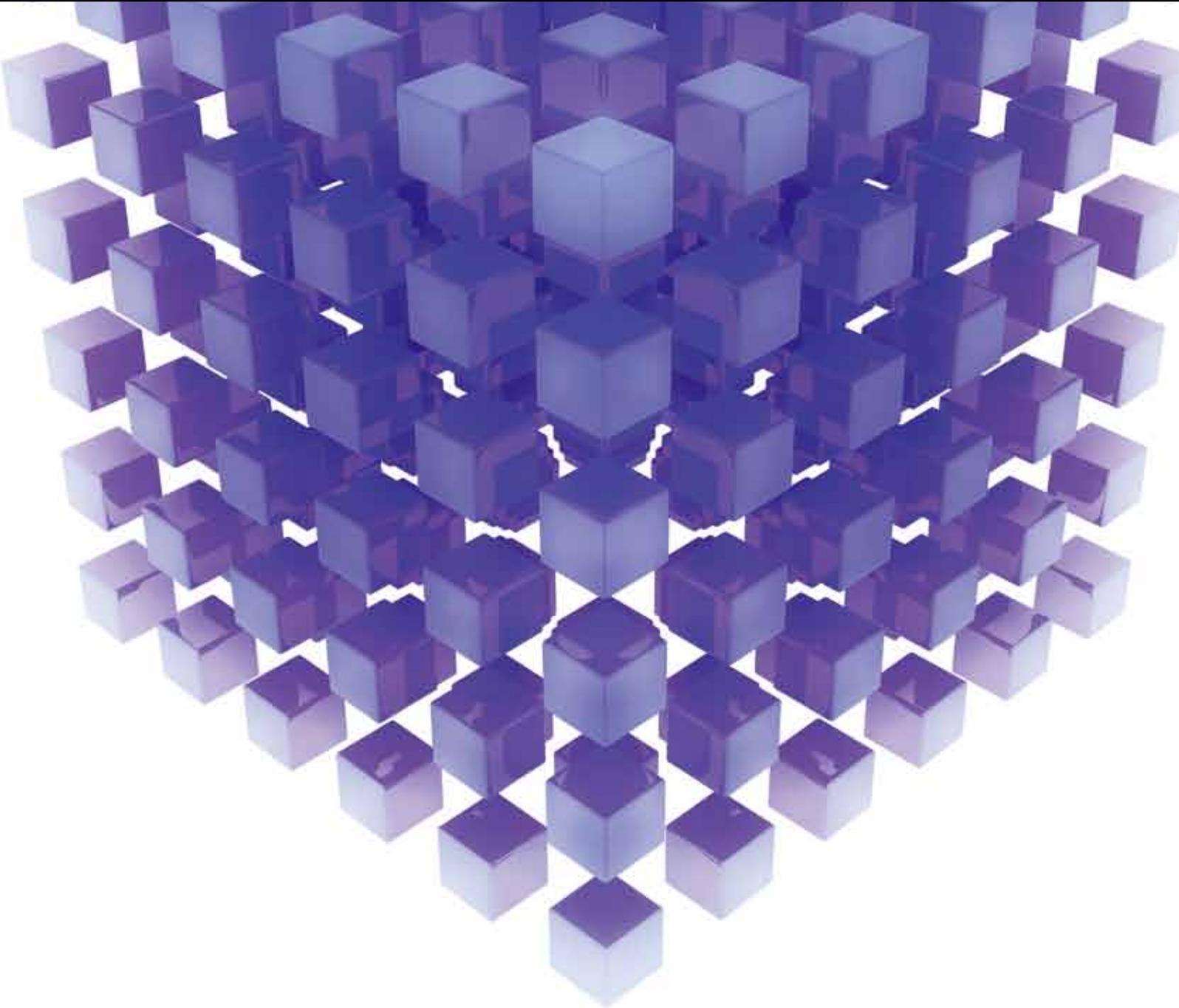


MATHEMATICAL PROBLEMS IN ENGINEERING

MODELING AND SIMULATION IN TRANSPORTATION ENGINEERING

GUEST EDITORS: WUHONG WANG, KLAUS BENGLER, GEERT WETS, AND HUIMIN NIU





Modeling and Simulation in Transportation Engineering

Mathematical Problems in Engineering

Modeling and Simulation in Transportation Engineering

Guest Editors: Wuhong Wang, Klaus Bengler, Geert Wets,
and Huimin Niu



Copyright © 2013 Hindawi Publishing Corporation. All rights reserved.

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

Editorial Board

- M. Abd El Aziz, Egypt
E. M. Abdel-Rahman, Canada
R. K. Abu Al-Rub, USA
Sarp Adali, South Africa
Salvatore Alfonzetti, Italy
Igor Andrianov, Germany
Sebastian Anita, Romania
W. Assawinchaichote, Thailand
Erwei Bai, USA
Ezzat G. Bakhoum, USA
José M. Balthazar, Brazil
R. K. Bera, India
C. Bérenguer, France
Jonathan N. Blakely, USA
Stefano Boccaletti, Spain
Stephane P.A. Bordas, USA
Daniela Boso, Italy
M. Boutayeb, France
Michael J. Brennan, UK
Salvatore Caddemi, Italy
Piermarco Cannarsa, Italy
Jose E. Capilla, Spain
Carlo Cattani, Italy
Marcelo M. Cavalcanti, Brazil
Diego J. Celentano, Chile
Mohammed Chadli, France
Arindam Chakraborty, USA
Yong-Kui Chang, China
Michael J. Chappell, UK
Kui Fu Chen, China
Xinkai Chen, Japan
Kue-Hong Chen, Taiwan
Jyh-Horng Chou, Taiwan
Slim Choura, Tunisia
Cesar Cruz-Hernandez, Mexico
Swagatam Das, India
Filippo de Monte, Italy
Antonio Desimone, Italy
Yannis Dimakopoulos, Greece
Baocang Ding, China
Joao B. R. Do Val, Brazil
Daoyi Dong, Australia
B. Dubey, India
Horst Ecker, Austria
M. Onder Efe, Turkey
Elmetwally Elabbasy, Egypt
A. Elías-Zúñiga, Mexico
Anders Eriksson, Sweden
Vedat S. Erturk, Turkey
Moez Feki, Tunisia
Ricardo Femat, Mexico
Robertt A. Valente, Portugal
C. Fuerte-Esquivel, Mexico
Zoran Gajic, USA
Ugo Galvanetto, Italy
Furong Gao, Hong Kong
Xin-Lin Gao, USA
Behrouz Gatmiri, Iran
Oleg V. Gendelman, Israel
Didier Georges, France
Paulo B. Gonçalves, Brazil
Oded Gottlieb, Israel
Fabrizio Greco, Italy
Quang Phuc Ha, Australia
M. R. Hajj, USA
Tony S. W. Hann, Taiwan
Thomas Hanne, Switzerland
K. R. (Stevanovic) Hedrih, Serbia
M.I. Herreros, Spain
Wei-Chiang Hong, Taiwan
Jaromir Horacek, Czech Republic
Huabing Huang, China
Chuangxia Huang, China
Gordon Huang, Canada
Yi Feng Hung, Taiwan
Hai-Feng Huo, China
Asier Ibeas, Spain
Anuar Ishak, Malaysia
Reza Jazar, Australia
Zhijian Ji, China
Jun Jiang, China
J. J. Judice, Portugal
Tadeusz Kaczorek, Poland
Tamas Kalmar-Nagy, USA
Tomasz Kapitaniak, Poland
Hamid Reza Karimi, Norway
Metin O. Kaya, Turkey
Nikolaos Kazantzis, USA
Farzad Khani, Iran
K. Krabbenhoft, Australia
Ren-Jieh Kuo, Taiwan
Jurgen Kurths, Germany
Claude Lamarque, France
Usik Lee, Korea
Marek Lefik, Poland
Stefano Lenci, Italy
Roman Lewandowski, Poland
Shanling Li, Canada
Ming Li, China
Jian Li, China
Shihua Li, China
Teh-Lu Liao, Taiwan
Panos Liatsis, UK
Shueei M. Lin, Taiwan
Yi-Kuei Lin, Taiwan
Jui-Sheng Lin, Taiwan
Yuji Liu, China
Wanquan Liu, Australia
Bin Liu, Australia
Paolo Lonetti, Italy
V. C. Loukopoulos, Greece
Junguo Lu, China
Chien-Yu Lu, Taiwan
Alexei Mailybaev, Brazil
Manoranjan K. Maiti, India
O. D. Makinde, South Africa
R. Martinez-Guerra, Mexico
Driss Mehdi, France
Roderick Melnik, Canada
Xinzhu Meng, China
Yuri V. Mikhlin, Ukraine
G. Milovanovic, Serbia
Ebrahim Momoniat, South Africa
Trung Nguyen Thoi, Vietnam
Hung Nguyen-Xuan, Vietnam
Ben T. Nohara, Japan
Sotiris K. Ntouyas, Greece
Gerard Olivar, Colombia
Claudio Padra, Argentina
Bijaya Ketan Panigrahi, India
Francesco Pellicano, Italy
Matjaz Perc, Slovenia
Vu Ngoc Phat, Vietnam
M. do Rosário Pinho, Portugal
A. Pogromsky, The Netherlands

Seppo Pohjolainen, Finland
Stanislav Potapenko, Canada
Sergio Preidikman, USA
Carsten Proppe, Germany
Hector Puebla, Mexico
Justo Puerto, Spain
Dane Quinn, USA
K. R. Rajagopal, USA
Gianluca Ranzi, Australia
Sivaguru Ravindran, USA
G. Rega, Italy
Pedro Ribeiro, Portugal
J. Rodellar, Spain
R. Rodriguez-Lopez, Spain
A. J. Rodriguez-Luis, Spain
Ignacio Romero, Spain
Hamid Ronagh, Australia
Carla Roque, Portugal
Rubén R. García, Spain
Manouchehr Salehi, Iran
Miguel A. Sanjuán, Spain
Ilmar F. Santos, Denmark
Nickolas S. Sapidis, Greece
E. J. Sapountzakis, Greece
Bozidar Sarler, Slovenia
Andrey V. Savkin, Australia
Massimo Scalia, Italy
Mohamed A. Seddeek, Egypt
A. P. Seyranian, Russia
Leonid Shaikhet, Ukraine

Cheng Shao, China
Bo Shen, Germany
Jian-Jun Shu, Singapore
Zhan Shu, UK
Dan Simon, USA
Luciano Simoni, Italy
Grigori M. Sisoiev, UK
Christos H. Skiadas, Greece
Davide Spinello, Canada
Sri Sridharan, USA
Rolf Stenberg, Finland
Changyin Sun, China
Jitao Sun, China
Xi-Ming Sun, China
Andrzej Swierniak, Poland
Yang Tang, Germany
Allen Tannenbaum, USA
Cristian Toma, Romania
Irina N. Trendafilova, UK
Alberto Trevisani, Italy
Jung-Fa Tsai, Taiwan
K. Vajravelu, USA
Victoria Vampa, Argentina
Josep Vehi, Spain
Stefano Vidoli, Italy
Xiaojun Wang, China
Dan Wang, China
Youqing Wang, China
Yongqi Wang, Germany
Cheng C. Wang, Taiwan

Moran Wang, China
Yijing Wang, China
Gerhard-Wilhelm Weber, Turkey
J. A. S. Witteveen, The Netherlands
Kwok-Wo Wong, Hong Kong
Ligang Wu, China
Zhengguang Wu, China
Gongnan Xie, China
Wang Xing-yuan, China
Xi Frank Xu, USA
Xuping Xu, USA
Jun-Juh Yan, Taiwan
Xing-Gang Yan, UK
Suh-Yuh Yang, Taiwan
Mahmoud T. Yassen, Egypt
Mohammad I. Younis, USA
Bo Yu, China
Huang Yuan, Germany
S.P. Yung, Hong Kong
Ion Zaballa, Spain
Ashraf M. Zenkour, Saudi Arabia
Jianming Zhan, China
Xu Zhang, China
Yingwei Zhang, China
Lu Zhen, China
Liancun Zheng, China
Jian Guo Zhou, UK
Zexuan Zhu, China
Mustapha Zidi, France

Contents

Modeling and Simulation in Transportation Engineering, Wuhong Wang, Klaus Bengler, Geert Wets, and Huimin Niu

Volume 2013, Article ID 768031, 2 pages

Traffic Volume Data Outlier Recovery via Tensor Model, Huachun Tan, Jianshuai Feng, Guangdong Feng, Wuhong Wang, and Yu-Jin Zhang

Volume 2013, Article ID 164810, 8 pages

Simulation-Based Genetic Algorithm towards an Energy-Efficient Railway Traffic Control,

Daniel Tuytens, Hongying Fei, Mohand Mezmaz, and Jad Jalwan

Volume 2013, Article ID 805410, 12 pages

A Neural Network Model for Driver's Lane-Changing Trajectory Prediction in Urban Traffic Flow,

Chenxi Ding, Wuhong Wang, Xiao Wang, and Martin Baumann

Volume 2013, Article ID 967358, 8 pages

Dynamic Cognitive Self-Organized TDMA for Medium Access Control in Real-Time Vehicle to Vehicle Communications, Mario Manzano, Felipe Espinosa, Ángel M. Bravo-Santos, Enrique Santiso,

Ignacio Bravo, and David Garcia

Volume 2013, Article ID 574528, 13 pages

Multiobjective Model for Emergency Resources Allocation, Zhaosheng Yang, Huxing Zhou, Xueying Gao, and Songnan Liu

Volume 2013, Article ID 538695, 6 pages

Study on Vehicle Track Model in Road Curved Section Based on Vehicle Dynamic Characteristics,

Ren Yuan-Yuan, Zhao Hong-Wei, Li Xian-Sheng, and Zheng Xue-Lian

Volume 2012, Article ID 818136, 17 pages

Traffic Congestion Evaluation and Signal Control Optimization Based on Wireless Sensor Networks: Model and Algorithms, Wei Zhang, Guozhen Tan, Nan Ding, and Guangyuan Wang

Volume 2012, Article ID 573171, 17 pages

Electrical Drive Radiated Emissions Estimation in Terms of Input Control Using Extreme Learning Machines, A. Wefky, F. Espinosa, L. de Santiago, P. Revenga, J. L. Lázaro, and M. Martínez

Volume 2012, Article ID 790526, 11 pages

Model for Microcirculation Transportation Network Design, Qun Chen and Feng Shi

Volume 2012, Article ID 379867, 11 pages

Adaptability Analysis of Service Facilities in Transfer Subway Stations, Liya Yao, Lishan Sun, Wuhong Wang, and Hui Xiong

Volume 2012, Article ID 701852, 12 pages

Optimal Sizing and Control Strategy Design for Heavy Hybrid Electric Truck, Yuan Zou, Dong-ge Li, and Xiao-song Hu

Volume 2012, Article ID 404073, 15 pages

A Compromise Programming Model for Highway Maintenance Resources Allocation Problem,

Hui Xiong, Qixin Shi, Xianding Tao, and Wuhong Wang

Volume 2012, Article ID 178651, 11 pages

Equivalent Mechanical Model for Lateral Liquid Sloshing in Partially Filled Tank Vehicles,

Zheng Xue-lian, Li Xian-sheng, and Ren Yuan-yuan

Volume 2012, Article ID 162825, 22 pages

Train Stop Scheduling in a High-Speed Rail Network by Utilizing a Two-Stage Approach, Huiling Fu,

Lei Nie, Benjamin R. Sperry, and Zhenhuan He

Volume 2012, Article ID 579130, 11 pages

Analysis of the Effect of the Length of Stop-Spacing on the Transport Efficiency of a Typically Formed Conventional Locomotive Hauled Passenger Train in China, Xuesong Feng, Keqi Wu, Haidong Liu,

Qipeng Sun, and Hanxiao Zhang

Volume 2012, Article ID 518493, 9 pages

Metropolis Parking Problems and Management Planning Solutions for Traffic Operation Effectiveness,

Yuejun Liu, Wuhong Wang, Chengxi Ding, Hongwei Guo, Weiwei Guo, Liya Yao, Hui Xiong, and Huachun Tan

Volume 2012, Article ID 678952, 6 pages

A Cellular Automaton Traffic Flow Model with Advanced Decelerations, Yingdong Liu

Volume 2012, Article ID 717085, 14 pages

Research on Intersection Signal Switching Model under Emergency Situation, Li Zongping, Bai Wei, and Gan Mi

Volume 2012, Article ID 186383, 14 pages

Study on the Road Network Connectivity Reliability of Valley City Based on Complex Network,

Yong-Sheng Qian, Min Wang, Hong-Xia Kang, Jun-Wei Zeng, and Yu-Fei Liu

Volume 2012, Article ID 430785, 14 pages

Modeling and Algorithms of the Crew Rostering Problem with Given Cycle on High-Speed Railway Lines, Zhiqiang Tian and Huimin Niu

Volume 2012, Article ID 214607, 15 pages

A Method for Queue Length Estimation in an Urban Street Network Based on Roll Time Occupancy Data, Dongfang Ma, Dianhai Wang, Yiming Bie, Feng Sun, and Sheng Jin

Volume 2012, Article ID 892575, 12 pages

A Node Model Capturing Turning Lane Capacity and Physical Queuing for the Dynamic Network Loading Problem, Mingxia Gao

Volume 2012, Article ID 542459, 14 pages

Modeling of Signal Plans for Transit Signal Priority at Isolated Intersections under Stochastic Condition, Lv Bin

Volume 2012, Article ID 650242, 10 pages

A Simulation-Based Dynamic Stochastic Route Choice Model for Evacuation, Xing Zhao, Gang Ren, Chao Fan, and Chen-Zi Ding

Volume 2012, Article ID 396248, 18 pages

Study on the Stochastic Chance-Constrained Fuzzy Programming Model and Algorithm for Wagon Flow Scheduling in Railway Bureau, Bin Liu

Volume 2012, Article ID 602153, 13 pages

Coordination Game Analysis through Penalty Scheme in Freight Intermodal Service, Jian Liu, Jianning Yu, and Yinzhen Li

Volume 2012, Article ID 247509, 14 pages

Research on the Behavior Characteristics of Pedestrian Crowd Weaving Flow in Transport Terminal, Liya Yao, Lishan Sun, Zhiyong Zhang, Shuwei Wang, and Jian Rong

Volume 2012, Article ID 264295, 9 pages

Editorial

Modeling and Simulation in Transportation Engineering

Wuhong Wang,¹ Klaus Bengler,² Geert Wets,³ and Huimin Niu⁴

¹ Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

² Department of Ergonomics, Technical University of Munich, 85747 Munich, Germany

³ Transportation Research Institute (IMOB), Hasselt University, 3590 Diepenbeek, Hasselt, Belgium

⁴ School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

Correspondence should be addressed to Wuhong Wang; wangwuhong@bit.edu.cn

Received 19 November 2013; Accepted 19 November 2013

Copyright © 2013 Wuhong 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.

