on bicycle free flow speed or crash risk are very useful and will provide the basis for improved bicycle traffic safety.

The previous studies on bicycle speed focus on the determination of bicycle free flow speed and speed distribution. Liu et al. [11] reported the mean of observed bicycle free flow speed was approximately 14 kph. Wei et al. [12] reported that the peak-hour free flow speeds of bicycle with and without separated barrier are 18.2 kph and 13.9 kph, respectively. According to Allen et al. [13], the bicycle free flow speed appears to be somewhere between 10 kph and 28 kph, with the majority of the observations being between 12 kph and 20 kph. Cherry [14] found free flow speeds of

²Department of Traffic Management Engineering, Zhejiang Police College, Hangzhou 310053, China

³College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China

Parameters types	Designation		Max	Min	Mean	SD
Cycleway features	Cycleway width (m)	CW	4.60	2.27	3.46	0.74
Traffic flow parameters	Bicycle flow (bicycles/h/m)	FB	1364	72	590	302
Bicycle types	% of BSEB	PBS	42.31%	3.75%	16.78%	5.79%
	% of SSEB	PSS	79.59%	29.61%	53.11%	10.56%
	% of male cyclists	PMC	92.75%	43.54%	65.64%	8.48%
Characteristics of cyclists	% of young cyclists	PYC	92.02%	22.50%	64.14%	14.13%
	% of middle-aged cyclists	PMAC	62.38%	4.80%	28.10%	11.82%
	% of loaded cyclists	PLC	30.00%	1.32%	11.19%	5.67%

TABLE 1: Descriptive statistics of model parameters.

bicycles in Shanghai were 18.2 kph and 13.0 kph for electric bicycles and classic bicycles, respectively; free flow speeds in Kunming were similar, at 17.9 kph for electric bicycles and 12.8 kph for classic bicycles. Lin et al. [15] found the free flow speeds for both electric bicycles and classic bicycles in Kunming were 21.86 kph and 14.81 kph, respectively. Similar results (21.86 kph for electric bicycles and 14.81 kph for classic bicycles) have also been found in Hangzhou by Jin et al. [16]. In terms of bicycle speed distribution, Dey et al. [17] proposed a speed distribution curve model under mixed traffic conditions, including both fast-moving vehicles (e.g., cars/jeeps, trucks/buses, two-wheelers, and three-wheelers) and slow-moving vehicles (e.g., bicycles and tractors). Lin et al. [15] used the lognormal distribution to fit the heterogeneous bicycle speed data. Wang et al. [18] analysed the impact of various factors on the speed of heterogeneous bicycle flow and used normal distribution to fit the bicycle speed samples.

Most studies emphasize modelling the relationships between free flow speed and such factors as geometric features, traffic characteristics, traffic control, environmental features, weather conditions, and driver's experience and characteristics [19–23]. However, the majority of existing models are only applicable to predict the speed for cars [23]. The impact factors on free flow speed for motorized vehicles are significantly different compared to bicycle traffic. To the best of our knowledge, there was little research focus on modelling the affecting factors on bicycles' free flow speed. The authors believe that this research would be helpful in evaluating and improving the safety of bicycle traffic flow, particularly at high heterogeneous bicycle flow locations.

The contribution of this paper is to develop artificial neural network (ANN) models to predict free flow speed for bicycle traffic with considering some impact factors. Four different models, namely, Model 1, Model 2, Model 3, and Model 4, were developed considering different categories of contributing factors. The characteristics of different models were analysed and compared. It is expected that the developed models may be useful for future prediction of bicycles' free flow speed or crash risk under different cycleway features, traffic conditions, bicycle types, and/or characteristics of cyclists.

2. Data Collection

2.1. Model Parameters. Selection of model parameters is a critical task to model the relationship between bicycle

free flow speed and its contributing factors. Based on the literature review and analysis of bicycle traffic flow [16], the input parameters of the proposed models could be divided into the following four groups: cycleway features (e.g., cycleway width, pavement condition, and geometric feature), traffic conditions (e.g., flow, speed, and density), bicycle types (e.g., electric bicycles), and characteristics of cyclists (e.g., age, gender, and alcohol consumption). Because the cycleway conditions are good and separated with motorized vehicle by barriers at the survey sites, the pavement conditions and geometric features have little effect on cyclists. Therefore, only cycleway width (CW) was considered in the proposed models. The traffic conditions category only includes bicycle flow per hour per meter (BF). Bicycle types in China typically consist of three categories: classic bicycle (CE), bicycle-style-electric-bicycle (BSEB), and shooterstyle-electric-bicycle (SSEB). The bicycle type parameters hence include percentage of BSEBs (PBS) and percentage of SSEBs (PSS). Considering the difficulty of bicycle data collection, four characteristic parameters of cyclists including percentage of male cyclists (PMC), percentage of young cyclists (PYC), percentage of middle-aged cyclists (PMAC), and percentage of loaded cyclists (PLC) were selected in this paper. The selected input parameters of models are listed in Table 1.

2.2. Data Survey. Field bicycle data used in this study were collected from eleven cycleways in Hangzhou, China. The widths of cycleway range from 2.27 to 4.60 m. All of the survey sites are straight and low gradient, located at least 100 meters away from intersection, and separated with motorized vehicle lane. The cameras were set up on the roadside of the cycleway to record the operation of bicycle traffic. Video surveillance application could record the movement of bicycles and the flow and speed can be automatically calculated. The other parameters (e.g., bicycle type and age and gender of cyclists) could be recorded and coded manually. In this paper, bicycle type consists of three categories: CE, BSEB, and SSEB. Cyclists' genders are easily distinguished and recorded by investigator. According to cyclists' age, the cyclists were divided into three groups: the young (under 40), the middleaged (between 40 and 60), and the elderly (over 60). The loaded cyclist means a cyclist who is carrying something (including an object or a person) on his/her bicycle. From the collected bicycle data, the descriptive statistics of model parameters can be found in Table 1. From the table, it can

Model parameters	3	Model 1	Model 2	Model 3	Model 4
Cycleway features	WB	√	√	√	√
Traffic flow parameters	FB	√	√	√	√
Bicycle types	PSS		√		√
	PBS		\checkmark		\checkmark
	PMC			√	√
Characteristics of cyclists	PYC			\checkmark	\checkmark
	PMAC			\checkmark	\checkmark
	PLC			\checkmark	\checkmark

Table 2: Selected input parameters for different ANN models.

be found that average bicycle flow is 590 bicycles/h/m, and the average percentages of BSEB, SSEB, male, young, middle-aged, and loaded cyclists are 16.78%, 53.11%, 65.64%, 64.14%, 28.10%, and 11.19%, respectively. Each parameter has wide range and is suitable for modelling the relationships with free flow speed of bicycle.

2.3. Estimation of Free Flow Speed for Bicycle Traffic. The free flow speed of bicycle flow is the speed of bicycles under low volumes and low densities and is the most important parameter for cycleway capacity estimation, LOS, and speed limit. Because it is difficult to determine which traffic conditions are of low volumes and densities, the 85th percentile speed is usually used as the free flow speed [24]. The 85th percentile speed of bicycle is the speed below which 85 percent of cyclists travel and is the most frequently used for speed limit design. The TRB special report also shows that the 85th percentile speed is an important descriptive statistic in evaluating road safety [25]. Therefore, in this study, we use the 85th percentile speed of bicycle flow as the free flow speed and the crash risk indicator for the evaluation of bicycle safety.

3. Artificial Neural Network Models

Artificial neural networks are a family of statistical learning models inspired by biological neural networks and are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown [26]. ANN models are widely used in modelling free flow speed of motorized vehicles [27–31].

3.1. Description of Models. An ANN model can include multiple input variables to predict multiple output variables. In this study, four ANN models were developed with or without considering different input variables such as bicycle types and characteristics of cyclists. The purpose of including different input variables in the modelling of bicycles' free flow speed and crash risk is to analyse and assess whether or not the selection of input variables will affect the performance of the developed ANN models.