Transportation problems including congestion, accidents and pollution emissions have become important topics, which have impact on social and economic developments. Recently, the mathematical modeling and simulation methods have received considerable attention regarding their potential as a powerful technique that can be widely applied to solve traffic and mobility operations.

This special issue presented an overview of the efforts for creating a safe, green, and intelligent transportation, especially the new achievement of intelligent transportation system (ITS) which has been reviewed in enough detail. This special issue aims at gathering the current state of the art of mathematical modeling and computational intelligence that is applied into transportation engineering. The papers submitted to this special issue have great contribution to the application of mathematical methods and models or should be the representative study in the field of transportation research. The main topics of the special issue include the following:

- (i) intelligent transportation system (ITS),
- (ii) logistics and transportation management,
- (iii) transit and rail systems operation,
- (iv) traffic flow and transportation model,
- (v) mathematical optimization in traffic engineering,
- (vi) traffic safety and emergence response,
- (vii) renewable energy sources in transportation,
- (viii) transport planning and road traffic,

- (ix) vehicle active safety and intelligent vehicle,
- (x) energy, transport policy, and economics,
- (xi) driving behavior and driver assistance system.

Since our daily life and work are closely related to traffic and mobility, transportation demand has increased dramatically. Following this increase, we additionally have to balance the increasing desire for mobility and efficiency with the societal concerns about traffic problems. Today's traffic demand is predominantly served by individual motorized vehicles, which are the primary means of transportation.

So far the significant efforts for improving transportation circumstance by using new technologies have been made. In recent years, ITS activities have focused on Green Intelligence Transportation System, which contributes to sustainability and environmental conservation compared to ITS technologies. More advanced transportation applications will materialize an accident-free traffic in the foreseeable future. Transportation engineering would no doubt move forward in the direction of safe, reliable, and economic techniques; further improvements will be made in the integration of driver assistance systems and transportation safety.

Aiming at summarizing the state-of-the-art of modeling and simulation in the context of modern transportation systems and the debate on the traffic problems, this special issue intends to introduce and discuss the developments and applications of modeling and simulation in traffic and mobility from the perspective of intelligent transportation systems. We hope that the papers in this special issue provide views of

the research of mathematical models and the development of advanced transportation engineering applications.

Wuhong Wang
Klaus Bengler
Geert Wets
Huimin Niu

Research Article

Traffic Volume Data Outlier Recovery via Tensor Model

Huachun Tan,¹ Jianshuai Feng,² Guangdong Feng,³ Wuhong Wang,¹ and Yu-Jin Zhang⁴

¹ Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

² Integrated Information System Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

³ Civil Aviation Engineering Consulting Company of China, Beijing 100621, China

⁴ Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Huachun Tan; htan8@wisc.edu

Received 22 September 2012; Accepted 24 February 2013

Academic Editor: Huimin Niu

Copyright © 2013 Huachun Tan 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.

Traffic volume data is already collected and used for a variety of purposes in intelligent transportation system (ITS). However, the collected data might be abnormal due to the problem of outlier data caused by malfunctions in data collection and record systems. To fully analyze and operate the collected data, it is necessary to develop a validate method for addressing the outlier data. Many existing algorithms have studied the problem of outlier recovery based on the time series methods. In this paper, a multiway tensor model is proposed for constructing the traffic volume data based on the intrinsic multilinear correlations, such as day to day and hour to hour. Then, a novel tensor recovery method, called ADMM-TR, is proposed for recovering outlier data of traffic volume data. The proposed method is evaluated on synthetic data and real world traffic volume data. Experimental results demonstrate the practicability, effectiveness, and advantage of the proposed method, especially for the real world traffic volume data.

1. Introduction

In order to alleviate the traffic congestion problem and facilitate the mobility in metropolises, large amounts of traffic information are collected as a part of intelligent transportation system (ITS) such as CVIS (Cooperative Vehicle Infrastructure System) in China. These collected traffic data have wide range of applications. The real time traffic information is provided to travelers to support their decision for making process on the optimal route choice [1]. As shown by the work of Kim et al. [2], the real time information can contribute to reduce the operation cost and maximize resource utilization. In addition to these applications, the collected data could be applied to maximize the utilization of the infrastructure for smooth flow of the traffic. One such application of real time traffic data is traffic information control [3]. On the other hand, several data mining techniques have been applied to mine time related association rules from traffic databases and their results have been used for traffic prediction such as the works of Williams et al. [4] and Xu et al. [5]. From the above discussion, it is concluded that the collected traffic data are essential for many potential applications in ITS.

In real world, the collected data are always corrupted due to noise values, especially outlier value, which may be caused by detector failures, communication problems, or any other hardware/software related problems. The presence of outlier data in the database would degrade significantly the quality as well as reliability of the data and might impede the effectiveness of ITS applications. Therefore, it is essential to fill the gaps caused by outlier data in order to fully explore the applicability of the data and realize the ITS applications.

While many different kinds of traffic data such as traffic volume, speed, and occupancy are collected, the focus of this research is on the traffic volume outlier data recovery. It is supposed that the detectors collecting traffic information are set up at road sections and the collected values represent the traffic volume for those road sections. The aim of this research is to recover the traffic volume outlier data for road sections.

Literature survey in the related field shows that several filtering recovery techniques have been applied to recover the outlier traffic data [6–8]. Filtering methods include techniques such as singular value decomposition, wavelet analysis, immune algorithm, and spectrum subtraction. Filtering methods formulate the traffic volume as time series

model and smooth the traffic waveform. These approaches recover the outlier data of day by day through spectrum analysis and feature information extracting. However, the traffic data through the same location is significantly similar from day to day and these approaches cannot utilize such characteristic. Pei and Ma [6] show that similarity is an important factor impacting on recovery performance. While the above methods consider only one mode similarity, the recovery performance is mainly dependent on the smooth threshold. Unfortunately, the smooth threshold is empirically determined.

In order to improve the recovery performance and consider the multidimension characteristic of traffic data, mining the multimode similarities will make a great contribution for recovering outlier value. Our approach is based on utilizing multimode correlations of traffic data; that is, traffic data have different correlation on different modes, such as week mode, day mode, and hour mode. More concretely, the feature of the proposed method is to recover the outlier value using the traffic volume information of the different modes. But, the problem is not so simple, because traffic volumes from many days might be corrupted by outlier data simultaneously. In order to consider the multiple outlier traffic volumes, we use tensor modeling the traffic volume.

In order to solve the traffic volume outlier data problem, we formulate the traffic volume recovery problem as a data recovery problem based on the assumption that the essential traffic volume is low- n -rank/low rank and the outliers are sparse. That is, the corrupted traffic volumes can formulated as

$$\mathcal{A} = \mathcal{L} + \mathcal{S}, \quad (1)$$

where \mathcal{A} is the observed traffic volume which is corrupted, \mathcal{L} is the recovered traffic volume, and \mathcal{S} represents the outliers. In the problem, the entrances of corrupted traffic data are unknown. One straight solution is optimizing the following problem under the assumption that the n -rank of \mathcal{L} is small and the corrupted outliers are sparse or bounded:

$$\begin{aligned} \min_{\mathcal{L}} \sum_i \mu_i \text{rank}_i(\mathcal{L}) \\ \text{s.t. } \|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F \leq \delta. \end{aligned} \quad (2)$$

The tensor recovery problem of (2) has been studied in recent years, which will be detailed in Section 3. In this paper, a new data recovery method based on tensor model called Alternating Direction Method of Multipliers for Tensor Recovery (ADMM-TR) is proposed to handle the outlier traffic volumes.

This paper makes three main contributions. (1) We use tensor to model the traffic volume and take advantage of the multiway characteristics of tensor, which could explore the multicorrelations of different modes in traffic data; and (2) we formulate the problem of the traffic volume outlier data recovery as a tensor recovery problem; (3) we proposed ADMM-TR algorithm by extending ADMM from matrix to tensor case to solve the formulated tensor recovery problem for traffic volume, and the convergence of ADMM-TR is

proved. It also should be noted that the proposed ADMM-TR method is different from [9], which reported that extended ADM for tensor recovery is proposed. In fact, they presented ADM for tensor completion, in which the data are missed and the entrances of missing data are known. While in this research, the objective is to recover the data which are corrupted including missing or noised and the entrances of corrupted values are unknown.

The paper is organized as follows. We present the review of tensor model in Section 2. Section 3 briefly reviews the related data recovery methods. In Section 4, tensor model for traffic volume is constructed, traffic data recovery problem is formulated, and an efficient algorithm is proposed to solve the formulation. Also a simple convergence guarantee for the proposed algorithm is given. In Section 5, we evaluated the proposed method on synthetic data and real world traffic volume data. Finally, we provide some concluding remarks in Section 6.

2. Notation and Review of Tensor Models

In this section, we adopt the nomenclature of Kolda and Bader's review on tensor decomposition [10] and partially adopt the notation in [11].

A tensor is the generalization of a matrix to higher dimensions. We denote scalars by lowercase letters (a, b, c, \dots), vectors as bold-case lowercase letters ($\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$) and matrices as uppercase letters (A, B, C, \dots). Tensors are written as calligraphic letters ($\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$).

An n -mode tensor is denoted as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. Its elements are denoted as $a_{i_1 \dots i_n}$, where $1 \leq i_k \leq I_k$, $1 \leq k \leq N$. The mode- n unfolding (also called matricization or flattening) of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is defined as unfold(\mathcal{A}, n) := $A_{(n)}$. The tensor element (i_1, i_2, \dots, i_N) is mapped to the matrix element (i_n, j) , where

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^N (i_k - 1) J_k, \quad \text{with } J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m. \quad (3)$$

Therefore, $A_{(n)} \in \mathbb{R}^{I_n \times J}$, where $J = \prod_{k=1, k \neq n}^N I_k$. Accordingly, its inverse operator fold can be defined as fold($A_{(n)}, n$) := \mathcal{A} .

The n -rank of an N -dimensional tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, denoted by r_n , is the rank of the mode- n unfolding matrix $A_{(n)}$:

$$r_n = \text{rank}_n(\mathcal{A}) = \text{rank}(A_{(n)}). \quad (4)$$

The inner product of two same-size tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is defined as the sum of the products of their entries, that is,

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} a_{i_1 \dots i_n} b_{i_1 \dots i_n}. \quad (5)$$

The corresponding Frobenius norm is $\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$. Besides, the l_0 norm of a tensor \mathcal{A} , denoted by $\|\mathcal{A}\|_0$, is the number of nonzero elements in \mathcal{A} and the l_1 norm is defined

as $\|\mathcal{A}\|_1 := \sum_{i_1, \dots, i_n} |a_{i_1, \dots, i_n}|$. It is clear that $\|\mathcal{A}\|_F = \|A_{(n)}\|_F$, $\|\mathcal{A}\|_0 = \|A_{(n)}\|_0$, and $\|\mathcal{A}\|_1 = \|A_{(n)}\|_1$ for any $1 \leq n \leq N$.

The n -mode (matrix) product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with a matrix $M \in \mathbb{R}^{J \times I_n}$ is denoted by $\mathcal{A} \times_n M$ and is size $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$. In terms of flattened matrix, the n -mode product can be expressed as

$$\mathcal{Y} = \mathcal{A} \times_n M \iff Y_{(n)} = MA_{(n)}. \quad (6)$$

3. Review of Data Recovery Methods

Recently, the problem of recovering the sparse and low-rank components with no prior knowledge about the sparsity pattern of the sparse matrix, or the rank of the low-rank matrix, has been well studied. Authors of [12] proposed the concept of “rank-sparse incoherence” and solved the problem by an interior point solver after being reformulated as a semidefinite problem. However, although interior point methods normally take very few iterations to converge, they have difficulty in handling large matrices. So this limitation prevented the usage of the technique in computer vision and the traffic volume recovery in this research.

To solve the problem for large scale matrices, Wright et al. [13] have adopted the iterative thresholding technique to solve the problem and obtained scalability properties. Lin et al. have proposed an accelerated proximal gradient (APG) algorithm [14] and applied techniques of augmented Lagrange multipliers (ALM) [15] to solve the problem. Yuan and Yang [16] have utilized the alternating direction method (ADM) which can be regarded as a practical version of the classical ALM method to solve the matrix recovery problem. The ADM method has been proved to have a pleasing convergence speed and results in [16] demonstrated its excellent performance.

Inspired by the idea of [16], this paper extends the sparse and low-rank recovery problem to tensor case, which is due to the fact that the multidimensional traffic data can be formulated into the form of tensor.

4. ADMM-TR for Traffic Volume Outlier Recovery

In this section, we show the solution of problem (2). The tensor model is firstly constructed for traffic volume in Section 4.1. Then we present the tensor recovery problem in Section 4.2. In Section 4.3, the classical ADMM approach is introduced. In Section 4.4, we convert the original problem into a constrained convex optimization problem which can be solved by the extended ADMM approach and present the details of the proposed algorithm. Also the convergence guarantees of the proposed algorithm are given in this section.

4.1. Tensor Model for Traffic Volume. The correlations of traffic volume data are critical for recovering the corrupted traffic volume data. Traditional methods mostly exploit part of correlations, such as historical or temporal neighboring correlations. The classic methods usually utilize the temporal

TABLE 1: The similarity coefficient of four modes.

Mode	Size	Similarity coefficient
Hour	6×12	0.9670
Day	7×288	0.8654
Week	7×288	0.9153
Link	4×288	0.8497

correlations of traffic data from day to day. For the single detector data, multiple correlations contain the relations of traffic data from day to day, hour to hour, and so forth. In addition, the spatial correlations exist in multiple detectors data.

In this paper, quantitative analysis of traffic data correlation is analyzed based on the traffic volume data downloaded from <http://pems.dot.ca.gov/>. The correlation coefficient applied to measuring the data correlation is given by [17]

$$s = \frac{\sum_{n \geq i > j \geq 1} R(i, j)}{n(n-1)/2}, \quad (7)$$

where n refers to the whole data points; $R(i, j)$ refers to the correlation coefficient matrix. Table 1 gives the results of correlation coefficient of four modes, which is hour, day, week, and month.

Conventional methods usually use day-to-day matrix pattern to model the traffic data. Although each mode of traffic data has a very high similarity, these methods do not utilize the multimode correlations, which are “Day \times Hour,” “Week \times Hour,” and “Link \times Hour,” simultaneously and thus may result in poor recovery performance.

To make full use of the multimode correlation and traffic spatial-temporal information, traffic data need to be constructed into multiway data set. Fortunately, tensor pattern based traffic data can be well used to model the multiway traffic data. This helps keep the original structure and employ enough traffic spatial-temporal information.

4.2. The Tensor Recovery Problem. The problem of (2) is NP hard, since it is not convex. Then, we use an approximation formulation, as shown in (8),

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S}} : & \|\mathcal{L}\|_* + \eta \|\mathcal{S}\|_1, \\ \text{s.t.} \quad & \|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F \leq \delta, \end{aligned} \quad (8)$$

where $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is the given matrix to be recovered; $\mathcal{L} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is the low-rank component of \mathcal{A} ; $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is the sparse component of \mathcal{A} . Compared with (2), (8) relaxes the constrain for recovering a low- n -rank tensor from a high-dimensional data tensor despite both small entry-wise noise and gross sparse errors.

Recently, Liu et al. [18] have proposed the definition of the nuclear norm of an n -mode tensor:

$$\|\mathcal{X}\|_* := \frac{1}{n} \sum_{i=1}^n \|X_{(i)}\|_*. \quad (9)$$

Based on this definition, the optimization in (8) can be written as

$$\min_{\mathcal{L}, \mathcal{S}} : \|\mathcal{L}\|_* + \eta \|\mathcal{S}\|_1 \equiv \frac{1}{n} \sum_{i=1}^n \lambda_i \|L_{(i)}\|_* + \frac{1}{n} \sum_{i=1}^n \eta_i \|S_{(i)}\|_1 \quad (10)$$

$$\text{s.t. } \|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F \leq \delta.$$

In order to recover $(\widehat{\mathcal{L}}, \widehat{\mathcal{S}})$, instead of directly solving (10), we solve the following dual problem:

$$\min_{\mathcal{L}, \mathcal{S}} : \frac{1}{2\gamma} \|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F^2 + \sum_{i=1}^n \lambda_i \|L_{(i)}\|_* + \sum_{i=1}^n \eta_i \|S_{(i)}\|_1. \quad (11)$$

The problem in (11) is still difficult to solve due to the interdependent nuclear norm and l_1 norm constraints. To simplify the problem, the formulation can be reformulated as follows:

$$\min_{\mathcal{L}, \mathcal{S}, M_i, N_i} : \frac{1}{2\gamma} \|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F^2 + \sum_{i=1}^n \lambda_i \|M_i\|_* + \sum_{i=1}^n \eta_i \|N_i\|_1 \quad (12)$$

$$\text{s.t. } P_i \mathcal{L} = M_i \quad P_i \mathcal{S} = N_i \quad \forall i,$$

where P_i is the matrix representation of mode- i unfolding (note that P_i is a permutation matrix; thus $P_i^T P_i = I$); $M_{(i)}$ and $N_{(i)}$ are additional auxiliary matrices of the same size as the mode- i unfolding of \mathcal{L} (or \mathcal{S}).

4.3. The Classical ADMM Approach. The classical alternating direction method of multipliers (ADMM) is for solving structured convex programs of the form:

$$\begin{aligned} \min_{x \in C_x, y \in C_y} : & f(x) + g(y) \\ \text{s.t. } & Ax + By = c, \end{aligned} \quad (13)$$

where f and g are convex functions defined on closed subsets C_x and C_y ; A , B , and c are matrices and vector of appropriate sizes. The segmented Lagrangian function of (13) is

$$\begin{aligned} L_A(x, y, w) = & f(x) + g(y) + \langle w, Ax + By - c \rangle \\ & + \frac{\beta}{2} \|Ax + By - c\|_2^2, \end{aligned} \quad (14)$$

where w is a Lagrangian multiplier vector and $\beta > 0$ is a penalty parameter.

The approach performs one sweep of alternating minimization with respect to x and y individually then updates the multiplier w ; at the iteration k the steps are given by [18, Equations (4.79)–(4.81)]:

$$\begin{aligned} x^{(k+1)} & \leftarrow \arg \min_{x \in C_x} L_A(x, y^{(k)}, w^{(k)}), \\ y^{(k+1)} & \leftarrow \arg \min_{y \in C_y} L_A(x^{(k+1)}, y, w^{(k)}), \end{aligned} \quad (15)$$

$$w^{(k+1)} \leftarrow w^{(k)} + \rho\beta (Ax^{(k+1)} + By^{(k+1)} - c),$$

where ρ is the step length. A convergence proof for the above ADMM algorithm was shown as follows.

Theorem 1 (See [19, Proposition 5.2]). *Assume that the optimal solution set X^* of (13) is nonempty. Furthermore, assume that C_x is bounded or else the matrix $A^* A$ is invertible. Then a sequence $\{x^{(k)}, y^{(k)}, w^{(k)}\}$ generated by (15) is bounded, and every limit point of $\{x^{(k)}\}$ is an optimal solution of the original of problem (13).*

4.4. ADMM Extension to Tensor Recovery. We observe (12) is well structured in the sense that the separable structure emerges in both the objective function and constraints. Thus, we propose an algorithm based on an extension of the classical ADMM approach for solving the tensor recovery problem by taking advantage of this favorable structure.

The augmented Lagrangian of (12) is

$$\begin{aligned} L_A(\mathcal{L}, \mathcal{S}, M_i, N_i) & = \frac{1}{2\gamma} \|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F^2 \\ & + \sum_{i=1}^n \left(\lambda_i \|M_i\|_* + \langle Y_i, P_i \mathcal{L} - M_i \rangle + \frac{\alpha_i}{2} \|P_i \mathcal{L} - M_i\|_F^2 \right) \\ & + \sum_{i=1}^n \left(\eta_i \|N_i\|_1 + \langle Z_i, P_i \mathcal{S} - N_i \rangle + \frac{\beta_i}{2} \|P_i \mathcal{S} - N_i\|_F^2 \right), \end{aligned} \quad (16)$$

where Y_i, Z_i are Lagrangian multipliers and $\alpha_i, \beta_i > 0$ are penalty parameters.

Then we can now directly apply ADMM with this augmented Lagrangian function.

Computing M_i . The optimal M_i can be solved with all other variables to be constant by the following subproblem:

$$\min_{M_i} : \lambda_i \|M_i\|_* + \langle Y_i, P_i \mathcal{L} - M_i \rangle + \frac{\alpha_i}{2} \|P_i \mathcal{L} - M_i\|_F^2. \quad (17)$$

As shown in [20], the optimal solution of (17) is given by

$$\widehat{M}_i = U_i D_{\lambda_i/\alpha_i}(\Lambda) V_i^T, \quad (18)$$

where $U_i \Lambda V_i^T$ is the singular value decomposition given by

$$U_i \Lambda V_i^T = P_i \mathcal{L} + \frac{Y_i}{\alpha_i}, \quad (19)$$

and the ‘‘shrinkage’’ operator $D_\tau(x)$ with $\tau > 0$ is defined as

$$D_\tau(x) = \begin{cases} x - \tau & \text{if } x > \tau \\ x + \tau & \text{if } x < -\tau \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

Computing N_i . The optimal N_i can be solved with all other variables to be the constants by the following subproblem:

$$\min_{N_i} : \eta_i \|N_i\|_1 + \langle Z_i, P_i \mathcal{S} - N_i \rangle + \frac{\beta_i}{2} \|P_i \mathcal{S} - N_i\|_F^2. \quad (21)$$

Input: n -mode tensor \mathcal{A}

Parameters: $\alpha, \beta, \gamma, \lambda, \eta, \rho$

(1) Initialization: $\mathcal{L}^{(0)} = \mathcal{A}, \mathcal{S}^{(0)} = 0, M_i = \mathcal{L}^{(0)}, N_i = 0, k = 1$

(2) Repeat until convergence

(3) for $i = 1$ to n

(4) $M_i^{(k+1)} = U_i D_{\lambda_i/\alpha_i}(\Lambda) V_i^T$
 where $U_i \Lambda V_i^T = P_i \mathcal{L}^{(k)} + Y_i^{(k)}/\alpha_i$

(5) $N_i^{(k+1)} = D_{\eta_i/\beta_i}(P_i \mathcal{S}^{(k)} + Z_i^{(k)}/\beta_i)$

(6) $Y_i^{(k+1)} = Y_i^{(k)} + \alpha_i(P_i \mathcal{L}^{(k)} - M_i^{(k+1)})$

(7) $Z_i^{(k+1)} = Z_i^{(k)} + \beta_i(P_i \mathcal{S}^{(k)} - N_i^{(k+1)})$

(8) end for

(9) $\mathcal{L}^{(k+1)} = \frac{\mathcal{A} - \mathcal{S}^{(k)} - \gamma \sum_{i=1}^n P_i^T (Y_i^{(k+1)} - \alpha_i M_i^{(k+1)})}{1 + \gamma \sum_{i=1}^n \alpha_i}$

(10) $\mathcal{S}^{(k+1)} = \frac{\mathcal{A} - \mathcal{L}^{(k+1)} - \gamma \sum_{i=1}^n P_i^T (Z_i^{(k+1)} - \beta_i N_i^{(k+1)})}{1 + \gamma \sum_{i=1}^n \beta_i}$

(11) $\alpha = \rho\alpha, \beta = \rho\beta$

(12) End

(13) $k = k + 1$

Output: n -mode tensor \mathcal{L}, \mathcal{S}

ALGORITHM 1: ADMM-TR: ADMM for tensor recovery.

By the well-known l_1 minimization [21], the optimal solution of (21) is

$$\widehat{N}_i = D_{\eta_i/\beta_i} \left(P_i \mathcal{S} + \frac{Z_i}{\beta_i} \right), \quad (22)$$

where D_τ is the ‘‘shrinkage’’ operation.

Computing \mathcal{L} . Now we fix all variables except \mathcal{L} and minimize L_A over \mathcal{L} . The resulting minimization problem is the minimization of a quadratic function:

$$\begin{aligned} \min_{\mathcal{L}}: L_A(\mathcal{L}) &= \frac{1}{2\gamma} \|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F^2 \\ &+ \sum_{i=1}^n \left(\langle Y_i, P_i \mathcal{L} - M_i \rangle + \frac{\alpha_i}{2} \|P_i \mathcal{L} - M_i\|_F^2 \right). \end{aligned} \quad (23)$$

The objective function is differentiable, so the minimizer \mathcal{L}_{\min} is characterized by $(\partial L_A(\mathcal{L}))/\partial \mathcal{L} = 0$. Thus, we obtain

$$\mathcal{L}_{\min} = \frac{\{\mathcal{A} - \mathcal{S} - \gamma \sum_{i=1}^n P_i^T (Y_i - \alpha_i M_i)\}}{(1 + \gamma \sum_{i=1}^n \alpha_i)}. \quad (24)$$

Computing \mathcal{S} . Now we fix all variables except \mathcal{S} and minimize L_A over \mathcal{S} . The resulting minimization problem is the minimization of a quadratic function:

$$\begin{aligned} \min_{\mathcal{S}}: L_A(\mathcal{S}) &= \frac{1}{2\gamma} \|\mathcal{A} - \mathcal{L} - \mathcal{S}\|_F^2 \\ &+ \sum_{i=1}^n \left(\langle Z_i, P_i \mathcal{S} - N_i \rangle + \frac{\beta_i}{2} \|P_i \mathcal{S} - N_i\|_F^2 \right). \end{aligned} \quad (25)$$

The objective function is also differentiable, so the minimizer \mathcal{S}_{\min} is characterized by $(\partial L_A(\mathcal{S}))/\partial \mathcal{S} = 0$. Thus, we have

$$\mathcal{S}_{\min} = \frac{\{\mathcal{A} - \mathcal{L} - \gamma \sum_{i=1}^n P_i^T (Z_i - \beta_i N_i)\}}{(1 + \gamma \sum_{i=1}^n \beta_i)}. \quad (26)$$

For comparing with RSTD [22], we also choose the difference of \mathcal{L} and \mathcal{S} in successive iterations against a certain tolerance as the stopping criterion. The pseudocode of the proposed ADMM-TR algorithm is summarized in Algorithm 1.

Theorem 2 (Convergence of ADMM-TR). *Assume that the optimal solution set X^* of (11) is nonempty. A sequence $\{\mathcal{L}^{(k)}, \mathcal{S}^{(k)}, M_i^{(k)}, N_i^{(k)}, Y_i^{(k)}, Z_i^{(k)}\}$ generated by our proposed ADMM-TR algorithm is bounded, and every limit point of $\{\mathcal{L}^{(k)}, \mathcal{S}^{(k)}\}$ is an optimal solution of the original problem (11).*

Proof. We check the assumptions of Theorem 1. C_x is not bounded, but $P_i^* P_i = I$ is a constant multiple of the identity operator. Thus Theorem 1 can also be applied to ADMM-TR and Theorem 2 can be derived. \square

5. Numerical Experiments

This section evaluates the empirical performance of the proposed algorithm on synthetic data and compares the results with RSTD (Rank Sparsity Tensor Decomposition) [22]. Also, experiments on traffic volume data outlier recovery illustrate the efficiency of the proposed method in traffic research filed.

We use the Lanczos algorithm for computing the singular values decomposition and adopt the same rule for predicting

TABLE 2: Comparison of ADMM-TR with RSTD for synthetic data where $\mathcal{A}_0 \in \mathbb{R}^{40 \times 40 \times 40}$, n -rank = [5, 5, 5].

spr	ADMM-TR				RSTD			
	RSE_ \mathcal{L}_0 ($e-3$)	RSE_ \mathcal{S}_0 ($e-3$)	# iter	Time (s)	RSE_ \mathcal{L}_0 ($e-3$)	RSE_ \mathcal{S}_0 ($e-3$)	# iter	Time (s)
0.05	4.3	4.3	133	13.7	4.6	4.6	208	17.8
0.15	9.8	4.2	162	20.8	10.5	4.7	235	33.6
0.25	12.3	4.3	514	44.9	15.9	5.2	676	51.0
0.35	56.5	9.5	654	53.3	67.5	11.4	737	57.5

TABLE 3: Comparison of ADMM-TR with RSTD for synthetic data where $\mathcal{A}_0 \in \mathbb{R}^{40 \times 40 \times 40}$, n -rank = [10, 10, 10].

spr	Algorithm: ADMM-TR				Algorithm: RSTD			
	RSE_ \mathcal{L}_0 ($e-3$)	RSE_ \mathcal{S}_0 ($e-3$)	# iter	Time (s)	RSE_ \mathcal{L}_0 ($e-3$)	RSE_ \mathcal{S}_0 ($e-3$)	# iter	Time (s)
0.05	4.4	2.0	236	26.4	4.7	2.5	338	29.9
0.15	4.7	2.1	417	42.1	5.2	2.5	603	50.8
0.25	8.9	2.9	664	60.8	12.0	4.2	663	53.6
0.35	17.3	5.2	981	80.6	21.1	6.5	1110	87.7

the dimension of the principal singular space as [22]. And the parameters are set as $\alpha = \beta = [I_1/I_{\max}, I_2/I_{\max}, \dots, I_n/I_{\max}]^T$ and $\gamma = 1/\text{sum}([I_1/I_{\max}, I_2/I_{\max}, \dots, I_n/I_{\max}]^T)$ for all experiments, where $I_{\max} = \max\{I_i\}$. η is set to $1/\sqrt{I_{\max}}$ as suggested in [23].

All the experiments are conducted and timed on the same desktop with an Pentium (R) Dual-Core 2.50 GHz CPU that has 4 GB memory, running on Windows 7 and MATLAB.

5.1. Synthetic Data. ADMM-TR and RSTD are tested on the synthetic data of size $40 \times 40 \times 40$. We generate the dimension r of a ‘‘core tensor’’ $\mathcal{C} \in \mathbb{R}^{r \times \dots \times r}$ which we fill with Gaussian distributed entries ($\sim \mathcal{N}(0, 1)$). Then, we generate matrices $U^{(1)}, \dots, U^{(N)}$, with $U^{(i)} \in \mathbb{R}^{n_i \times r}$ whose elements are also i.i.d. Gaussian random variables ($\sim \mathcal{N}(0, 1)$) and set

$$\mathcal{L}_0 = \mathcal{C}_{\times 1} U^{(1)}_{\times 2 \dots \times N} U^{(N)}. \quad (27)$$

The entries of sparse tensor \mathcal{S}_0 are independently distributed, each taking on value 0 with probability $1-spr$ and each taking on impulsive value with probability spr . We apply the proposed algorithm to the tensor $\mathcal{A}_0 = \mathcal{L}_0 + \mathcal{S}_0$ to recover \mathcal{L} and \mathcal{S} and compare with RSTD. For these experiments, two cases of n -rank are investigated, n -rank = [5, 5, 5] and n -rank = [10, 10, 10]. Table 1 presents the average results (across 30 instances) for different spr .