Table 2 lists four categories input parameters used for each model. Cycleway features and traffic flow parameters have been proved to have important effect on free flow speed; hence, these two parameters are always included in four models. Model 2 and Model 3 include the input parameters of bicycle types and characteristics of cyclists,

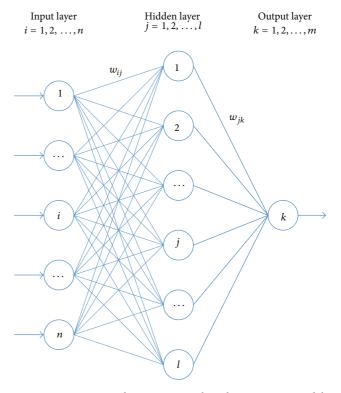


Figure 1: Structure and notations in a three-layer BPANN model.

respectively; meanwhile, Model 4 selects both bicycle types and characteristics of cyclists as the input variables. The tick marks (\checkmark) in Table 2 indicate that the input parameters are included in the modelling process as input variables. The developed four ANN models will provide selectivity and flexibility for considering suitable input variables in the prediction of free flow speed of bicycle flow.

3.2. Network Architectures. A back propagation ANN (BPANN) model, as shown in Figure 1, was introduced for modelling bicycle free flow speed. BPANN model is one of the most well-known ANN models applied in many areas [26]. The goal and motivation for developing the backpropagation algorithm are to find a way to train a multilayered neural network such that it can learn the appropriate internal representations to allow it to learn any arbitrary mapping of input to output.

The three-layer BPANN architecture of this study is listed in Figure 1. Multilayer BPANN is a layered parallel processing system consisting of input layer, output layer, and hidden layer [32]. According to Figure 1, i, j, and k are subscripts for input, hidden, and output layers, respectively. The number of input and output parameters and hidden nodes is n, m, and l, respectively. The number of nodes in input and output layers (n and m) corresponds to the number of input variables and output variables. The number of nodes in the hidden layer (l) should be determined by the network designer and number of input and output variables. The weight factors for hidden layer and output layer are ω_{ij} and ω_{jk} , respectively. The values n and m are problem dependent. In this study, the values of n are 2, 4, 6, and 8 for Models 1, 2, 3, and 4. The output parameter is free flow speed, and thus the value m is one.

The number of nodes in hidden layer has significant effect on the performance of BPANN models. According to the previous research, the number of nodes in hidden layer should meet the following conditions:

$$l = \sqrt{n+m} + a,\tag{1}$$

where a is an integer between 0 and 10. Because the four models have different numbers of input variables, we use the same number of nodes in hidden layer for comparison. Therefore, in this paper, the number of nodes in hidden layer was set as 10 for all of models.

- 3.3. BPANN Algorithm Process. The backpropagation learning algorithm for ANN can be divided into two phases: propagation and weight update. The detailed algorithm processes are listed as follows.
- 3.3.1. Initialization of BPANN. The input variables are expressed as $\mathbf{x} = (x_1, x_2, \dots, x_n)$; the input vector in hidden layer is $\mathbf{hi} = (hi_1, hi_2, \dots, hi_l)$; the output vector in hidden layer is $\mathbf{ho} = (ho_1, ho_2, \dots, ho_l)$; the input vector in output layer is $\mathbf{yi} = (yi_1, yi_2, \dots, yi_m)$; the output vector in output layer is $\mathbf{yo} = (yo_1, yo_2, \dots, yo_m)$; and the expected output result is $\mathbf{do} = (d_1, d_2, \dots, d_m)$. The error function of each output sample is expressed as

$$e = \frac{1}{2} \sum_{k=1}^{m} (do_k - yo_k)^2,$$
 (2)

where *e* is the error of each output sample.

Interconnecting weights were assigned some random numbers. The given precision ε and maximum learning number M are set.

3.3.2. Calculating the Inputs and Outputs in Hidden Layer. The input and output values in hidden layer could be calculated by the following equations:

$$hi_j = \sum_{i=1}^n \omega_{ij} x_i - b_j$$
 $j = 1, 2, ..., l$
 $ho_j = f(hi_j)$ $j = 1, 2, ..., l$

$$yi_k = \sum_{i=1}^{l} \omega_{jk} ho_j - b_o \quad k = 1, 2, ..., m$$

 $yo_k = f(yi_k) \quad k = 1, 2, ..., m,$
(3)

where b_h and b_o are the critical values of neurons in hidden layer and in output layer, respectively. $f(\cdot)$ is the logarithmic sigmoid transfer function which is represented by the following equation:

$$f(x) = \frac{1}{1 + e^{-x}} \quad 0 < f(x) < 1. \tag{4}$$

3.3.3. Calculating Partial Derivative of Error Function. The partial derivative of the error function for each neuron in output layer can be expressed as follows:

$$\frac{\partial e}{\partial \omega_{jk}} = \frac{\partial e}{\partial y i_k} \frac{\partial y i_k}{\partial \omega_{jk}}$$

$$\frac{\partial y i_k}{\partial \omega_{jk}} = \frac{\partial \left[\sum_{j=1}^l \omega_{jk} h o_j - b_o\right]}{\partial \omega_{jk}} = h o_j$$

$$\frac{\partial e}{\partial y i_k} = \frac{\partial \left[\left(1/2\right) \sum_{k=1}^m \left(d o_k - y o_k\right)^2\right]}{\partial y i_k}$$

$$= -\left(d o_k - y o_k\right) y o_k'$$

$$= -\left(d o_k - y o_k\right) f'\left(y i_k\right) - \delta_k.$$
(5)

The partial derivative of the error function for each neuron in hidden layer can be also expressed as follows:

$$\frac{\partial e}{\partial \omega_{ij}} = \frac{\partial e}{\partial h i_{j}} \frac{\partial h i_{j}}{\partial \omega_{ij}}$$

$$\frac{\partial h i_{j}}{\partial \omega_{ij}} = \frac{\partial \left[\sum_{i=1}^{n} \omega_{ij} x_{i} - b_{j}\right]}{\partial \omega_{ij}} = x_{i}$$

$$\frac{\partial e}{\partial h i_{j}} = \frac{\partial \left[\left(1/2\right) \sum_{k=1}^{m} \left(do_{k} - yo_{k}\right)^{2}\right]}{\partial ho_{j}} \frac{\partial ho_{j}}{\partial hi_{j}}$$

$$= -\left(\sum_{k=1}^{m} \delta_{k} \omega_{jk}\right) f'\left(hi_{j}\right) - \delta_{j}.$$
(6)

3.3.4. Adjustment of Interconnecting Weights. The adjustment of interconnecting weights for hidden layer and output layer is expressed as

$$\Delta\omega_{jk} = -\eta \frac{\partial e}{\partial \omega_{jk}} = \eta \delta_k h o_j$$

$$\omega_{jk}^{N+1} = \omega_{jk}^N + \eta \delta_k h o_j$$

$$\Delta\omega_{ij} = -\eta \frac{\partial e}{\partial \omega_{ij}} = -\eta \frac{\partial e}{\partial h i_j} \frac{\partial h i_j}{\partial \omega_{ij}} = \eta \delta_j x_i$$

$$\omega_{ij}^{N+1} = \omega_{ij}^N + \eta \delta_j x_i,$$
(7)

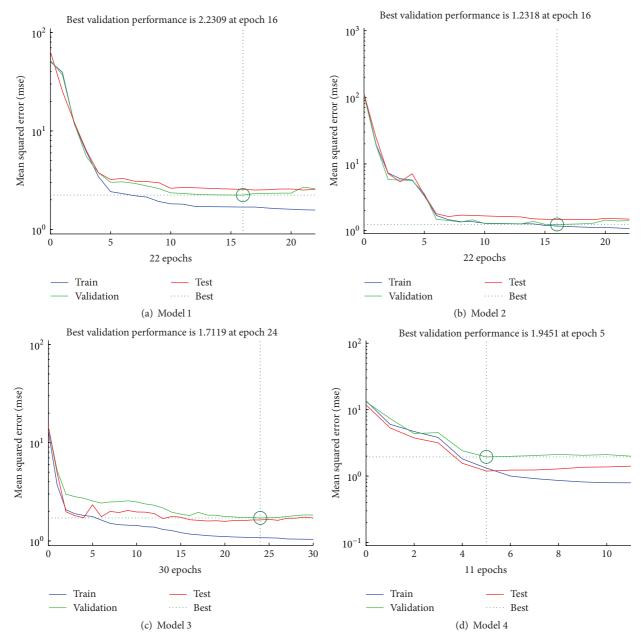


FIGURE 2: Learning performance of four BPANN models.

where $\Delta \omega_{ij}$ and $\Delta \omega_{jk}$ are the changes of weight values for hidden and output layers; N is the number of iterations; η is the learning rate, a parameter selected for the magnitude of change in interconnecting weights.

3.3.5. Calculating the Total Error. The total error of all training samples can be calculated as

$$E = \frac{1}{2} \sum_{t=1}^{T} \sum_{k=1}^{m} \left[do_k(t) - yo_k(t) \right]^2,$$
 (8)

where t is the serial number of training samples and T is the number of training samples.

3.3.6. Iteration Termination Conditions. If $E < \varepsilon$ or the number of iterations is larger than preset maximum learning

number M, then stop the ANN algorithm and output the results. Otherwise, return to the second step and begin the next learning iteration.

4. Results and Discussion

The BPANN codes were developed using a commercial software named MATLAB. The field bicycle data for training, validation, and testing are collected from eleven bicycle paths in Hangzhou, China [33]. 459 samples were collected, and 70% of total samples (321 samples) were used for training, 15% (69 samples) for validation, and 15% (69 samples) for testing. Detailed descriptive statistics of field data can be found in [16] and Table 1. The BPANN with 2-10-1, 4-10-1, 6-10-1, and 8-10-1 architectures for Models 1–4 are trained and validated.

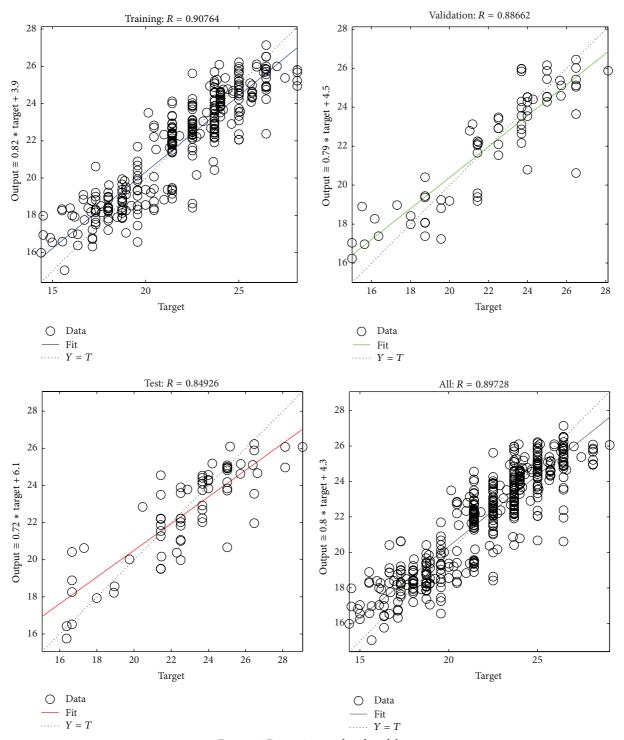


FIGURE 3: Regression results of model 1.

The trained models were tested by 69 samples which were not used in the training and validation stages.

Before training, in order to improve the training performance of the BPANN, it is often useful to scale the field input variables so that they always fall within a specified range. Therefore, in this study, field sample data is normalized in the range [0 1] by using the following formula:

$$nx_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}.$$
 (9)

The strength of each training, validation, and testing stage was evaluated by calculating the error and regression coefficient *R*. Learning performance plots of four BPANN models are shown in Figure 2, and the regression analysis

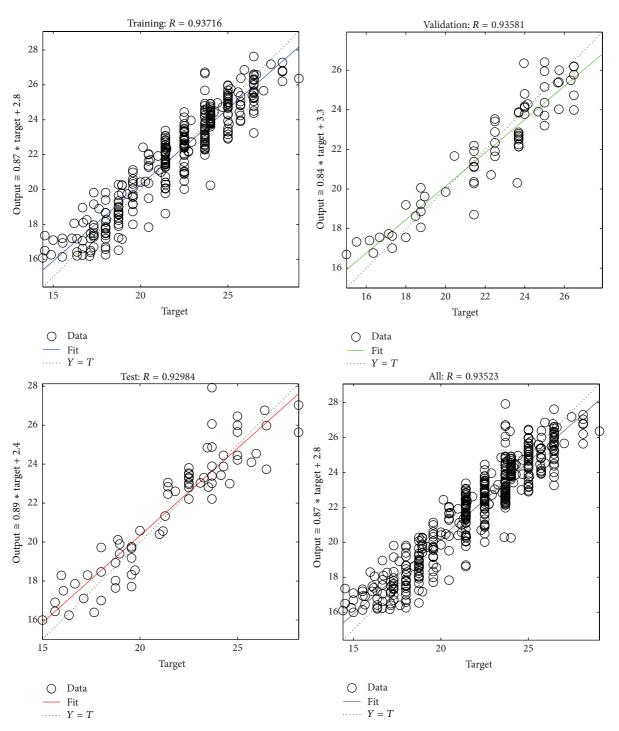


FIGURE 4: Regression results of model 2.

plots of four models for training, validation, and testing are presented in Figures 3–6.

The performance indicators, the mean absolute percentage error (MAPE) and the root mean square error (RMSE), for the testing samples were proposed [34]. These two indicators are given by the following equations:

$$MAPE = \frac{1}{M} \sum_{j=1}^{M} \left| \frac{\widehat{V}_{f}(j) - V_{f}(j)}{V_{f}(j)} \right| \times 100\%,$$

$$RMSE = \sqrt{\frac{1}{M} \sum_{j=1}^{M} \left[\widehat{V}_{f}(j) - V_{f}(j) \right]^{2}},$$
(10)

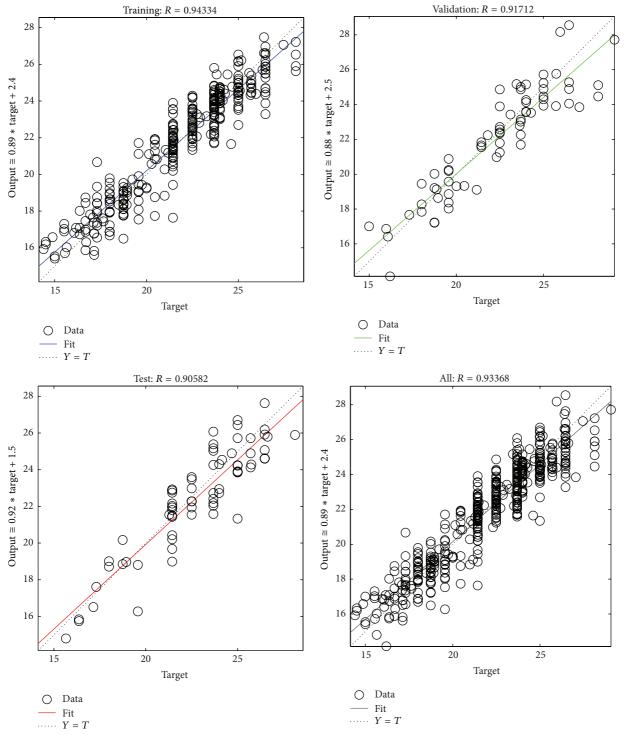


FIGURE 5: Regression results of model 3.

where $\widehat{V}_f(j)$ is the predicted free flow speed of bicycle for the jth testing sample; $V_f(j)$ is the observed free flow speed for the jth testing sample; M is the number of testing samples.

The correlation coefficient (R^2) , MAPEs, and RMSEs of four models are listed in Table 3, and the observed and predicted free flow speeds are illustrated in Figure 7. From the figure and the table, we have the following findings:

TABLE 3: Prediction errors for different models.

Model 1 0.72	5.25%	1.60
		1.00
Model 2 0.85	4.54%	1.21
Model 3 0.82	4.60%	1.28
Model 4 0.87	4.13%	1.09

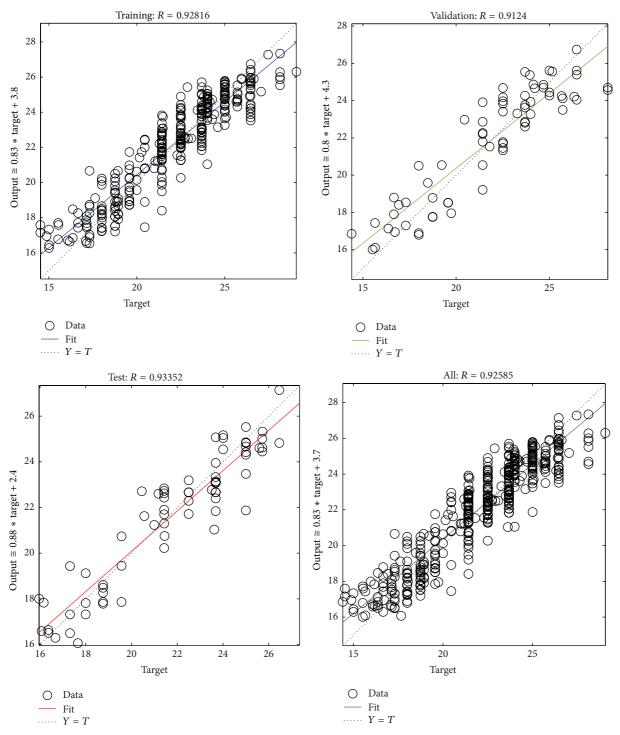


FIGURE 6: Regression results of model 4.