The quality of recovery is measured by the relative square error (RSE) to \mathcal{L}_0 and \mathcal{S}_0 , defined to be

$$\begin{aligned} \text{RSE}_{\mathcal{L}_0} &= \frac{\|\widehat{\mathcal{L}} - \mathcal{L}_0\|_F}{\|\mathcal{L}_0\|_F}, \\ \text{RSE}_{\mathcal{S}_0} &= \frac{\|\widehat{\mathcal{S}} - \mathcal{S}_0\|_F}{\|\mathcal{S}_0\|_F}. \end{aligned} \quad (28)$$

Tables 2 and 3 show that the proposed algorithm (ADMM-TR) is about 10 percent faster than RSTD proposed

in [22] and achieves better accuracy in terms of relative square error. Though both of the algorithms involve computing a SVD per iteration, we observe that the proposed algorithm take much fewer iterations than RSTD to converge to the optimal solution.

The more impulsive entries are added, that is, for higher value of spr , the less probable it becomes for the tensor recovery problem. In addition, the problem becomes more sophisticated when the n -rank is higher for the ground truth tensors of the same size. In Table 1, different spr is set for two tensor cases. And results show the recovered accuracy decreases as the spr grows for a certain case. In particular, the recovered accuracy for the tensor $\mathcal{A}_0 \in \mathbb{R}^{40 \times 40 \times 40}$ with n -rank = [5, 5, 5] decreases sharply when the spr is up to 40%, while the phenomenon occurs for tensor $\mathcal{A}_0 \in \mathbb{R}^{40 \times 40 \times 40}$ with n -rank = [10, 10, 10] when the spr is about 25%. This is due to that relative low rank and low sparse ratio are precondition of the tensor recovery problem.

5.2. Traffic Volume Data. To evaluate the performances of the proposed method in traffic volume data outlier recovery, a complete traffic volume data set is used as the test data set. We use the data of a fixed point in Sacramento County which is downloaded from <http://pems.dot.ca.gov/>. The traffic volume data are recorded every 5 minutes. Therefore, a daily traffic volume series for a loop detector contains 288 records, and the whole period of the data lasts for 16 days, that is, from August 2 to August 17, 2010.

Based on multiple correlations of the traffic volume data, we model the data set as a tensor model of size $16 \times 24 \times 12$ which stands for 16 days, 24 hours in a day, and 12 sample intervals (i.e., recorded by 5 minutes) per hour. The ratios of outlier data are set from 5% to 15% and the outlier data are produced randomly. All the results are averaged by 10 instances.

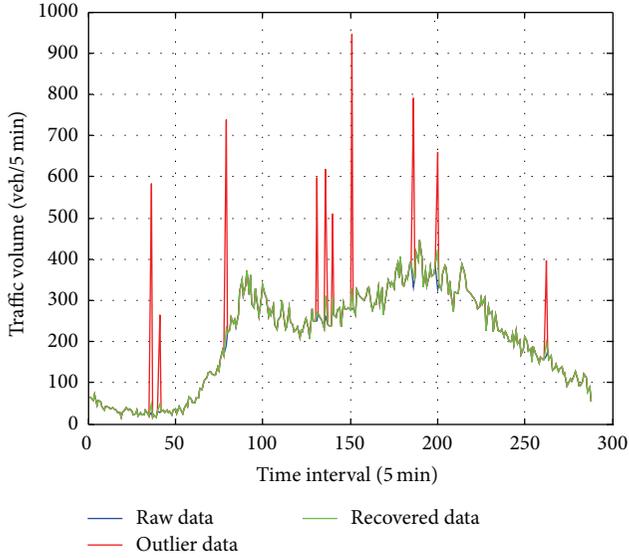


FIGURE 1: Comparisons with raw traffic volume data, data corrupted by outliers with 5% ratio, and data recovered by ADMM-TR.

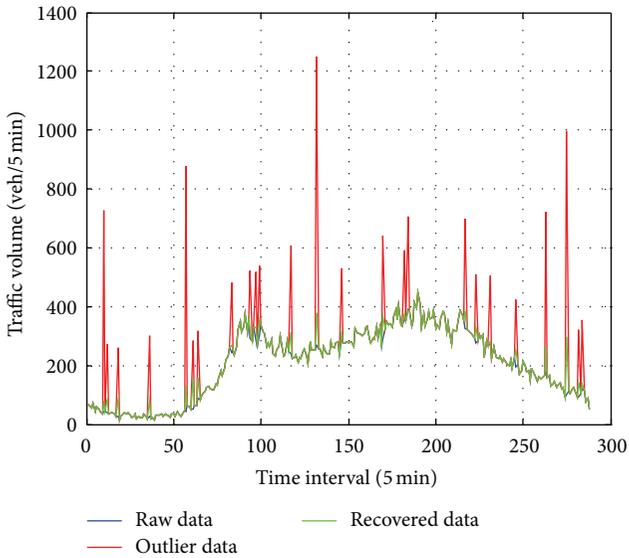


FIGURE 2: Comparisons with raw traffic volume data, data corrupted by outliers with 10% ratio, and data recovered by ADMM-TR.

For the real world data, we mainly pay attention to the correctness of the recovered traffic volume data. Thus the quality of recovery is measured by the relative square error (RSE) to \mathcal{L}_0 and Mean Absolute Percentage Error (MAPE) to \mathcal{L}_0 , defined to be

$$\begin{aligned} \text{RSE}_{\mathcal{L}_0} &= \frac{\|\widehat{\mathcal{L}} - \mathcal{L}_0\|_F}{\|\mathcal{L}_0\|_F}, \\ \text{MAPE}_{\mathcal{L}_0} &= \frac{1}{M} \sum_{m=1}^M \left| \frac{t_r^{(m)} - t_e^{(m)}}{t_r^{(m)}} \right|, \end{aligned} \quad (29)$$

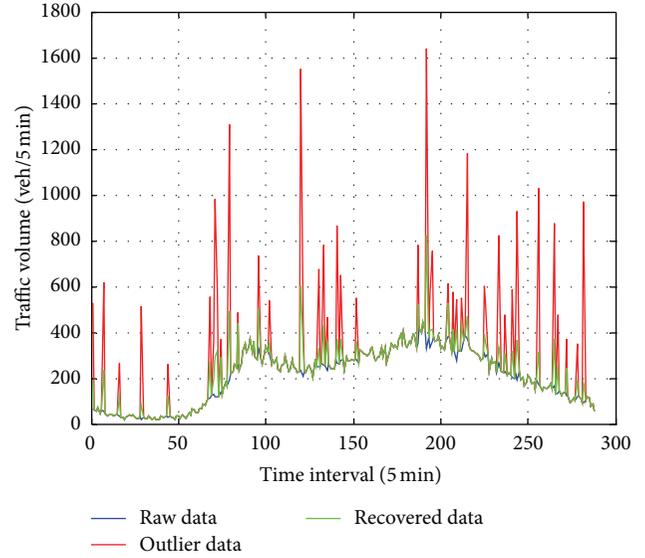


FIGURE 3: Comparisons with raw traffic volume data, data corrupted by outliers with 15% ratio, and data recovered by ADMM-TR.

TABLE 4: Traffic volume outlier data before and after being recovered.

spr	Before recovery		After recovery	
	$\text{RSE}_{\mathcal{L}_0}$ ($e-3$)	$\text{MAPE}_{\mathcal{L}_0}$ ($e-3$)	$\text{RSE}_{\mathcal{L}_0}$ ($e-3$)	$\text{MAPE}_{\mathcal{L}_0}$ ($e-3$)
0.05	0.5005	0.2136	0.0961	0.0314
0.10	0.7066	0.5008	0.1417	0.1027
0.15	0.8850	0.8777	0.2221	0.2351

where $t_r^{(m)}$ and $t_e^{(m)}$ are the m th elements which stand for the known real value and recovered value, respectively. M denotes the number of recovered traffic volumes.

Table 4 presents the relative errors of traffic volume outlier data before and after being recovered by ADMM-TR. The results show that the $\text{RSE}_{\mathcal{L}_0}$ and $\text{MAPE}_{\mathcal{L}_0}$ for traffic volume data corrupted by outlier data are about 5 times than the data after being recovered by ADMM-TR. Figures 1, 2, and 3 present the profiles of traffic volume data for a day. The results show that ADMM-TR could recover the traffic volume outlier data with perfect performance.

6. Conclusions

In this paper, we concentrate on the mathematical problem in traffic volume outlier data recovery and proposed a novel tensor recovery method based on alternating direction method of multipliers (ADMM). The proposed algorithm can automatically separate the low- n -rank tensor data and sparse part. The experiments show that the proposed method is more stable and accurate in most cases and has excellent convergence rate. Experiments on real world traffic volume data demonstrate the practicability and effectiveness of the proposed method in traffic research domain.

In the future, we would like to investigate how to automatically choose the parameters in our algorithm and explore

additional applications of our method in traffic research domain.

Acknowledgments

This research was supported by NSFC (Grant No. 61271376, 61171118, and 91120015), Beijing Natural Science Foundation (4122067). The authors would like to thank Professor Bin Ran from the University of Wisconsin-Madison and Yong Li from the University of Notre Dame for the suggestive discussions.

References

- [1] K. Tamura and M. Hirayama, "Toward realization of VICS—vehicle information and communications system," in *Proceedings of the IEEE-IEE Vehicle Navigation and Informations Systems Conference*, pp. 72–77, October 1993.
- [2] S. Kim, M. E. Lewis, and C. C. White III, "Optimal vehicle routing with real-time traffic information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 178–188, 2005.
- [3] P. Mirchandani and L. Head, "A real-time traffic signal control system: architecture, algorithms, and analysis," *Transportation Research Part C*, vol. 9, no. 6, pp. 415–432, 2001.
- [4] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transportation Research Record*, no. 1644, pp. 132–141, 1998.
- [5] J. Xu, X. Li, and H. Shi, "Short-term traffic flow forecasting model under missing data," *Journal of Computer Applications*, vol. 30, no. 4, pp. 1117–1120, 2010.
- [6] Y. Pei and J. Ma, "Real-time traffic data screening and reconstruction," *China Civil Engineering Journal*, vol. 36, no. 7, pp. 78–83, 2003.
- [7] S. Chen, W. Wang, and W. Li, "Noise recognition and noise reduction of real-time traffic data," *Dongnan Daxue Xuebao*, vol. 36, no. 2, pp. 322–325, 2006.
- [8] B.-G. Daniel, J. D.-P. Francisco, G.-O. David et al., "Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 7, pp. 530–545, 2010.
- [9] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, no. 2, Article ID 025010, 2011.
- [10] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [11] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Transactions of Image Processing*, vol. 1, no. 2, pp. 244–250, 1992.
- [12] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.
- [13] J. Wright, Y. Peng, Y. Ma, and A. Ganesh, "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," in *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS '09)*, pp. 2080–2088, Vancouver, Canada, 2009.
- [14] A. Ganesh, Z. Lin, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast algorithms for recovering a corrupted low-rank matrix," in *Proceedings of the 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP '09)*, pp. 213–216, December 2009.
- [15] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of a corrupted low-rank matrices," *Mathematical Programming*, 2009.
- [16] X. Yuan and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," Tech. Rep., Department of Mathematics, Hong Kong Baptist University, 2009.
- [17] L. Qu, Y. Zhang, J. Hu, L. Jia, and L. Li, "A BPCA based missing value imputing method for traffic flow volume data," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '08)*, pp. 985–990, June 2008.
- [18] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," in *Proceedings of the International Conference on Computer Vision (ICCV '09)*, 2009.
- [19] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation*, Prentice-Hall, 1989.
- [20] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [21] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1 -minimization: methodology and convergence," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1107–1130, 2008.
- [22] Y. Li, J. Yan, Y. Zhou, and J. Yang, "Optimum subspace learning and error correction for tensors," in *Proceedings of the 11th European Conference on Computer Vision (ECCV '10)*, Crete, Greece, 2010.
- [23] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, article 11, 2011.

Research Article

Simulation-Based Genetic Algorithm towards an Energy-Efficient Railway Traffic Control

Daniel Tuytens,¹ Hongying Fei,² Mohand Mezmaz,¹ and Jad Jalwan¹

¹ Department of Mathematics and Operational Research, University of Mons, 9 Rue de Houdain, B7000 Mons, Belgium

² School of Management, Shanghai University, No. 99 Shangda Road, Shanghai 200444, China

Correspondence should be addressed to Hongying Fei; feifmnot@gmail.com

Received 11 October 2012; Revised 29 December 2012; Accepted 29 December 2012

Academic Editor: Wuhong Wang

Copyright © 2013 Daniel Tuytens 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 real-time traffic control has an important impact on the efficiency of the energy utilization in the modern railway network. This study is aimed to develop an energy-efficient railway traffic control solution for any specified railway. In other words, it is expected to define suitable driving profiles for all the trains running within a specified period through the targeted network with an objective to minimize their total energy consumption. How to optimize the train synchronization so as to benefit from the energy regenerated by electronic braking is also considered in this study. A method based on genetic algorithm and empirical single train driving strategies is developed for this objective. Six monomode strategies and one multimode strategy are tested and compared with the four scenarios extracted from the Belgian railway system. The results obtained by simulation show that the multi-mode control strategy overcomes the mono-mode control strategies with regard to global energy consumption, while there is no firm relation between the utilization rate of energy regenerated by dynamic braking operations and the reduction of total energy consumption.

1. Introduction

In modern railway system, most of the energy required by trains is supplied by the electric network. In recent years, the reduction of energy consuming has become one of the main concerns of the railway managers, and thus more and more projects have been kicked off in this domain, such as RailEnergy (cf., <http://www.railenergy.org/>) and GreenRail (cf., <http://www.logisticsinwallonia.be/en/greenrail/>).

In practice, a targeted railway network is composed of tracks, stations or junctions, and trains running within this network according to a seasonal timetable. Stations and junctions are normally defined as referenced operational points in the railway system, and they will be abbreviated as OcpRefs in the rest of this paper. The seasonal timetable defines for every day during a certain period, normally one year, which trains run along which track at which time. All the information previously mentioned above is determined by the railway system management committee and cannot be changed except when perturbations occur.

Nowadays, at the traffic management center, the working place where dispatchers monitor the real-time railway status and solve traffic problems when necessary, traffic information is normally visualized in track diagrams on large distant panels or on computer screens. Train dispatchers monitor the railway occupation status and control the network by automatic or manual remote interlocking system. As far as we are concerned, today's railway management systems are mainly focused on supporting train dispatchers in solving disturbances and conflicts when they occur, but few, though some work has been observed in the literature, such as the study of Lüthi [1], can help dispatchers make optimal decision in practice in solving those problems and provide real-time traffic control solutions with an objective to minimize the network energy consumption. In fact, when a targeted railway network is defined and distributed to the related staff, the railway staff generally works in an empirical way because few supporting systems have been employed in traffic management centers for improving dispatchers' capability to deal with such processes.

Since the energy consumption of the railway network depends on the real-time operations of the trains that run within it, the optimization about the driving profile of each train running along their journey within the network is really critical for an energy-efficient traffic control strategy. This study is aimed at proposing suitable driving profiles for the trains that run in one railway network so as to reduce the overall energy consumption of the targeted railway network. The proposed solution can be used as a decision-aid tool when the railway managers apply their traffic control strategies.

In order to take necessary real-time traffic control constraints into consideration, several hypotheses are used in this study.

- (i) All the trains running through their journeys must respect the seasonal timetable, which is defined by the infrastructure manager. In the seasonal timetable, the time needed by each train passing through each OcpRef is determined. In consequence, no safety constraint is considered in this study though safety-based driving behavior is very important [2].
- (ii) Speed limits are predefined for all railway tracks.
- (iii) The type of one train passing through a certain OcpRef can be “pass” or “stop,” which is also determined in the seasonal timetable. For the OcpRef of type “pass,” the train can pass by at any speed not higher than the predefined speed limit; for the OcpRef of type “stop,” the train must arrive at and leave from the OcpRef according to the indicated arrival and departure times with a short stay.