- (1) It is seen that all four BPANN models predict free flow speed with less errors, and the absolute speed differences are less than 2 kph. The results indicate that these models are all excellent in predicting the free flow speed. Model 1 including minimum input variables also performs well in predicting the free flow speed.
- (2) It can be also found that Model 2 and Model 3 have higher accuracies than Model 1. It is evident that the inclusions of bicycle types and characteristics of cyclists greatly improves the performance of Model 2 and Model 3 compared to Model 1. Different from motorized vehicles, characteristics of cyclists can be observed and analysed. Model 3 shows that using the

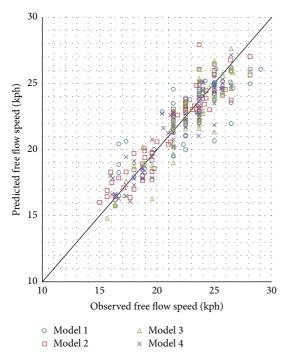


FIGURE 7: Observed and predicted free flow speed for different models.

input variables of cyclists' characteristics produces a slightly higher rate of accuracy compared to Model 1.

- (3) Comparing Model 2 and Model 3, it can be found that the performance of Model 2 is better than that of Model 3. This implies that bicycle type has more significant effect on bicycles' free flow speed and crash risk than characteristics of cyclists. Due to the higher speed of electric bicycles, the free flow speed and bicycle crash risk have significant correlation on the percentage of electric bicycles. Therefore, the management and speed limit for electric bicycles are very important to improve the safety of bicycle path.
- (4) Considering both input categories of bicycle types and characteristics of cyclists, the performance of Model 4 for testing dataset is the best. The MAPE and RMSE of testing data are 4.13% and 1.09 kph, respectively. This model provides us with the theoretical foundation for analysing the impact factors on the free flow speed and crash risk of bicycle traffic flow.

5. Conclusions

Free flow speed of bicycle traffic flow is a very important parameter for determining the speed limit of cycleway and evaluating the crash risk of bicycle traffic flow. The developed BPANN models in this paper are expected to be a useful and robust method to help traffic engineers improve the safety of bicycle traffic flow. Therefore, four different models with or without considering the impact factors (e.g., bicycle types and characteristics of cyclists) are used to predict the free flow speed and crash risk of heterogeneous bicycle traffic flow.

The BPANN models have been trained, validated, and tested using MATLAB software. As mentioned in results of testing datasets, the correlation coefficients (R^2) of four models by using adaptive learning have been obtained as 0.72, 0.85, 0.82, and 0.87, respectively, for expected outputs. The results imply that the proposed ANN methods have acceptable accuracies in predicting free flow speed of bicycles, and the considered bicycle types and characteristics of cyclists will effectively improve the accuracy of the prediction models. The study is limited to predicting the free flow speed only considering four categories factors. Other parameters such as percentage of passing, geometric features, and environmental features may be included for modelling in future work.

Conflict of Interests

The authors declare that there is no conflict of commercial or associative interests regarding the publication of this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 51278220 and 51208462), the Fundamental Research Funds for the Central Universities (2014QNA4018), the Projects in the National Science & Technology Pillar Program (2014BAG03B05), and the Key Science and Technology Innovation Team of Zhejiang Province (2013TD09).

References

- [1] Y. Kuang, X. Qu, and S. Wang, "A tree-structured crash surrogate measure for freeways," *Accident Analysis & Prevention*, vol. 77, pp. 137–148, 2015.
- [2] X. Qu, Y. Kuang, E. Oh, and S. Jin, "Safety evaluation for expressways: a comparative study for macroscopic and microscopic indicators," *Traffic Injury Prevention*, vol. 15, no. 1, pp. 89–93, 2014.
- [3] S. Jin, X. Qu, and D. Wang, "Assessment of expressway traffic safety using gaussian mixture model based on time to collision," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1122–1130, 2011.
- [4] X. Qu, Y. Yang, Z. Liu, S. Jin, and J. Weng, "Potential crash risks of expressway on-ramps and off-ramps: a case study in Beijing, China," *Safety Science*, vol. 70, pp. 58–62, 2014.
- [5] J. P. Schepers, E. Fishman, P. den Hertog, K. Klein Wolt, and A. L. Schwab, "The safety of electrically assisted bicycles compared to classic bicycles," *Accident Analysis and Prevention*, vol. 73, pp. 174–180, 2014.
- [6] Traffic Management Bureau Ministry of Public Security in China, Statistical Yearbook of Road Traffic Accidents in China, Traffic Management Bureau Ministry of Public Security in China, 2003–2012.
- [7] B.-J. Park, Y. Zhang, and D. Lord, "Bayesian mixture modeling approach to account for heterogeneity in speed data," *Transportation Research B: Methodological*, vol. 44, no. 5, pp. 662–673, 2010
- [8] L. Aarts and I. van Schagen, "Driving speed and the risk of road crashes: a review," *Accident Analysis and Prevention*, vol. 38, no. 2, pp. 215–224, 2006.

- [9] A. A. M. Aljanahi, A. H. Rhodes, and A. V. Metcalfe, "Speed, speed limits and road traffic accidents under free flow conditions," *Accident Analysis and Prevention*, vol. 31, no. 1-2, pp. 161– 168, 1998.
- [10] C. Hamann and C. Peek-Asa, "On-road bicycle facilities and bicycle crashes in Iowa, 2007–2010," Accident Analysis and Prevention, vol. 56, pp. 103–109, 2013.
- [11] X. Liu, L. D. Shen, and F. Ren, "Operational analysis of bicycle interchanges in Beijing, China," *Transportation Research Record*, vol. 1396, pp. 18–21, 1993.
- [12] H. Wei, J. Huang, and J. Wang, "Models for estimating traffic capacity on urban bicycle lanes," in *Proceedings of the 76th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 1997.
- [13] D. P. Allen, N. Rouphail, J. E. Hummer, and J. S. Milazzo II, "Event frequency and LOS to mean and standard deviation of bicycle speeds," *Transportation Research Record*, vol. 1636, pp. 29–36, 1998.
- [14] C. R. Cherry, Electric two-wheelers in china: analysis of environmental, safety, and mobility impacts [Ph.D. thesis], University of California, Berkeley, Calif, USA, 2007.
- [15] S. Lin, M. He, Y. Tan, and M. He, "Comparison study on operating speeds of electric bicycles and bicycles: experience from field investigation in Kunming, China," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2048, pp. 52–59, 2008.
- [16] S. Jin, X. Qu, D. Zhou, C. Xu, D. Ma, and D. Wang, "Estimating cycleway capacity and bicycle equivalent unit for electric bicycles," *Transportation Research A: Policy and Practice*, vol. 77, pp. 225–248, 2015.
- [17] P. P. Dey, S. Chandra, and S. Gangopadhaya, "Speed distribution curves under mixed traffic conditions," *Journal of Transportation Engineering*, vol. 132, no. 6, pp. 475–481, 2006.
- [18] D. Wang, D. Zhou, S. Jin, and D. Ma, "Characteristics of mixed bicycle traffic flow on conventional bicycle path," in *Proceedings* of the 94th Annual Meeting of the Transportation Research Board, Washington, DC, USA, 2015.
- [19] K. Fitzpatrick, P. Carlson, M. A. Brewer, M. D. Wooldridge, and S. P. Miaou, "Design speed, operating speed, and posted speed practices," NCHRP Report 504, Transportation Research Board, 2003.
- [20] D. R. Taylor, S. Muthiah, B. T. Kulakowski, K. M. Mahoney, and R. J. Porter, "Artificial neural network speed profile model for construction work zones on high-speed highways," *Journal of Transportation Engineering*, vol. 133, no. 3, pp. 198–204, 2007.
- [21] P. X. McKelvey, T. L. Maleck, N. Stamatiadis, and D. K. Hardy, "Highway accidents and the older driver," *Transportation Research Record*, vol. 1172, pp. 47–57, 1998.
- [22] Y. M. Najjar, R. W. Stokes, and E. R. Russell, "Setting speed limits on Kansas two-lane highways: neuronet approach," *Transportation Research Record*, vol. 1708, pp. 20–27, 2000.
- [23] Transportation Research Board, Highway Capacity Manual (HCM), National Research Council, Washington, DC, USA, 2010.
- [24] Y. Hou, C. Sun, and P. Edara, "Statistical test for 85th and 15th percentile speeds with asymptotic distribution of sample quantiles," *Transportation Research Record*, vol. 2279, pp. 47–53, 2012.
- [25] Transportation Research Board, Special Report 254 Managing Speed: Review of Current Practice for Setting and Enforcing Speed Limits, National Research Council, Washington, DC, USA, 1998.

- [26] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [27] D. Singh, M. Zaman, and L. White, "Modeling of 85th percentile speed for rural highways for enhanced traffic safety," Tech. Rep. FHWA-OK-11-07, Oklahoma Department of Transportation, 2011
- [28] A. D'Andrea, F. Carbone, S. Salviera, and O. Pellegrino, "The most influential variables in the determination of V85 speed," *Procedia—Social and Behavioral Sciences*, vol. 53, pp. 633–644, 2012.
- [29] S. K. S. Abbas, M. A. Adnan, and I. R. Endut, "Exploration of 85th percentile operating speed model on horizontal curve: a case study for two-lane rural highways," *Procedia Social and Behavioral Sciences*, vol. 16, pp. 352–363, 2011.
- [30] A. Lobo, C. Rodrigues, and A. Couto, "Estimating percentile speeds from maximum operating speed frontier," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2404, pp. 1–8, 2014.
- [31] D. Singh, M. Zaman, and L. White, "Neural network modeling of 85th percentile speed for two-lane rural highways," *Trans*portation Research Record, vol. 2301, pp. 17–27, 2014.
- [32] A. Agarwal and R. D. Singh, "Runoff modelling through back propagation artificial neural network with variable rainfallrunoff data," *Water Resources Management*, vol. 18, no. 3, pp. 285–300, 2004.
- [33] D. Zhou, C. Xu, D. Wang, and S. Jin, "Estimating capacity of bicycle path on urban roads in Hangzhou, China," in *Proceedings of the 94th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2015.
- [34] S. Jin, D.-H. Wang, C. Xu, and D.-F. Ma, "Short-term traffic safety forecasting using Gaussian mixture model and Kalman filter," *Journal of Zhejiang University: Science A*, vol. 14, no. 4, pp. 231–243, 2013.

Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 761818, 8 pages http://dx.doi.org/10.1155/2015/761818

Research Article

Risk and Resilience Analysis of Complex Network Systems Considering Cascading Failure and Recovery Strategy Based on Coupled Map Lattices

Fuchun Ren, Tingdi Zhao, and Hongli Wang

School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

Correspondence should be addressed to Tingdi Zhao; ztd@buaa.edu.cn

Received 3 July 2015; Accepted 13 September 2015

Academic Editor: Mark Leeson

Copyright © 2015 Fuchun Ren 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.

Risk and resilience are important and challenging issues in complex network systems since a single failure may trigger a whole collapse of the systems due to cascading effect. New theories, models, and methods are urgently demanded to deal with this challenge. In this paper, a coupled map lattices (CML) based approach is adopted to analyze the risk of cascading process in Watts-Strogatz (WS) small-world network and Barabási and Albert (BA) scale-free network, respectively. Then, to achieve an effective and robust system and provide guidance in countering the cascading failure, a modified CML model with recovery strategy factor is proposed. Numerical simulations are put forward based on small-world CML and scale-free CML. The simulation results reveal that appropriate recovery strategies would significantly improve the resilience of networks.

1. Introduction

In modern society, many real-world systems, such as Internet, transportation network, and power grid system [1–3], can be described by complex networks. They have made a great contribution to social development. However, despite the benefits to people's life, risk or disadvantage of complex network system is also serious since a local failure may trigger a breakdown of the whole system due to cascading effect; many real-world examples have been witnessed around the world, for example, the power grid of North America in 2003 [4], in which a fault of three extrahigh voltage transmissions led to a cascading effect in power system affecting about 50 million people and caused an economic loss of 4 billion to 10 billion, the 2003 SARS plague [5], and the 2008 global financial crisis [6].

The traumatic experiences bring an urgent need of the study of risk and safety of complex network system which has been a hot topic for scientific researchers in recent years [7–10]. Amongst these researches, cascading failure of complex networks has been one of the hottest topics. Numbers of important aspects of cascading failure in complex networks

have been discussed by scholars recently. Many of them provide a good research on the assessment and modeling of cascading effect [11, 12].

From the perspective of risk management, safety and resilience engineering, hazard, and risk should be identified and controlled timely and effectively. However, with the increasing of scale and complexity of modern complex network systems, the responsibility to manage an effective and safe system has been heavier than ever before. New theories, models, and methods are urgently demanded to protect the systems from an unacceptable risk condition.

Coupled map lattices (CML) are a dynamical system with discrete time, discrete space, and continuous state variables [13] which are initially used to study spatiotemporal chaos and then widely applied in various fields including but not limited to biology, mathematics, and engineering. In CML, the dynamical elements are situated at discrete spatial points while the time is discrete and the state is continuous. Each spatial element is coupled to its neighbors. Recently, CML has been widely investigated to model the dynamical behaviors and cascading failure in complex systems [14–18]. To counter the impact of cascading failure and provide an

effective prevention of large-scale breakdown, strategies such as redistribution and restoration have been discussed [19–22]; however there is few considering recovery strategy of failed node in the process of cascading failure. In other words, risk recognition and analysis are just a prime step; actions can be taken to achieve an acceptable resilience of the whole system [23].

The remainder of this paper is organized as follows. Section 2 proposes the risk and resilience analysis approach based on CML and a modified CML model with recovery strategy factor is discussed. In Sections 3 and 4, the smallworld CML and scale-free CML are adopted, respectively, to analyze the cascading failure and the effect of recovery strategy. Finally, Section 5 provides a conclusion of this paper.

2. Risk and Resilience Analysis Based on Coupled Map Lattices

2.1. Risk Analysis of Cascading Failure Based on Coupled Map Lattices. Coupled map lattices (CML) are a dynamical system with discrete time, discrete space, and continuous state variables; typically, CML of N nodes can be described as follows:

$$x_{i}(t+1) = \left| (1-\varepsilon) f\left(x_{i}(t)\right) + \frac{\varepsilon \sum_{j=1, j\neq i}^{N} a_{ij} f\left(x_{j}(t)\right)}{k(i)} \right|,$$
 (1)

where $x_i(t)$ is the state variable of the ith node at the tth time step. The connection information of the nodes is given by an adjacency matrix $A=(a_{ij})_{N\times N}$. If there is an edge between node i and node j, $a_{ij}=a_{ji}=1$; otherwise, $a_{ij}=a_{ji}=0$. k(i) denotes the degree of node i. $\varepsilon\in(0,1)$ represents the coupling strength. The function f defines the local dynamics which is chosen in this work as the chaotic logistic map f(x)=4x(1-x), where $f(x)\in(0,1)$ if $x\in(0,1)$. Absolute value notation in (1) is used to guarantee that the state of each node is always nonnegative. The ith node is regarded to be in a normal state at the tth time step if $0< x_i(t)<1$. In contrast, if $x_i(m)\geq 1$ at the tth time step, the tth node is regarded to be failed and t and t is t0.

However, to simulate and analyze the cascading failure based on the CML, an external perturbation $R \ge 1$ is added to a specific node b at the mth time step as in

$$x_{b}(m) = \left| (1 - \varepsilon) f \left(x_{b}(m - 1) \right) + \frac{\varepsilon \sum_{j=1, j \neq b}^{N} a_{bj} f \left(x_{j}(m - 1) \right)}{k(b)} \right| + R.$$
(2)

Then, $x_b(m) \ge 1$ at the mth time step; other nodes would be impacted according to (1) afterwards and a cascading failure would be triggered.