In one word, one traffic control strategy is defined in this study as a set of real-time driving profiles, where each driving profile corresponds to one train involved in train journeys predefined in the seasonal timetable. Each train’s driving profile, along the journey between two successive OcpRefs, is obtained by an industrial simulator provided by one well-known train manufacturer. This simulator has been approved and used by this company. Our objective is to optimize, from a global view, this set of driving profiles so as to minimize the total energy consumption for the targeted railway network within a certain period.

This paper is organized as follows: at first, a typical framework of the centralized railway traffic management system and the main idea of train synchronization are introduced, and then a brief review of the related work has been made. In Section 4, definitions of mono-train driving and multi-train traffic control strategies are described. Afterwards, the proposed genetic algorithm for generating energy-efficient traffic control strategy is detailed in Section 5. The sixth part consists of simulated results for different scenarios with different traffic control strategies. At the end of the paper, conclusions and some further discussions are made with possible ideas and methods for future researches.

2. Railway Traffic Management System and Train Synchronization

The railway traffic control is a dynamic process undertaken by the train dispatchers to manage the railway in real time. In general, the railway traffic management system is centralized and has two fundamental components: traffic management center and traffic management territory [3]. Trains moving within a territory are supervised and controlled by the associated train dispatcher at the traffic management center.

Regarded as the kernel of the railway traffic management system, the dispatchers at the traffic management center are responsible for the regulation of railway traffic within their associated territory with the help of traffic control system, which helps the dispatchers to observe the status of the their associated territories (e.g., the occupation of line sections, location of trains, position of switches and aspects of signals, etc.) and collect necessary information in a continuous manner. Ideally, the traffic control system should also be able to help dispatchers to make decisions and communicate with both the upper level decision makers and the front line level operational staff in accordance with the rules and the regulations predefined by the railway authority. As mentioned in Figure 1, the train dispatchers supervise the status of their associated territories with real-time data collected by the traffic control system from the associated territories (e.g., availability of railway lines, stations and sidings, the status of switching between different operation systems (OSs), performances of signals and railway traffic, etc.). Additional information about the real-time traffic status may be collected or transferred by personal communication equipment. When unplanned events (e.g., trains’ unscheduled stops, new trains added to the current schedule, etc.) occur, the involved train dispatcher tries to configure and solve the problem, according to predefined rules and regulations of the railway traffic management, within acceptable time. When necessary, this train dispatcher has to communicate with the associated decision makers before making the final decision and transferring it to all involved people.

The main and traditional objective of the traffic control is not only to avoid conflicts between trains but also to restore the disrupted railway traffic as soon as necessary. In recent years, the third objective has arisen: synchronizing trains’ real-time operations so as to reduce the railway energy consumption in a dynamic perspective.

Although it seems that the traffic control center is not a major energy-consuming part of the railway system, it plays an important role in railway energy consumption management because the decisions made by dispatcher(s) impact directly on real-time operations of running trains, which are considered as the biggest energy-consuming units.

In addition, it is only the traffic control center that could make the decision from both global and dynamic perspectives so that train drivers would be able to “synchronize” their actions for not only avoiding unnecessary energy consumption but also reutilizing the energy produced by the dynamic braking nearby.

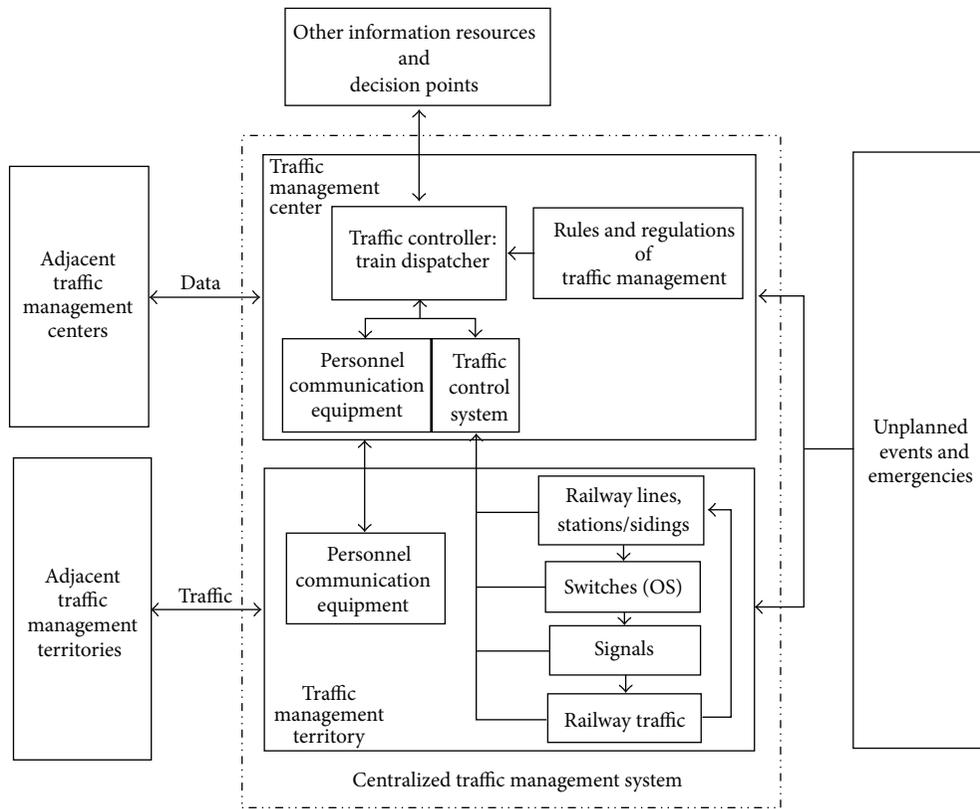


FIGURE 1: Interrelationship amongst the main components in the centralized traffic management system.

Train synchronization management is mainly concerned with the interactions between different trains in order to minimize the total energy consumption of the targeted railway network.

An ideal train synchronization solution should respect at first necessary constraints, such as the network security and the train punctuality, and meanwhile it should also minimize the network energy cost from a global perspective.

In the modern railway network, most of the trains are equipped with electric braking equipment. Dynamic braking is a general term used to describe the use of an electric motor as a generator to dissipate energy. This type of braking is more precisely described by one of the two terms: regenerative braking and rheostat braking. In regenerative braking, the electricity can be either transmitted through overhead catenary wires or an electrified third rail or can be stored onboard through the use of a flywheel, battery, or other energy storage system while Rheostat braking occurs when the produced electrical energy is run through resistors and dissipated as heat energy [4], since dynamic brakes only produce a retarding force when the wheels are rolling, and this force decreases as the rotational speed approaches zero. The wheel will be providing less and less braking force when its rotation slows down and finally the braking force becomes zero when it stops turning all together [5]; the dynamic braking is used to slow down the train rather than stop it. As shown by the two examples in Figure 2, when the regenerative braking is realized, the current generated by

the regenerated electric energy can be utilized by the trains running in the same substation over the same time period (such as those periods indicated by two-way arrows with solid lines). If no traction operations are produced in that period, the regenerated energies are lost in the electricity network except that some special devices are equipped to reserve them (such as those periods indicated by two-way arrows with dashed lines). In consequence, it is important to dynamically synchronize operations of trains running in the same substation so as to benefit as much as possible from the regenerated energy.

In general, a traffic control strategy is represented by a set of driving strategies that are defined for the trains that run within the targeted railway network. Two different traffic control strategies for three trains are shown in Figure 2, where the intervals indicated by two-way arrows (both dashed and solid lines) are the periods during which energies are generated by dynamic operations (descending arrows with solid lines) because trains run at a higher speed than their thresholds (vd_1 , vd_2 , and vd_3 , resp.) in those periods. Energies regenerated within the intervals indicated by the solid line two-way arrows can be reutilized by nearby trains for their traction operations (ascending and horizontal arrows), while those regenerated within the intervals indicated by dashed line two-way arrows cannot be reutilized by the nearby trains because no traction operation is produced. Neither energy consumption nor energy regeneration is observed during coasting phases (descending arrows with dashed lines).

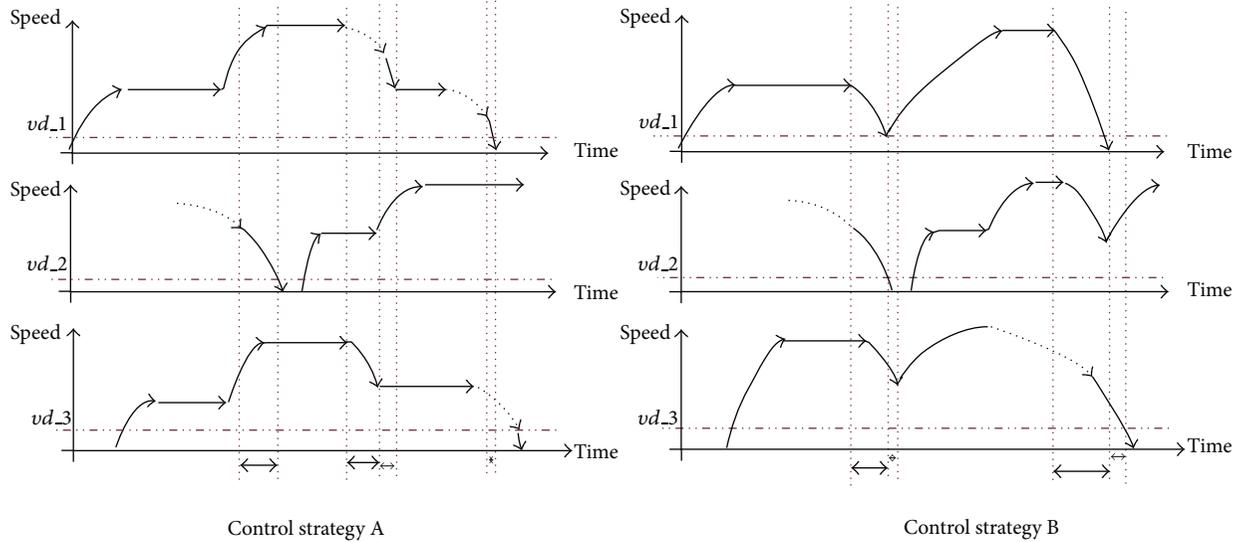


FIGURE 2: Two different control strategies for three trains.

It is observed in Figure 2 that the control strategy A may consume less energy than that of the control strategy B because more regenerated energies are utilized by the traction operations of the trains nearby; in other words, the train synchronization efficiency of control strategy A is higher than that of the control strategy B. Since most of the modern trains are equipped with dynamic braking devices, an energy-efficient traffic control strategy should not only reduce the single train's energy consumption but also increase the efficiency of the train synchronization. The objective is to minimize the global energy consumption of the targeted railway network.

3. A Brief Review of the Related Work

Although railway traffic control is an important topic in the modern railway management system, few results about the studies on railway traffic control, especially real-time energy-efficient traffic control for general railway system, have been published in the literature. The complexity of the railway traffic control system, related to the dynamic characteristic of the real-time train synchronization problem, can be regarded as the main reason.

According to the literature, the majority of the recent studies on railway traffic control are focused on avoiding conflicts within the railway network (e.g., [6–10]) or rescheduling trains for reducing negative impacts of disruptions (e.g., [11–14]). As one of the most lately raised topics in the railway traffic control, no clear definition can be found in the literature about real-time train synchronization, and few publications are dealing with the exact problem.

Nevertheless, some published results are observed in the literature about implicit train synchronization within a targeted railway network with an objective of minimizing the global energy consumption [15–22], though limitations are always observed.

- (1) Most of the studies simplify the problem by using average pacing velocities [15, 16] or constant speeds [17–20] to estimate the energy consumption of the trains along their journeys.
- (2) Some studies took into account the acceleration time between two different speeds when changing from one section to another but still suppose a constant speed for the train within the time interval [18].
- (3) Many studies were focused on avoiding train conflicts to reduce energy consumption [16, 18] while considering only the time reservation or margins of speeds instead of proposing driving profiles.
- (4) Determining time reserves first, and then optimizing the speed profiles in the second stages is considered reasonable and observed in several studies [17, 19, 21], while those studies consider the driving profile either in an independent way [17] or by means of setting constant speed for each section along the journey [19], or adjusting train running time instead of defining suitable driving profiles so as to reduce power peaks [21].
- (5) Although a well-developed integrated real-time rescheduling framework may help to save energy by reducing the number of unnecessary signal influences [1], the reutilization of the energy regenerated by dynamic braking was not considered.

In one word, a set of studies have been focused on the energy-efficient train scheduling, but few of them take into account both the train synchronization and the specific driving profile. For example, although Albrecht [21] took into account the energy regenerated by dynamic braking for the calculation of energy consumption, he just tried to modify the running time alongside the journey instead of proposing proper driving strategies for train drivers to reduce the power peak.

In fact, the railway energy-efficient traffic control is one of the most difficult optimization problems because not only the velocity profile for each running train should be dynamically defined but also all mandatory constraints keep changing. In addition, the energy consumption depends not only on the driving strategies but also on the infrastructure's condition (slope, curve, etc.), driving environment (tunnels, bridges, or open air), and the train's configuration (locomotive, weight, etc.). In this study, these impacts are considered as parameters for calculating energy consumption of the train with a specified driving strategy by using an industrial simulation tool.

As for resolution methods, decomposition methods (e.g., [17, 19, 21]), heuristics (e.g., [17, 18, 20]), and meta-heuristics (e.g., [16, 22]) are mainly used as optimization methods in the literature, while some other methods, such as simulation (e.g., [15, 20]), local strategy (e.g., [15]), and fuzzy rule-based methods (e.g., [14]) are also observed. In fact, even if some real-time traffic control issues can be theoretically solved by some exact methods when all the parameters are determined, the considerable running time of exact methods makes them impractical in real application because of the complexity of such problems. That is why there is no room for time-consuming method in this case, and most of the researchers are interested in meta-heuristics or heuristics. The decomposition idea looks efficient, but the problem raised is how to reduce the error amplification during the resolution procedure. Furthermore, it is important to find an efficient decomposition perspective.

In order to obtain real-time energy-efficient traffic control strategy within reasonable execution time, we are focused on heuristic and meta-heuristic methods. Furthermore, encouraged by the good performance of methods based on genetic algorithm (e.g., [21, 22]), we have developed a method based on genetic algorithm and simulation output to solve the traffic control problem.

4. Construction of Traffic Control Strategies

As previously mentioned, a traffic control strategy consists of driving strategies defined for all the involved trains. With a given driving strategy, the train driver will be informed about the details of the driving profile suggested to him during the following journey. In fact, a driving strategy is mainly determined by two important parameters.

- (i) Traction coefficient: coefficient of pulling effort of train's engine. When the engine performs a full traction effort, the coefficient is 1; otherwise, it is a real number between 0 and 1.
- (ii) Driving mode: it represents a rule used to determine a sequence of driving operations, such as traction, coasting, cruising, and braking operations, which are performed by the driver along the trajectory.

In this study, a traffic control strategy is defined as a set of mono-train driving strategies applied to all the trains involved in train journeys observed in the targeted railway network. According to the literature, several energy-saving

mono-train driving strategies have been proposed by using dynamic programming [23] or evolutionary algorithm [24]. Since the energy consumption depends not only on how a train performs but also on the environment, such as the railway network conditions, it is impossible to define a universal optimal driving strategy for any train under any conditions, and it takes considerable time to calculate for each train its optimal strategy under the specific conditions, which are not suitable for real-time multi-train control problem because the details of the updated driving strategies have to be sent to train drivers in a short time; otherwise, the decisions will be no more available because all the trains are running all the time within the network and their situations change rapidly. Furthermore, since some empirical driving modes are normally proposed by experts, who are working in the railway system, to train drivers, those driving modes have been already realized and known well by them. In consequence, the proposed traffic control strategy proposed in this study is based on empirical driving modes, which can save time of calculation and are familiar enough to get accepted by train drivers.

It should be mentioned that empirical driving modes mostly based on the maximum principle as Howlett and Pudney stated [25], and the experimental results of Miyatake and Ko [23] mentioned that the maximum principle is not quite appropriate for the use of power recovery, and therefore a power-reduced driving mode, where the traction effort is reduced to 80%, is constructed to each studied empirical driving mode with an aim at getting solutions with acceptable quality.

4.1. Empirical Mono-Train Driving Modes. In this study, the three most used empirical mono-train driving modes are employed.

- (i) Mode 1: cruising mode. Only one universal sequence of the train operations is observed within any of the railway segments that compose the journey: traction, cruising, and braking operations whereby the driver has to reduce the train's speed.
- (ii) Mode 2: coasting mode. Starting with a traction operation, the driver performs alternatively traction and coasting operations along a specified journey, where upper and lower bounds of the running speed are specified. The conditions involved in operation changes are as follows.
 - (a) If the speed of the train is lower than the upper bound, the traction operation is implemented until the speed reaches the upper limit, and then the coasting operation is applied until the speed of the train reaches the lower bound, where the train will be powered again. These two kinds of operations are implemented alternatively until the train passes into a section with different upper and lower bounds of the speed. A braking operation may take place if the coasting operation is not enough to reduce train's speed to the upper bound of the speed at the next section.