Topological structure (static) and restoration strategy (dynamic) of the complex network would be fundamental to cascading effect, spread threshold, and scope [11, 21, 23]. In

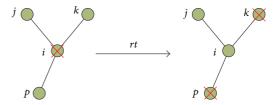


FIGURE 1: An example of the cascading process with recovery strategy.

CML, it is the coupling strength ε and external perturbation R that get an important impact [16]. However, coupling strength ε is something like topological factor and external perturbation R is something like external attack that is not controllable once happened. Even though restoration strategy in weighted networks is discussed, there are few discussions about the recovery strategy in CML; that is, the failed units would be treated as failed evermore in the cascading process. So a factor of recovery strategy against cascading failure should be added in the research of cascading failure in CML.

2.2. A Modified CML Model with Recovery Strategy Factor. As illustrated above, if the ith node got failed at the mth time step, that is, $x_i(m) \geq 1$, then the state variable of the ith node is identically equal to 0 after the mth time step; that is, $x_i(t) \equiv 0$, t > m. But in actual world, recovery strategies would be implemented at proper time after the malfunction, as illustrated in Figure 1; when the ith node got failed at the mth time step it would be recovered to normal after rt time step while the states of its neighbors, that is, kth and pth nodes, have been changed to failed at the (m + rt)th time step. The timeliness and effectiveness of recovery would make a critical influence on the cascading process. In cascading, we take the reduction of failure rate and recovery number of failed nodes as the two metric parameters of the effectiveness of recovery.

In the following sections, three different recovery strategies with certain recovery time and one hybrid strategy with random recovery time are adopted to illustrate the effect of recovery strategy.

- (1) Strategy 1: the failed node would be recovered immediately at the next time step (rt = 1).
- (2) Strategy 2: the failed node would be recovered after two time steps (rt = 2).
- (3) Strategy 3: the failed node would be recovered after three time steps (rt = 3).
- (4) Strategy 4: the failed node would be recovered after a random time step ranging from 1 to 3 ($rt \in \{1, 2, 3\}$).

3. Analysis in Small-World Coupled Map Lattices

A small-world network is characterized by shorter average path length and higher clustering coefficient; that is, most

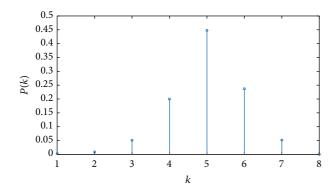


FIGURE 2: Node degree distribution of the small-world network model.

nodes only connect to their nearest-neighbor nodes while only a few nodes have long-range connections with relatively distant nodes [24]. It is proved that there are many real-world network systems subject to small-world feature, such as the transportation system [25, 26] and electrical power grid [27, 28]. In this section, analyses in small-world CML are put forward to investigate the effectiveness of recovery strategies in different conditions.

Firstly, the small-world network is generated according to Watts-Strogatz (WS) model [24] by changing the rewiring probability p based on a K-regular network with N nodes arranged on a ring and 2K edges per node. It is noteworthy that the WS network will be a random network when the rewiring probability p=1 and will be a regular network when the rewiring probability p=0. Here we take N=1000, K=2, and p=0.3. The node degree distribution of the generated small-world network is depicted in Figure 2.

3.1. Cascading Failure Affected by Random External Perturbation R

3.1.1. When No Recovery Strategy Is Considered. In the small-world CML, the coupling coefficient of CML is given: $\varepsilon=0.3$ and then a random node failure is triggered by 5 different external perturbations $R\in(1,5)$. As shown in Figure 3, the cascading failure in small-world CML would be triggered after the initial failure when $R\geq 1$, and the cascading process could be approximately divided into 3 phases: $t_1\in(1,3)$, the failure rate increases slowly; $t_2\in(4,6)$, the failure rate increases rapidly; and $t_3\in(6,\infty)$, the number of nodes would be stable. Noteworthily, the failure rate is just about 10% when R=2 and would not be over 40% for other R, implying that a few long-range connections are enough for a single node failure to trigger a cascading of failure on network in a few steps, but the scale is controllable in some way.

3.1.2. When a Recovery Strategy Is Added. However, in actual life, it is unrealistic and irresponsible to let the disaster spread unbridledly. In contrast, a small-world CML with a recovery strategy is simulated to illustrate the difference. Here, strategy

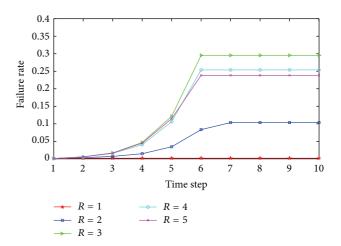


FIGURE 3: The relationship between failure rate and time step with different perturbations *R* in small-world CML with no recovery strategy.

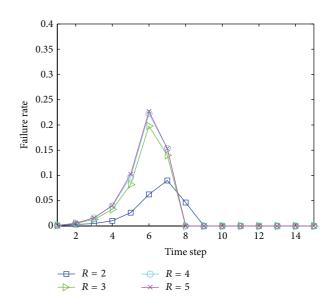


FIGURE 4: The relationship between failure rate and time step with different perturbations *R* in small-world CML with certain recovery strategy.

2 that is introduced in Section 2.2 with a certain recovery time is employed. Simulation result is shown in Figure 4.

In comparison with the result in Figure 3, the peak failure rates are reduced, and there is an obvious decreasing phase at the 6th or the 7th time step and then after about 2 time steps all of the nodes would return to normal. For each scenario, the recovery number of failed nodes is shown in Figure 5.

As calculated, the recovery number re_number = 122, 236, 267, and 273, respectively. It could be found that the *re_number* is almost the same except for a little rising trend when the external perturbation $R \ge 3$, but the recovery number would get a sharp increasing (nearly doubled) when the external perturbation is changed from R = 2 to R = 3.

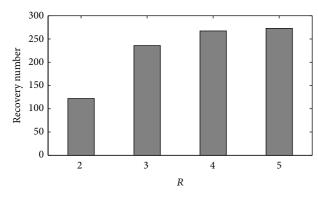


FIGURE 5: The relationship between recovery number of failed nodes *re_number* and perturbation *R*.

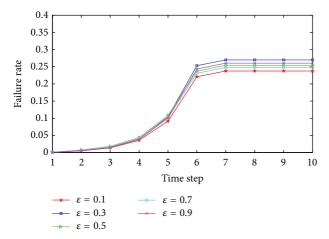


FIGURE 6: The relationship between failure rate and time step with different coupling coefficients ε in small-world CML with no recovery strategy.

3.2. Cascading Failure Affected by Coupling Coefficient ε

3.2.1. When No Recovery Strategy Is Considered. The coupling coefficient ε represents the coupling strength of CML. In this part, its effect on cascading failure is discussed. Firstly, a perturbation R = 5 is added to a randomly selected node to trigger the cascading process, and ε takes the values of 0.1, 0.3, 0.5, 0.7, and 0.9, respectively. As shown in Figure 6, cascading process can also be approximately divided into 3 phases: $t_1 \in (1,4)$, the failure rate increases slowly; $t_2 \in$ (4,6), the failure rate increases rapidly; and $t_3 \in (6,\infty)$, the number of nodes would be stable. However, it is shown that the cascading failure process is little affected by coupling coefficient ε as the final failure rates are all about 25% (less than 5% range) and the number of failed nodes will get into stability at the 7th time step. This may be explained by the fact that small-world model is highly coupled intrinsically with a higher clustering coefficient, and then the effect of coupling coefficient ε of CML would contribute little in the small-world CML model.

3.2.2. When a Recovery Strategy Is Added. In contrast, to illustrate the effect of recovery strategy in cascading failure

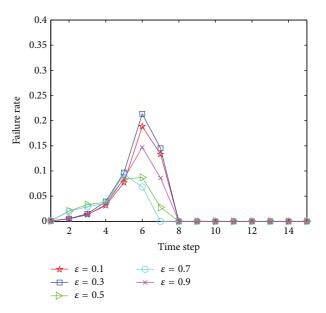


FIGURE 7: The relationship between failure rate and time step with different coupling coefficients ε in small-world CML with certain recovery strategy.

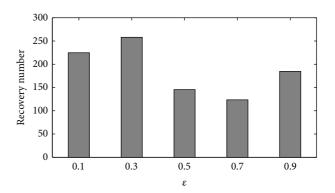


Figure 8: The relationship between recovery number of failed nodes re_number and coupling coefficient ε .

of different coupling coefficients ε , the small-world CML with recovery strategy is modeled; as introduced in Section 2, strategy 2 with a certain recovery time is applied. The simulation result is shown in Figure 7.

In comparison with the result in Figure 6, the peak failure rates are all reduced, and all of the nodes would return to normal before the 8th time step. However, the result shown in Figure 7 reflects a difference of effect degree of different coupling coefficients ε under the recovery strategy. The failure rate would be largely reduced (more than 50%) when $\varepsilon=0.5$ and 0.7. Further, as shown in Figure 8, the recovery number $re_number=225, 258, 146, 124,$ and 185, respectively.

It is clear that the *re_number* is also fewer when $\varepsilon = 0.5$ and 0.7. This would be very helpful to provide guidance for failure prevention in small-world CML.

3.3. Cascading Failure Affected by Recovery Strategy. As illustrated above, an added recovery strategy in small-world

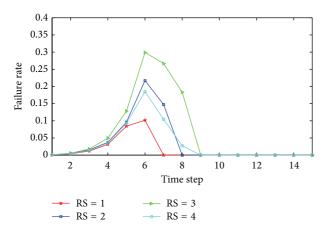


FIGURE 9: The relationship between failure rate and time step with recovery strategies in small-world CML.

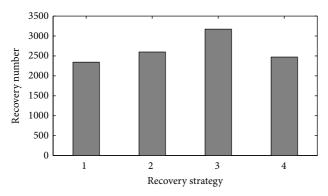


FIGURE 10: The relationship between recovery number and recovery strategy in small-world CML.

CML could significantly improve the robustness of network; a further study of the effect of different recovery strategy is put forward. As introduced in Section 2.2, four different recovery strategies are simulated, respectively (shown in Figure 9). For each simulation, the same perturbation R=5 is added to a randomly selected node to trigger the cascading process and coupling coefficients take the same value $\varepsilon=0.3$.

Together with the recovery number re_number = 2344, 2599, 3172, and 2472 (shown in Figure 10), it could be found that the small-world CML would achieve a lowest peak of failure rate with the lowest recovery number by adopting recovery strategy 1 (RS1); that is, the failed node would be recovered immediately at the next time step. This result reveals that in small-world CML a timely recovery would be the best prevention from cascading failure since the failure could spread very fast due to the small-world feature of network.

4. Analysis in Scale-Free Coupled Map Lattices

Many real-world network systems could be described by scale-free network model, such as the Internet, WWW [29]. The scale-free networks with degree distribution P(k) give a power-law behavior; that is, $P(k) \sim ck^{-\gamma}$, where P(k) is

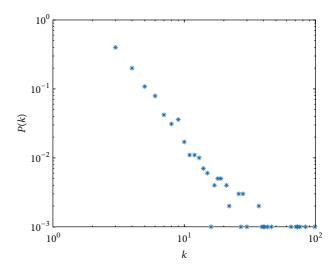


Figure 11: Node degree distribution of the scale-free network model.

the probability that the degree of a node in the network is equal to k and γ is scale-free networks exponent assigned to a positive real number. Barabási and Albert (BA) argued that the generation of networks in the scale-free structure is based on two rules, growth and preferential attachment [29], according to which the BA scale-free model is constructed with the parameters $n=1000,\ m_0=3,\ {\rm and}\ \Delta m=3;\ P(k)\sim ck^{-\gamma}$ with c=6.3 and $\gamma=2.5$. The degree distribution of the adopted scale-free network model is depicted in Figure 11 on log-log scales.

4.1. Cascading Failure Affected by Random External Perturbation R

4.1.1. When No Recovery Strategy Is Considered. In the scale-free CML, a random external perturbation R ranging from 1 to 5 (five different levels) is added to a random node to trigger the cascading process and the coupling coefficient of CML is given: $\varepsilon=0.3$. As shown in Figure 12, the cascading failure would be triggered when $R\geq 1$, and the cascading process can be approximately divided into 3 phases: $t_1\in (1,3\sim 4)$, the failure rate increases slowly; $t_2\in (3,6\sim 8)$, the failure rate increases rapidly; and $t_3\in (6\sim 8,\infty)$, the number of nodes would be stable. What is more, the failure rate would be 100% when $R\geq 4$ and more than 60% even $R\geq 2$, revealing that some nodes with high node degree are enough for a single node failure to cause a collapse of the whole network in a few steps.

4.1.2. When a Recovery Strategy Is Added. In contrast, the scale-free CML with a recovery strategy is adopted to illustrate the difference. Here, strategy 2 with a certain recovery time is employed. Simulation result is shown in Figure 13.

In comparison with the result in Figure 12, the peak failure rate would be all reduced since the recovery strategy works earlier at the second phase of cascading process when failure rate would increase rapidly.

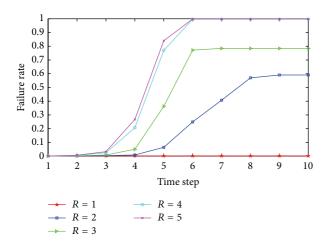


FIGURE 12: The relationship between failure rate and time step with different perturbations R in scale-free CML with no recovery strategy.

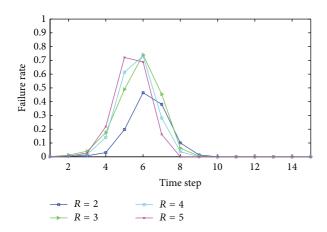


FIGURE 13: The relationship between failure rate and time step with different perturbations *R* in scale-free CML with certain recovery strategy.

As calculated in Figure 14, re_number = 601, 991, 915, and 913, respectively. It could be found that the re_number is nearly the same when the external perturbation $R \ge 3$ while it is very few when R = 2. It can also be explained by the earlier work of recovery strategy effect in countering the cascading failure.

4.2. Cascading Failure Affected by Coupling Coefficient ε

4.2.1. When No Recovery Strategy Is Considered. In this section, the effect on cascading failure of coupling coefficient ε in scale-free CML is studied. Here, ε takes the values of 0.1, 0.3, 0.5, 0.7, and 0.9, respectively, and the external perturbation R=5 to trigger the cascading failure (shown in Figure 15). The cascading process appears to be 3 different phases approximately: that is, $t_1 \in (1,3)$, the failure rate increases slowly; $t_2 \in (3,6)$, the failure rate increases rapidly; and $t_3 \in (6,\infty)$, the number of nodes would be stable. Approximately, the failure rate would be 100% when

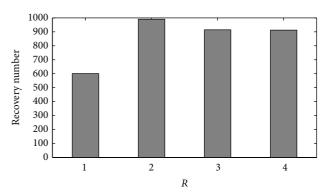


FIGURE 14: The relationship between recovery number of failed nodes *re_number* and perturbation *R*.

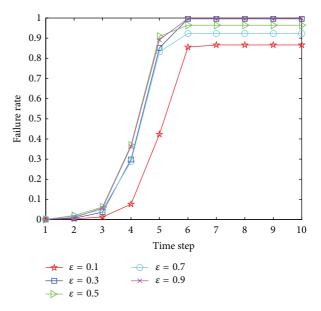


Figure 15: The relationship between failure rate and time step with different coupling coefficients ε in scale-free CML with no recovery strategy.

 ε = 0.3, 0.9 and 90% ~95% when ε = 0.5, 0.7 and only little higher than 80% when ε = 0.1, implying that the cascading failure process would be affected by coupling coefficient ε in scale-free CML in some way.

4.2.2. When a Recovery Strategy Is Added. In contrast, scale-free CML with recovery strategy 2 is modeled to illustrate the effect of recovery strategy with different coupling coefficients ε. The simulation result is shown in Figure 16.

In comparison with the result in Figure 15, the peak failure rates are reduced relatively, which is obvious (about 20%) when $\varepsilon=0.3$ and 0.9 and nearly did not change when $\varepsilon=0.1,0.5,0.7$. All the nodes would return to normal before the 8th time step. Further, the recovery number of failed nodes is shown in Figure 17.

As calculated, re_number = 993, 1015, 1008, 950, and 877, respectively. It could be found that the *re_number* is nearly the same with different coupling coefficients.

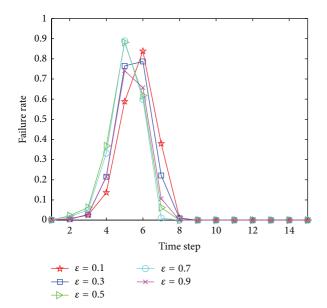


FIGURE 16: The relationship between failure rate and time step with different coupling coefficients ε in scale-free CML with certain recovery strategy.