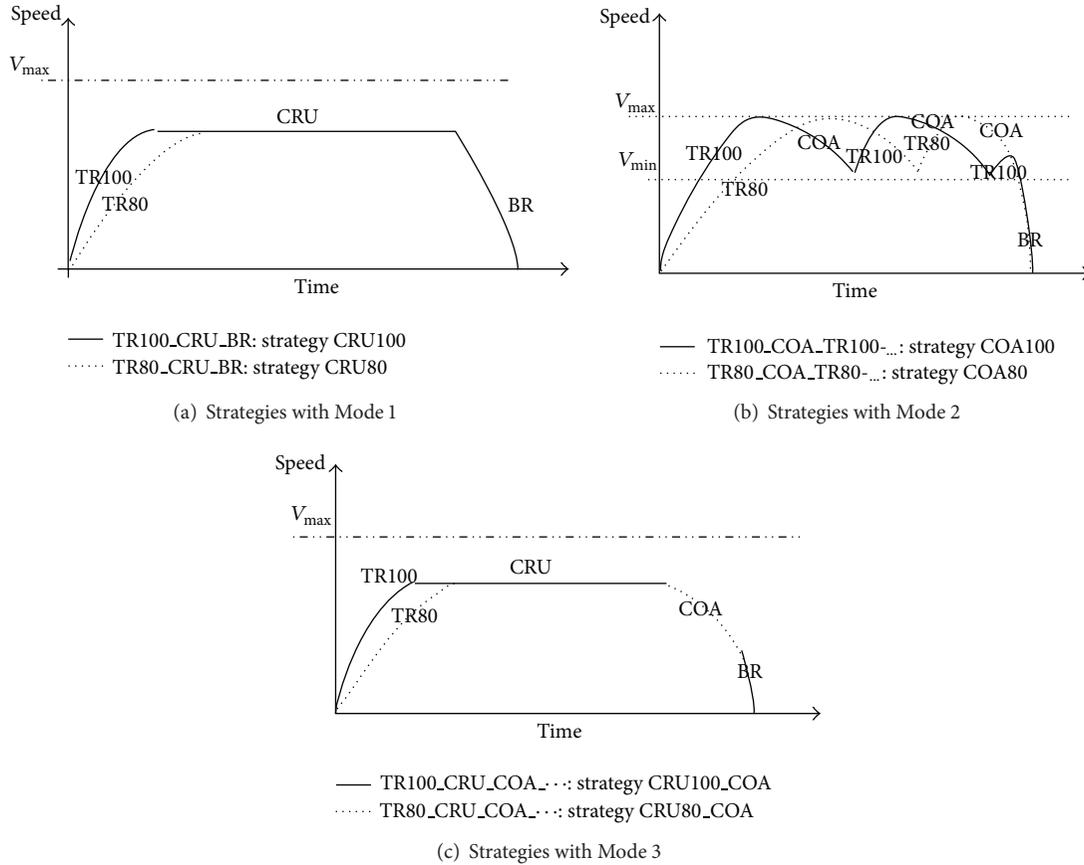


FIGURE 3: Mono-train driving strategies.

In that case, the braking operation will force the train to reduce its speed to a reasonable value.

- (iii) Mode 3: Cruising-coasting mode. In general, a train running with this mode is conducted by a set of operations in the order of traction, cruising, and coasting within each segment of its journey. Braking operation can be only observed when the train shows a risk of exceeding the upper bound of the speed limit or when it is beyond the capacity of the coasting operation to slow down the train to the specified arrival speed at its destination.

It is obvious that one train may have different driving profiles with different traction efforts even with the same mono-train driving mode. According to the practice, the two most performed traction efforts are applied to generate mono-train driving strategies: full traction effort (100%) and reduced traction effort (80%).

As shown in Figure 3, six combinations of mono-train driving mode and traction efforts are defined as basic mono-train driving strategies for generating multi-train traffic control strategy in this study by combing three basic empirical modes with two possibilities of traction effort, respectively: CRU100, CRU80, COA100, COA80, CRU100_COA, and CRU80_COA, where the first three letters represent

the involved driving mode and the numbers stand for the rate of traction effort. For example, CRU100 represents the combination of Mode 1 and full traction effort.

4.2. Traffic Control Strategies Based on Empirical Mono-Train Driving Strategies. Considering that the mono-train driving strategies used for different trains during their journeys can be either universal or different, in this study, seven traffic control strategies in total are defined: six of them are monomode strategies, where one universal driving mode is defined for all the trains during their journeys, and the remaining one is multimode strategy, where different mono-train driving strategies, selected among six mono-train driving strategies mentioned previously, are used to generate driving profile for the trains observed in the targeted network.

5. The Genetic Algorithm for Generating Energy-Efficient Traffic Control Strategy

5.1. Framework of the Proposed Genetic Algorithm. As shown in Figure 4, the procedure of the proposed genetic algorithm, used for generating energy-efficient traffic control strategy, is as follows.

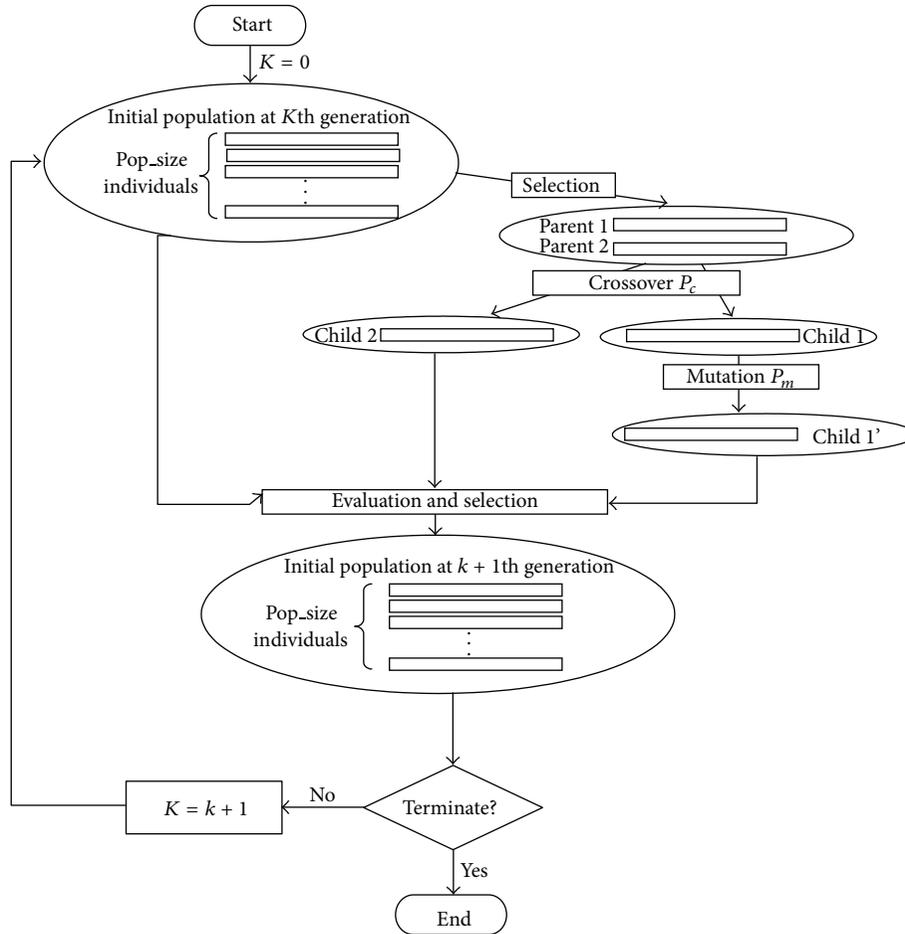


FIGURE 4: Framework of the proposed genetic algorithm.

Step 1. Construct an initial population. Each individual of the population represents a feasible traffic control strategy, that is, a set of feasible driving profiles proposed for trains involved in the targeted network within the specified period. All necessary constraints, such as the passing time noted in the timetable, and speed limits set along the line, are respected in each individual. The coding strategy of one individual is detailed in Section 5.2.

Step 2. Calculate each individual's fitness value. The fitness value is used to confirm that the smaller an individual's fitness is, the more likely it will be selected. In this study, the global energy consumption of a traffic control solution is used as its fitness.

Step 3. Select randomly two individuals by making a roulette wheel of the fitness array. These two individuals will be used as parents for the crossover operator.

Step 4. Recombine the selected parents with a predefined crossover probability, P_c , and generate two new individuals; if no crossover takes place, copy the original information of these parents to their children.

Step 5. Randomly select one newly generated child. Mutate it with a predefined mutation, probability P_m , to generate a new individual.

Step 6. Evolve the population for the next iteration by making roulette wheel selection.

Step 7. The individual, who has the best fitness among all the current individuals, replaces one individual, which is randomly selected from the current selected population. This step confirms that the present best individual will be a member of the new population.

Step 8. If the iteration number or the runtime reaches the predetermined limit or the current population is converged (its rate of convergence is larger than a predefined threshold), this procedure is terminated and the one with the best fitness is reported as the final traffic control solution; otherwise, go to Step 2.

5.2. Coding Strategy. Make a train journey represent a train running between two successive OcpRefs within a period specified in the periodical timetable. Suppose that N train

$V_s(0), C(0), M(0), V_f(0)$...	$V_s(i), C(i), M(i), V_f(i)$...	$V_s(N), C(N), M(N), V_f(N)$
------------------------------	-----	------------------------------	-----	------------------------------

FIGURE 5: Structure of an individual representing a traffic control strategy for a railway network containing N train journeys.

journeys are observed passing through the targeted railway network in a constructed scenario. One traffic control strategy, named *sd* an individual, is coded as a structural array that consists of N elements. Each element, named as a gene, represents a specified train journey. At the operational level, the features of one train journey, such as positions of departure and arrival, running duration, and characteristics of railway lines along the journey and the used locomotive, are already determined, and thus the energy consumption of this train journey is determined. When all the train journeys are well defined, the global energy consumption of a specified traffic control strategy can be obtained by summing up the energy consumed by all train involved in the defined train journeys. It should be mentioned that because both network configuration and locomotives are planned in the tactical decision level while our study is focused on the operational level, only four of those parameters can be updated in this study to get an energy-efficient traffic control strategy. These variables are defined as follows:

$V_s(i)$: passing speed of train involved in the train journey i at the begin of its running section;

$V_f(i)$: passing speed of train involved in the train journey i at the end of its running section;

$C(i)$: coefficient of traction effort for train involved in the train journey i ;

$M(i)$: type of driving mode for train involved in train journey i .

According to the individual structure shown in Figure 5, the train involved in train journey i ($0 \leq i < N$) is planned to pass through its departure and arrival points at the speeds $V_s(i)$ and $V_f(i)$, respectively. During its journey, the rate of traction effort is $C(i)$, and the driving mode is $M(i)$.

5.3. Variation Operators. The aim of the variation operators is to maintain genetic diversity from one generation of a population to the next, while we overcome the infeasibility of the new individual generated by the variation operation with respect to the constraints.

In the proposed genetic algorithm, two kinds of variation operators are defined: one-point crossover and point mutation. When crossover occurs, a single crossover point on both parents' chromosomes is selected and the corresponding data beyond that point in either chromosome is swapped between two parents, while only the gene at the selected point is updated with the newly generated data for mutation.

Although more than one factor of the gene can be varied to generate a new one because each gene contains four elements (driving mode, rate of traction effort, and departure and arrival speeds), only one of them, randomly chosen, is varied within the allowed range at one iteration. The range of variation can be either the predefined candidate list or the

limited value. For example, the driving mode varies within the range of {CRU, COA, or CRU_COA}; the rate of traction effort can be either 1.0 or 0.8; the departure or arrival speed should be either fixed as zero at an OcpRef of type "stop" or varied from zero to the speed limitation at an OcpRef of type "pass."

Traditionally, after swapping the corresponding parts of two parents during crossover operation, two children are generated. When mutation occurs, one new individual is generated by updating the selected gene. Those newly generated solution can be used as candidate(s) for generating the new population. However, in this study, the traditional variation operators may generate infeasible solutions that do not satisfy any more some constraints. In consequence, some additional operations must be guaranteed that all the solutions generated by the variation operators are feasible ones.

Feasibility check is necessity when train's departure or arrival speed is updated. A feasible solution should satisfy the following constraints.

- (i) Constraint 1: the predefined running time should be respected.
- (ii) Constraint 2: each train's speed should be always below the speed limitation of the associated section.
- (iii) Constraint 3: the arrival and departure speeds of one train at an OcpRef should be the same.

Constraint 1 may be violated when the driving mode is transferred from CRU to COA or the rate of traction effort is reduced; in this case, the actual choice will not be accepted, and the other solution will be generated with different variation factor until Constraint 1 is satisfied. If there are no more candidates in the list, keep the original solution.

Constraint 2 can always be respected if the new value is generated from a bounded range.

The feasibility of the new solution with regard to Constraint 3 depends on both the selected variation element and the type of OcpRefs. If both OcpRef are of the type "stop," the new solution will have no chance to violate this constraint; if one OcpRef is of the type "pass," a state that the involved train can pass through this OcpRef without stop, a further check should be made by the comparison between the departure and the arrival speeds of the involved train at OcpRefs of the type "pass" in both new children generated by crossover operation (or the new individual generated by the mutation operation) with the arrival speed of this train in the preceding section and its departure speed in the following section, respectively. If train's departure speed is always equal to its arrival speed at the same OcpRef, the new solution is accepted as a feasible one; otherwise, replace the arrival speed or the departure speed of the involved train in the adjacent section by the corresponding value associated with the selected gene

TABLE 1: Brief configuration of different scenarios.

Number of scenario	Main station	Period	Number of train journeys
1	Gent-St-Pieters	07h00–07h30	52
2	Gent-St-Pieters	11h00–11h30	58
3	Namur	07h00–07h30	39
4	Namur	11h00–11h30	22

when necessary so as to ensure that the train's arrival speed is the same as the departure speed when it passes through an OcpRef.

Considering that the mutation operator used in this study is much more complex than the traditional point mutation, the steps taken by the dedicated mutation operator are detailed in the following for a better understanding.

Step 1. Set the initial candidate list of variation factors. Theoretically, the set should contain all the four elements of the gene, that is, $\{V_s, M, C, V_f\}$. However, the candidate list may be reduced according to the types of OcpRefs associated with the selected point because once an OcpRef is of the type “stop,” the train must execute “stop” at this point, that is, its speed being fixed as zero. In consequence, V_s and V_f at the OcpRef of the type “pass” should be excluded from the list of candidates for variation operator.

Step 2. Randomly select one element from the candidate list for the mutation operation.

Step 3. Vary the selected element to construct a new solution.

- (i) If M is selected as variation factor, the driving mode of the selected gene will be replaced by another driving mode randomly selected among three basic empirical driving modes from 1 to 3.
- (ii) If C is selected, the selected gene's actual traction effort will be updated by a value randomly selected from 1 and 0.8.
- (iii) If V_s is selected, the selected gene's actual departure speed will be replaced by a randomly generated value, which is positive and respects the speed limitation.
- (iv) If V_f is selected, the arrival speed of the involved gene will be updated with a positive value randomly generated between 0 and the corresponding speed limitation.

Step 4. check the feasibility of the newly generated solution; if the solution generated at Step 3 is feasible, the mutation operator is terminated. If an infeasible solution is generated, it may result from a reduced traction effort or the change of speeds at OcpRef of the type “pass.” The corrections are made as follows:

- (i) If C is varied from 1 to 0.8, first make $C = 1$, and then exclude M from the candidate list for variation operation. If the candidate list becomes empty, this

operation will be terminated without variation; otherwise, go to Step 2 with a reduced candidate list.

- (ii) If V_s or V_f is selected as variation factor, update the V_f of the involved train in the precedent section or its V_s in the following section to avoid the conflict caused by defining different speeds for the train at the same OcpRef. Once the speed conflicts are solved, the mutation is terminated with a feasible solution.