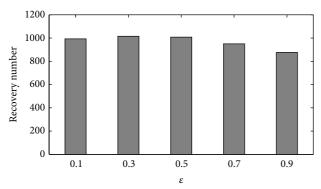


FIGURE 17: The relationship between recovery number of failed nodes re_number and coupling coefficient ε .

4.3. Cascading Failure Affected by Recovery Strategy. In this section, a further study of effects of different recovery strategy in scale-free CML is put forward. Four different recovery strategies with the same perturbation R=5 and coupling coefficients taking the same value $\varepsilon=0.3$ are simulated, respectively. As shown in Figure 18, recovery strategy 1 (RS1) would extremely lower the peak failure rate by about 40% compared with Figure 12.

Further, as calculated in Figure 19, the recovery number re_number = 1226, 1010, 900, and 990, respectively, revealing that the *re_number* would increase with the decreasing of recovery time. So, in this way, a tradeoff between failure rate and recovery number would be necessary in the selection of recovery strategy in scale-free CML.

5. Conclusions

This paper analyzes the cascading failure in Watts-Strogatz (WS) small-world network and Barabási and Albert (BA)

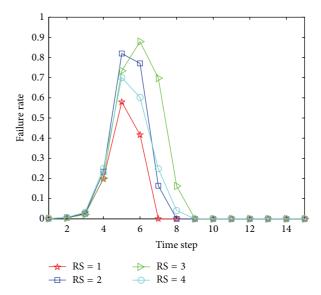


FIGURE 18: The relationship between failure rate and time step with recovery strategies in scale-free CML.

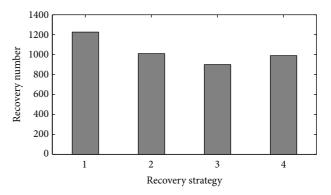


FIGURE 19: The relationship between recovery number and recovery strategy in scale-free CML.

scale-free network based on coupled map lattices (CML), respectively. A modified CML model with recovery strategy factor is proposed to manage the risk of cascading failure. Cascading failure triggered by random external perturbation R, coupling coefficient ε , and recovery strategy RS is simulated and discussed, and the main findings and contributions of this paper are as follows:

- (1) For both small-world CML and scale-free CML, the cascading failure would be triggered after the initial failure when $R \geq 1$, and the cascading process could be approximately divided into 3 phases; however, the failure rate is just about 10% when R = 2 and would not be over 40% for other R in small-world CML while it would be 100% when $R \geq 4$ and 80% even $R \geq 2$ in scale-free CML.
- (2) The cascading failure would be little affected by coupling coefficient ε as the final failure rates are all about 25% (less than 5% range) in small-world CML but a little more in scale-free CML (about 20% range).

(3) For both small-world CML and scale-free CML, when a recovery strategy is added, cascading failure would be significantly affected with an obvious reduction of failure rate in all conditions. But it reveals a difference that a timely recovery in small-world CML would be the best strategy by achieving a lowest peak of failure rate with the lowest recovery number while a tradeoff between failure rate and recovery number should be considered in the selection of recovery strategy in scale-free CML since recovery number would increase with the decreasing of recovery time.

The findings of this paper provide important theory basis for guidance in countering the cascading failure, managing the hidden potential risk of networks, and helping achieve an effective and robust system.

Conflict of Interests

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

Acknowledgments

This paper was supported by the National Basic Research Program of China (973 Program) (2014CB744904). In addition, the authors would like to thank the anonymous reviewers helping to improve the paper.

References

- [1] 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.
- [2] R. Guimera, S. Mossa, A. Turtschi, and L. N. Amaral, "The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 22, pp. 7794–7799, 2005.
- [3] G. Caldarelli, Scale-Free Networks: Complex Webs in Nature and Technology, OUP Catalogue, 2007.
- [4] United States-Canada Power System Outage Task Force, Final Report on the August 14th, 2003 Blackout in the United States and Canada: Causes and Recommendations, United States Department of Energy and Canadian Department of Natural Resources, 2004.
- [5] P. Wallis and B. Nerlich, "Disease metaphors in new epidemics: the UK media framing of the 2003 SARS epidemic," *Social Science and Medicine*, vol. 60, no. 11, pp. 2629–2639, 2005.
- [6] E. Helleiner, "Understanding the 2007-2008 global financial crisis: Lessons for scholars of international political economy," *Annual Review of Political Science*, vol. 14, pp. 67–87, 2011.
- [7] W. H. Chen, Q. Y. Jiang, Y. J. Cao, and Z. Han, "Risk-based vulnerability assessment in complex power systems," *Power System Technology*, vol. 29, no. 4, p. 3, 2005.
- [8] E. Zio, "From complexity science to reliability efficiency: a new way of looking at complex network systems and critical infrastructures," *International Journal of Critical Infrastructures*, vol. 3, no. 3-4, pp. 488–508, 2007.
- [9] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman, "Complex systems analysis of series of blackouts: cascading

- failure, critical points, and self-organization," *Chaos*, vol. 17, no. 2, Article ID 026103, 2007.
- [10] J. Lorenza, S. Battiston, and F. Schweitzer, "Systemic risk in a unifying framework for cascading processes on networks," *The European Physical Journal B*, vol. 71, no. 4, pp. 441–460, 2009.
- [11] L. Dueñas-Osorio and S. M. Vemuru, "Cascading failures in complex infrastructure systems," *Structural Safety*, vol. 31, no. 2, pp. 157–167, 2009.
- [12] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.
- [13] K. Kaneko, "Overview of coupled map lattices," *Chaos. An Interdisciplinary Journal of Nonlinear Science*, vol. 2, no. 3, pp. 279–282, 1992.
- [14] X. F. Wang and J. Xu, "Cascading failures in coupled map lattices," *Physical Review E*, vol. 70, no. 5, Article ID 056113, 2004
- [15] J. Xu and X. F. Wang, "Cascading failures in scale-free coupled map lattices," *Physica A: Statistical Mechanics and its Applications*, vol. 349, no. 3-4, pp. 685–692, 2005.
- [16] D. Cui, Z. Gao, and X. Zhao, "Cascades in small-world modular networks with CML'S method," *Modern Physics Letters B*, vol. 21, no. 30, pp. 2055–2062, 2007.
- [17] D. Cui, Z.-Y. Gao, and X.-M. Zhao, "Cascades with coupled map lattices in preferential attachment community networks," *Chinese Physics B*, vol. 17, no. 5, pp. 1703–1708, 2008.
- [18] X. G. Chen, J. Zhou, and Z. T. Zhu, "Cascading failures study of urban traffic system based on CML," *Mathematics in Practices and Theory*, vol. 39, no. 7, pp. 79–84, 2009.
- [19] L. Zhao, K. Park, and Y.-C. Lai, "Attack vulnerability of scalefree networks due to cascading breakdown," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 70, no. 3, Article ID 035101, 2004.
- [20] J. Ash and D. Newth, "Optimizing complex networks for resilience against cascading failure," *Physica A*, vol. 380, no. 1-2, pp. 673–683, 2007.
- [21] M. Babaei, H. Ghassemieh, and M. Jalili, "Cascading failure tolerance of modular small-world networks," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 58, no. 8, pp. 527–531, 2011.
- [22] C. Liu, D. Li, E. Zio, and R. Kang, "A modeling framework for system restoration from cascading failures," *PLoS ONE*, vol. 9, no. 12, Article ID e112363, 2014.
- [23] M. Ouyang and Z. Wang, "Resilience assessment of interdependent infrastructure systems: with a focus on joint restoration modeling and analysis," *Reliability Engineering & System Safety*, vol. 141, pp. 74–82, 2015.
- [24] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'smallworld' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [25] V. Latora and M. Marchiori, "Is the Boston subway a small-world network?" *Physica A: Statistical Mechanics and Its Applications*, vol. 314, no. 1–4, pp. 109–113, 2002.
- [26] K. A. Seaton and L. M. Hackett, "Stations, trains and small-world networks," *Physica A*, vol. 339, no. 3-4, pp. 635–644, 2004.
- [27] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the US power grid," *Safety Science*, vol. 47, no. 10, pp. 1332– 1336, 2009.
- [28] J.-W. Wang and L.-L. Rong, "Robustness of the western United States power grid under edge attack strategies due to cascading failures," *Safety Science*, vol. 49, no. 6, pp. 807–812, 2011.
- [29] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.