6. Experimental Results

6.1. Data. Experimental data used in this study are supplied by the Belgian railway infrastructure manager who handles the train traffic control in Belgium. Most of the Belgian railway lines are electrified. The majority of the electrified railway lines use 3000 volt (DC) overhead power supply, and only some high-speed lines are electrified at 25000 volt (AC). Since the high-speed lines are separated from the normal lines by using special railways and this study is focused on the normal lines, all the experimental data used here are concerned with railway lines electrified at 3000 volt (DC). In addition, because of the loss of energy when it is transported in the wire, it is reasonable to suppose that the energy produced by the dynamic braking of one train can only be reutilized by the trains with a distance of less than 30 kilometers from it.

In this study, as shown in Figure 6, two railway networks are constructed by selecting railway lines around two railway stations in Belgium, Gent-St-Pieters and Namur. The former is a main station in a relatively dense network, while the latter is a normal one according to the periodical timetable of the year of 2011. For each defined railway network, two sets of train journeys are selected for the specified periods: train journeys observed from 07h00 to 07h30 and those observed from 11h00 to 11h30. In consequence, as shown in Table 1, four scenarios are in total constructed with train parts selected and cleaned according to different conditions and constraints. It should be mentioned that since the distance between the two selected stations is far more than 60 km, it is possible to construct two distinct railway networks by using tracks around these two stations, as shown in Figure 6.

The parameters of the genetic algorithm are set as follows:

- (i) crossover probability: 0.8,
- (ii) mutation probability: 0.1,
- (iii) rate of convergence: 95%,
- (iv) maximum runtime: 2 minutes,
- (v) maximum number of iterations: 500.

6.2. Numerical Results. Four scenarios, mentioned in Section 6.1, are tested with six monomode control strategies and one multimode strategy. Since genetic algorithm is a meta-heuristic, the final solution obtained is rather an approximate solution than an optimal one, and therefore ten trails are executed for each scenario.

As shown in Table 2, only two kinds of traffic control strategies can be feasible solutions for all constructed scenarios within all trials: one uses CRU100 strategy as the universal

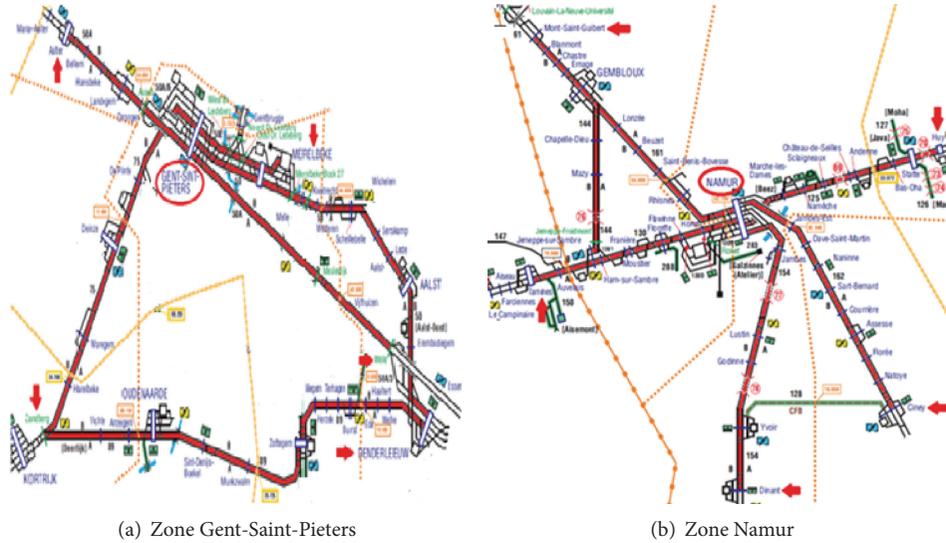


FIGURE 6: Railway networks constructed around two selected stations in Belgium.

TABLE 2: Energy consumed at pantographs.

Number of scenarios	Total energy consumed at pantographs (unit: kwh)			Total energy consumed at pantographs for traction operations (unit: kwh)		
	CRU100	CRU100_COA	MULTI	CRU100	CRU100_COA	MULTI
1	5972.17	3935.95	3932.58	6670.52	4404.61	4396.08
2	9444.85	—	7642.59	10163.06	—	8230.83
3	3828.01	3040.84	2744.41	4527.23	3575.10	3288.28
4	2059.21	1410.12	1263.64	2460.54	1676.83	1505.93

mono-train strategy and the other is a multimode control strategy (MULTI), where each train part has its proper mono-train driving strategy. The traffic control strategy, where CRU100_COA is used as the universal mono-train strategy for all the train parts, can obtain feasible solutions for most of the constructed scenarios but not for all. In comparison with total energies consumed by feasible solution obtained with different traffic control strategies, MULTI traffic control strategy can always obtain the solutions with minimal energy consumption (about 70% of that consumed by solutions obtained with CRU100 traffic control strategy).

It is observed that the total energy consumed at pantographs is less than the total energy consumed at pantographs for traction operations for all the obtained solutions; that is, the energy regenerated by dynamic braking operations is more or less used among different trains.

A further analysis is made about the relation between the reutilization of energy generated by dynamic braking operations and the total energy consumption.

As shown in Table 3, it is observed that high utilization rate of energy regenerated by dynamic braking operations is observed in all solutions. However, there is no firm relation between the utilization rate of energy regenerated by dynamic braking operations and the reduction of total energy consumption.

TABLE 3: Reutilization of energy regenerated by dynamic braking operations.

	Rate of total energy consumption to energy consumed by traction operations (%)	Utilization rate of energy regenerated by dynamic braking operations (%)
CRU100	87.68	92.32
CRU100_COA	86.17	81.75
MULTI	87.42	80.25

In consequence, it is concluded that a proper combination of different mono-train driving strategies can help reduce the total energy consumption of the targeted network. A high utilization rate of energy regenerated by dynamic braking operations is not necessary for minimizing the total energy consumption. The proposed method can propose a set of energy-efficient driving profiles for train parts observed in the targeted network according to experimental results.

7. Conclusions and Perspectives

This study is aimed to propose a method that can obtain an energy-efficient traffic control strategy for any targeted

railway network. Since little work observed in the literature to optimize the synchronization of trains' operations though it is very important in the real world, we are focused on this aspect in this study. In addition, considering that the time consuming method is not practical for real-time traffic control, a method based on genetic algorithm is proposed to generate a proper traffic control solution within reasonable runtime.

According to the experimental results, it is concluded that the proposed method can generate energy-efficient traffic control solution, where the driving profile of a train part is defined by a suitable mono-train driving strategy that can be different from the driving strategy used by another train part in the same railway network.

Furthermore, it is concluded that although a high utilization rate of energy regenerated by dynamic braking operations can always be observed in the energy-efficient traffic control solution, the optimization of mono-train driving profile is more important than the reutilization of energy generated by dynamic braking operations because a higher value of the latter is not necessary for a traffic control solution to have the minimal total energy consumption.

In one word, in order to minimize the total energy consumption of a targeted railway network, it is important to not only propose ecodriving driving profile for each train part but also synchronize the operations of different train parts to introduce a high utilization rate of the energy regenerated by dynamic braking operations.

It should be mentioned that this study does not yet consider the estimation of the disturbances in the railway network though the quality of the solution depends on the input parameters that predict the real status of the railway network. Therefore, it is interesting to develop the future study based on the prediction of the status of railway network at the point of time when the proposed solution would be implemented.

In addition, some other objectives such as punctuality are also important in practice, and therefore multi-objective optimization should be an interesting research topic as well. The driving modes of trains should be provided in a real-time manner, whereas it takes the current algorithm considerable execution time to get results. In consequence, it is important to reduce the running time of the algorithm. In fact, the parallelization of the algorithm may be helpful to find good results in a reasonable timeframe.

Furthermore, several works about synchronization of trains in mass transit systems are observed in the literature. Though much more constraints should be taken into account in the open railway system, the approaches proposed in the former can give some ideas about train synchronization for the latter, thus how to get benefit from the successful experiences in the mass transit systems will be another interesting topic of research in the future.

Acknowledgment

This work is supported by the Belgian Walloon Region through the GreenRail Project.

References

- [1] M. Lüthi, "Improving the efficiency of heavily used railway networks through integrated real-time rescheduling," Dissertation Eidgenössische Technische Hochschule 18615, 2009.
- [2] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [3] I. Şahin, "Railway traffic control and train scheduling based on inter-train conflict management," *Transportation Research B*, vol. 33, no. 7, pp. 511–534, 1999.
- [4] B. L. Judy and A. V. Johansson, *Dynamic Braking, in Diesel-Electrics: How To Keep'em Rolling*, Simmons-Boardman, New York, NY, USA, 1954.
- [5] H. A. John, *The Railroad: What it Is, What it Does*, Simmons-Boardman, Omaha, Neb, USA, 2000.
- [6] D. Huisman, L. G. Kroon, R. M. Lentink et al., "Operations research in passenger railway transportation," *Statistica Neerlandica*, vol. 59, no. 4, pp. 467–497, 2005.
- [7] J. Törnquist, "Computer-based decision support for railway traffic scheduling and dispatching: a review of models and algorithms," in *Proceedings of 5th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS '05)*, 2005.
- [8] A. Lova, P. Tormos, F. Barber, L. Ingolotti, M. A. Salido, and M. Abril, "Intelligent train scheduling on a high-loaded railway network," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4359, pp. 219–232, 2007.
- [9] J. Rodriguez, "A constraint programming model for real-time train scheduling at junctions," *Transportation Research B*, vol. 41, no. 2, pp. 231–245, 2007.
- [10] R. Acuna-Agost, *Mathematical Modeling and Methods for Rescheduling Trains Under Disrupted Operations*, University of Avignon and the Vaucluse, Avignon, France, 2009.
- [11] J. Jespersen-Groth, D. Potthoff, J. Clausen et al., "Disruption management in passenger railway transportation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5868, pp. 399–421, 2009.
- [12] S. Mladenovic and M. Cangalovic, "Heuristic approach to train rescheduling," *Yugoslao Journal of Operations Research*, vol. 17, no. 1, pp. 9–29, 2007.
- [13] J. Törnquist and J. A. Persson, "N-tracked railway traffic rescheduling during disturbances," *Transportation Research B*, vol. 41, no. 3, pp. 342–362, 2007.
- [14] A. Tazoniero, R. Gonçalves, and F. Gomide, "Decision making strategies for real-time train dispatch and control," in *Analysis and Design of Intelligent Systems using Soft Computing Techniques*, P. Melin, O. Castillo, E. G. Ramirez, and W. Pedrycz, Eds., vol. 41 of *Advances in Intelligent and Soft Computing*, pp. 195–204, 2007.
- [15] M. J. Dorfman and J. Medanic, "Scheduling trains on a railway network using a discrete event model of railway traffic," *Transportation Research B*, vol. 38, no. 1, pp. 81–98, 2004.
- [16] T. Albrecht, "The influence of anticipating train driving on the dispatching process in railway conflict situations," *Networks and Spatial Economics*, vol. 9, no. 1, pp. 85–101, 2009.
- [17] M. Mazzarello and E. Ottaviani, "A traffic management system for real-time traffic optimisation in railways," *Transportation Research B*, vol. 41, no. 2, pp. 246–274, 2007.

- [18] A. Mascis, D. Pacciarelli, and M. Pranzo, "Scheduling models for short-term railway traffic optimisation," *Lecture Notes in Economics and Mathematical Systems*, vol. 600, pp. 71–90, 2008.
- [19] G. Caimi, M. Fuchsberger, D. Burkolter, and T. Hermann, "Conflict-free train scheduling in a compensation zone exploiting speed profile," in *Proceedings of the 3rd International Seminar on Railway Operations Modelling and Analysis Rail*, February 2009.
- [20] Y. Ding, Y. Bai, F. M. Liu, and B. H. Mao, "Simulation algorithm for energy-efficient train control under moving block system," in *2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009*, pp. 498–502, USA, April 2009.
- [21] T. Albrecht, "Reducing power peaks and energy consumption in rail transit systems by simultaneous train running time control," in *Proceedings of the 9th international conference on Computer Aided Design, Manufacture and Operation in the Railway and Other Advanced Mass Transit Systems (COMPRAIL '04)*, J. Allan, R. J. Hill, C. A. Brebbia, G. Sciutto, and S. Sone, Eds., *Computers in Railways VIII*, pp. S885–S894, WIT Press, Dresden, Germany, May 2004.
- [22] Z. Miao, K. Fu, and F. Yang, "A hybrid genetic algorithm for the multiple crossdocks problem," *Mathematical Problems in Engineering*, vol. 2012, Article ID 316908, 18 pages, 2012.
- [23] M. Miyatake and H. Ko, "Optimization of train speed profile for minimum energy consumption," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 5, no. 3, pp. 263–269, 2010.
- [24] R. Chevrier, G. Marlière, and J. Rodriguez, "Saving energy in railway management with an evolutionary multiobjective algorithm: application and case study," in *Proceedings of the 9th World Congress on Railway Research*, Lille, France, May 2011.
- [25] P. G. Howlett and P. J. Pudney, *Energy-Efficient Train Control*, Springer, Berlin, Germany, 1995.

Research Article

A Neural Network Model for Driver's Lane-Changing Trajectory Prediction in Urban Traffic Flow

Chenxi Ding,^{1,2} Wuhong Wang,¹ Xiao Wang,¹ and Martin Baumann²

¹ Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

² Institut für Verkehrssystemtechnik, Deutsche Zentrum für Luft- und Raumfahrt, Lilienthalplatz 7, 38108 Braunschweig, Germany

Correspondence should be addressed to Wuhong Wang; wangwuhong@bit.edu.cn

Received 13 September 2012; Accepted 10 November 2012

Academic Editor: Huimin Niu

Copyright © 2013 Chenxi Ding 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 neural network may learn and incorporate the uncertainties to predict the driver's lane-changing behavior more accurately. In this paper, we will discuss in detail the effectiveness of Back-Propagation (BP) neural network for prediction of lane-changing trajectory based on the past vehicle data and compare the results between BP neural network model and Elman Network model in terms of the training time and accuracy. Driving simulator data and NGSIM data were processed by a smooth method and then used to validate the availability of the model. The test results indicate that BP neural network might be an accurate prediction of driver's lane-changing behavior in urban traffic flow. The objective of this paper is to show the usefulness of BP neural network in prediction of lane-changing process and confirm that the vehicle trajectory is influenced previously by the collected data.

1. Introduction

The most crucial road traffic problems requiring solutions have been the reduction of traffic accidents and traffic congestion [1]. Increasing the safety level of driving in traffic, especially the safety of maneuvers such as lane-changing and overtaking, is one of the key technologies for the Intelligent Transportation System (ITS) to achieve congestion-free and accident-free traffic situations [2]. Every year, traffic accidents result in approximately 1.2 million fatalities worldwide; without new prevention measures, this number could increase by 65% over the next two decades [3]. Researchers estimate that lane-changing crashes account for 4% to 10% of all vehicle crashes in the USA [4]. Although, this value is not very high, the delay time it causes accounts for 10% of the total time caused by all traffic accidents.

Due to the importance of driving behavior to vehicle safety, many researchers have attempted to model driving behavior. A general trend in the study of modeling driving behavior is the greater application of computational artificial intelligence. Because the driver's mental and physical behavior is nondeterministic and highly nonlinear, it is

difficult for traditional methods to embody this kind of uncertain relationship. The Artificial Neural Networks, fuzzy logic theory, and dynamic Bayesian networks, which include well-known hidden Markov models, have attracted many researchers to do relative research. Kumagai et al. focused on the prediction of drivers' intentions of stopping the car at an intersection with their current and historical maneuvers based on a simple dynamic Bayesian Network [5]. Tezuka et al. developed a method to infer driver behavior with a driving simulator to evaluate continuous time-series steering angle data at the time of lane-changing. The proposed method used a static type conditional Gaussian model on Bayesian Networks [6]. Kuge et al. proposed hidden Markov models (HMMs) using observations of vehicle parameters and lane positions to model trajectories [7]. Sathyanarayana et al. proposed a method to model driver behavior signals using hidden Markov models. The hierarchical framework and initial results can encourage more investigations into driver behavior signal analysis and related safety systems employing a partitioned submodule strategy [8]. Pentland and Andrew proposed that many human behaviors can be accurately described as a set of dynamic models sequenced

together by a Markov chain. They considered the human as a device with a large number of internal mental states and used the dynamic Markov models to recognize human behaviors from sensory data and to predict human behaviors a few seconds into the future [9]. Macadam and Johnson demonstrated the use of elementary neural networks (a two-layer back propagation) to represent the driver's steering behavior in double lane-changing maneuvers and S-curve maneuvers. Due to the limited data source for neural networks, it was concluded that the adaptive nature of neural networks should be used for modeling driver steering behavior under a variety of operation scenarios [10]. Cheng et al. used a Back-Propagation neural network as a controller for an automated vehicle system. Camera images were used as inputs to the neural network [11]. Tomar et al. proposed a method to give the future lane-changing trajectories accurately for discrete patches using a multilayer perceptron (MLP) [12]. The proposed multilayer perception network is a simple single input, single output based on a single hidden layer and is used for training, testing, and prediction of the vehicle trajectories.

Most of the prior research based on Neural Network or DBN (Dynamic Bayesian Network) have recognized temporal information, inferred current states, or simply detected the action after it has begun, but it has not predicted future states. Most studies until now have analyzed driver behavior by offline processing of data due to the limited datasets. Besides, some studies are only able to train itself and predict the future positions of a lane-changing vehicle in certain discrete sections of the path only and not over the complete lane change path. Therefore, our research is intended to real-time predict lane-changing trajectory based on a time delay Back-Propagation Neural Network (BPNN). We proposed two-layer Tansig and Linear BP neural network to deal with n -input, single-output problems. The position, velocity, acceleration, and time headway of the vehicle were used as the inputs of the model. The future lane-changing trajectory was considered as the desired output of the trained network. We mixed the data of the different path sections as inputs to train the network in order to make its applications more wide. To validate the model, we employed NGSIM trajectory data and a smoothing method. Simulation results demonstrate the effectiveness of the proposed model.

In this research, we attempted to answer the following questions based on the proposed model. Is it possible to infer future driving states in different path sections using one neural network? How long of the prediction time is stable? Is it necessary to combine various input data for the prediction, and which one (or combination) is the more informative indicator? How to post-processed the raw data effectively? Which network is suitable for online analysis, BP Neural Network or Elman Network?

2. Traditional Lane-Changing Model Development

A major problem in developing a driver model lies on the requirement of suitable quantitative formulations by means of mathematical and statistical theories [13–15].

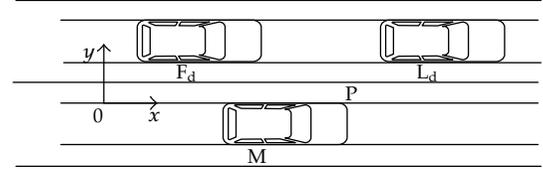


FIGURE 1: Prelane-changing configuration showing position of subject vehicle M.

Lane-changing behavior has been widely studied from the viewpoint of traffic flow theory. A typical lane-changing algorithm which is used for microscopic traffic simulation is described in this section [16].

Figure 1 illustrates the lane-changing situation for vehicle M. The vehicles L_d , F_d , and M represent the leading vehicle in the destination lane, the following vehicle in the destination lane, and the subject vehicle, respectively. The trajectories of subject vehicle M can be constructed by a two-dimensional coordinate system.

The objective of this section is to use a simple lane-changing model to define the lateral acceleration, the lateral velocity, and the lateral distance of subject vehicle M during a certain time-interval $[t_0, T]$ in Figure 2. The time-intervals for the lane-changing maneuver are shown in Figure 2 and the times $t_0 - T$ are defined as follows.

- (i) At time t_0 , the subject vehicle M starts lane-changing maneuver and sets $t_0 = 0$.
- (ii) At time t_{adj} , the subject vehicle M adjusts successfully for lateral acceleration.
- (iii) At time $t_C + t_{adj}$, the subject vehicle M arrives at the marginal collision point.
- (iv) At time $t_{lat} + t_{adj}$, the subject vehicle M finishes the lateral acceleration.
- (v) At time T , the subject vehicle M finishes the lane-changing maneuver.

Note that with the exception of the subject vehicle, the lateral and longitudinal accelerations of the other vehicles are assumed to be 0. Figure 3 illustrates the motion of vehicle M during a lane-changing or merge maneuver. $\theta(t)$ is the angle between the tangent of the lane changing trajectory at the time t and the X-axis. h is the lane width. H is the total lateral displacement for vehicle M. Memar et al. developed a sinusoidal pattern of lateral acceleration of the subject vehicle [17]. Instantaneous lateral acceleration $a_{lat}(t)$ is given by (1). Where H is the total lateral displacement for the subject vehicle M, t is the elapsed time, and t_{lat} and t_{adj} are defined as above. According to (1), the lateral acceleration is positive within the first half of the lateral displacement, that is, $H/2$, and negative in the second half.

On basis of the lateral acceleration, the lateral velocity and the lateral distance traveled of the front-left corner P during a lane-changing can be derived by successive integration, given in (2) and (3).

As indicated above, traditional lane-changing models do not consider the uncertainties and perceptions in human

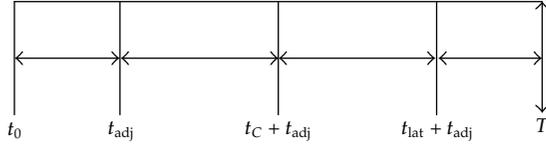


FIGURE 2: Time-intervals of lane-changing maneuver.

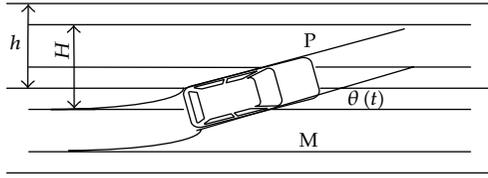


FIGURE 3: The vehicle M's motion during a lane-changing/merge maneuver.

behavior that are involved in modeling lane-changing. Therefore, research that develops lane-changing model from the artificial intelligence viewpoint is very important, such as the artificial neural network method introduced in the next section. The random term in network is attributable to a specific individual and the unimportant variables

$$a_{\text{lat}}(t) = \begin{cases} 2\pi H/t_{\text{lat}}^2 \times \sin\left(\frac{2\pi}{t_{\text{lat}}}(t - t_{\text{adj}})\right), & t_{\text{adj}} \leq t \leq t_{\text{lat}} + t_{\text{adj}}, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

$$v_{\text{lat}}(t) = \begin{cases} -H/t_{\text{lat}} \times \cos\left(\frac{2\pi}{t_{\text{lat}}}(t - t_{\text{adj}})\right) + \frac{H}{t_{\text{lat}}}, & t_{\text{adj}} \leq t \leq t_{\text{lat}} + t_{\text{adj}}, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

$$y_{\text{lat}}(t) = \begin{cases} H, & t \geq t_{\text{lat}} + t_{\text{adj}}, \\ -H/2\pi \times \sin\left(\frac{2\pi}{t_{\text{lat}}}(t - t_{\text{adj}})\right) + \frac{H}{t_{\text{lat}}}(t - t_{\text{adj}}), & t_{\text{adj}} \leq t \leq t_{\text{lat}} + t_{\text{adj}}, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

3. Neural Network Model for Lane-Changing Trajectory Prediction

It is necessary that Advanced Driver Assistance System (ADAS) is provided with concepts and techniques that enable prediction of future situations. In order to optimize warning and control strategy of driver assistance systems, determine assistance systems more exactly, an analysis of driving behavior prediction under real-time traffic conditions is very essential to carry out driving erroneous analysis at the microscopic level. As an example, early recognition of a lane-changing behavior would help to adapt the warning and control strategy of Forward Collision Avoidance Assistance Systems and Lane Departure Warning (LDW) Systems.

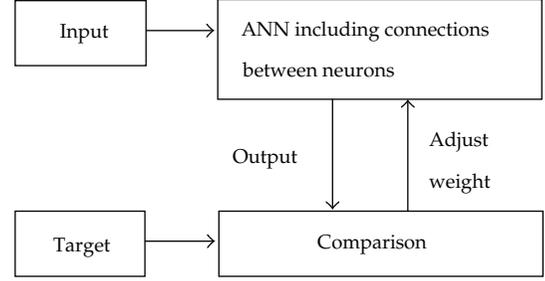


FIGURE 4: Conceptual operation of ANN model.

Therefore, our research objective was the development of a neural network-based model for collision avoidance systems in the case of lateral lane-changing and longitudinal car following. Consistent with this primary goal, different advanced neural network models were employed and compared for the same problem.

3.1. Artificial Neural Network. Artificial Neural Networks (ANNs) are massive parallel adaptive networks of simple nonlinear computing elements. These elements are called neurons and are intended to model some functionalities of the human nervous system in order to take advantage of its computational strength [18]. As a commonly used nonlinear function approximation tool, artificial neural network has shown great advantages in forecasting, pattern identification, optimization techniques, and signal processing for its nonlinear, flexible, and valid self-organization properties. A variety of problem areas are modeled using ANN [19–21] and in many instances, ANN has provided superior results compared to the conventional modeling techniques.

The basic model of ANN consists of computational units, which are a highly simplified model of the structure of the biological neural network [22]. Conceptual operation of ANN is shown in Figure 4. ANN is regarded as a black box that takes a weighted sum of all inputs and computes an output value using a transformation or output function. The output value is propagated to many other units via connections between units.

In general, the output function is a linear function in which a unit becomes active only when its net input exceeds the threshold of the unit, or it is a sigmoid function which is a nondecreasing and differentiable function of the input. Computational units in an ANN model are hierarchically structured in layers and depend upon the layer in which a unit resides. The units are called input, hidden, or output units. There are many input units, and some are dependent on the others. The output units are dependent on all input units. A hidden unit is used to augment the input data in order to support any required function from input or output. The inputs and outputs can be discrete or continuous data values. The input and output could also be stochastic, deterministic, or fuzzy.

In order to store a pattern in a network, it is necessary to adjust the weights of the connections in the network. The set of all weights on all connections in a network form a weight vector. The process of computing appropriate weights is called

a learning law or learning algorithm. The learning process of ANN can be thought of as a reward and punishment mechanism [23], whereby when the system reacts appropriately to an input, the related weights are strengthened. In this case, it is possible to generate outputs, which are similar to those corresponding to the previously encountered inputs. On the contrary, when undesirable outputs are produced, the related weights are reduced. The model learns to give a different reaction when similar inputs occur, thus updating the system towards producing desirable results, whilst the undesirable ones are “punished.”

Back-Propagation (BP) neural network, a typical case of neural networks, is used most widely and is more mature than other networks. BP neural network models consist of an input layer, one or several hidden layers, and an output layer. The typical BP neuron model is shown in Figure 5. When a set of input values and corresponding desired output values are supplied to the network, the transferred value is propagated from the input layer through hidden layers to the output layer. The neural network tries to learn the input-output parameter relationship process by adapting its free parameters.

Mathematical expression of a BP neural network is defined as

$$Y_j = f\left(\sum_{i=1}^n w_{ji}x_i - b_j\right) = f(n_j), \quad (4)$$

where column vector \mathbf{X} is input vector, row vector \mathbf{W}_j is the connection weight vector for neuron j , b_j is the threshold of the output, n_j represents the input of neurons, and the function f is the transfer function.

3.2. BP Neural Network Model for Lane-Changing Trajectory Prediction. For safe driving, it is necessary that the drivers perceive the relevant objects of a situation, comprehend the meaning of these objects to form a holistic understanding of the current situation, and predict the future development of the situation [24]. However, it is difficult for traditional lane-changing behavior model to embody the uncertainty in the series of cognitive behavior of drivers. Unlike the classic mathematic methods, BP networks can approximate the specific inputs and outputs relationship without a certain model.

Our research has attempted to set up a BP neural network model for lane-changing trajectory prediction and approximate the simulation of vehicle-to-vehicle interactions during the lane-changing process. Figure 6 shows a simple lane-changing maneuver.

In this study, an n -input, single-output based time delay BP neural network with two hidden layers is used for training, testing, and prediction of the vehicle trajectories. There are 40 sets of training samples consisted of four variables: the prior position, velocity, acceleration of the object vehicle, and the time headway. Each input variable consists of 1 second contiguous time history sets (or 10 frames). The desired output is the prediction of the next state 1 second in the lane-changing process by delaying time. The network was trained based on the Levenberg-Marquardt algorithm and the mean-squared error was examined. Network weight values are then

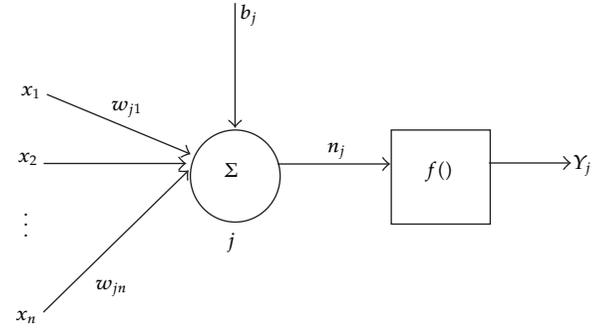


FIGURE 5: Typical BP neuron model.

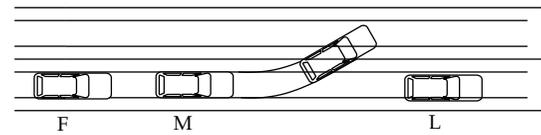


FIGURE 6: The vehicles M, F, and L during a lane-changing maneuver.

iteratively adjusted until output errors are minimized. Once appropriate training data were collected, learning procedure could be implemented in order to reach similar performance with neural networks. Note that we mixed the training data of the different path sections as inputs to make the applications more wide.

4. Simulation Results and Discussion

4.1. Date Collection Based on Driving Simulation Test. In the simulation test, a driving simulator was used to collect driving behavior data including vehicle movements and maneuver operations. Separate computers were used to generate vehicle motion calculations and the front view displays. Subjects were instructed to use the driving simulator executing lane-changing and overtaking maneuvers on a two-lane road. There were 32 male drivers and 8 female drivers, their ages range from 24 to 50 years old and driving experience range from 1 year to 23 years. All the training data for the BP neural network was obtained from a driving simulator.

4.2. Results and Discussion of BP Neural Network Model. The BP neural network model used to predict lane-changing trajectory is shown in Figure 7. The network model consists of an input layer, two hidden layers, and an output layer. A nonlinear sigmoid defines each neuron's activation function in the first layer. The second layer neurons are linear. The biases b is attributable to a specific individual, uncertainties, and the unimportant variables.

In this paper, we only show tested results of lane-changing lateral trajectories as an example. Figure 8 shows the tested lateral trajectory from a driving simulator and the predicted lateral trajectory from BP neural network model. The prediction of the vehicle trajectory has large error in the initial phase because of the small number of previous samples.

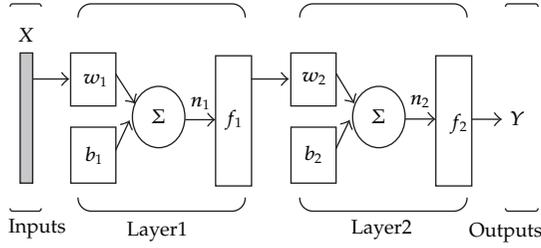


FIGURE 7: BP neural network with two hidden layers.

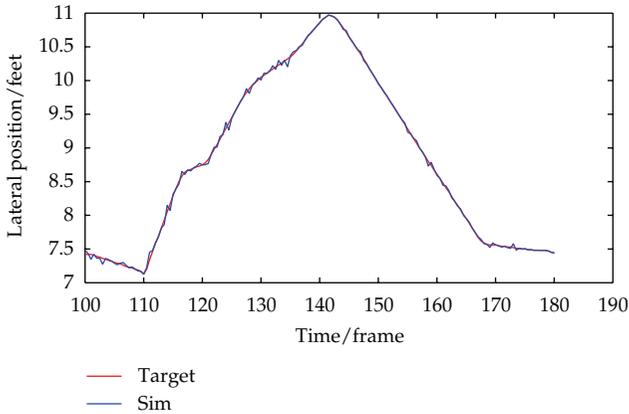


FIGURE 8: Predicted lateral trajectory of the vehicle based on BPNN with 25 hidden nodes.

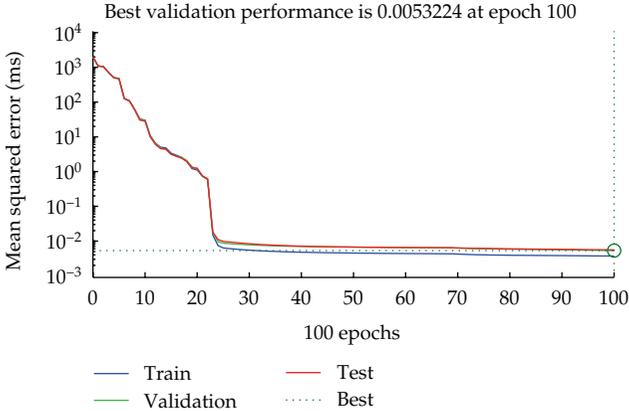


FIGURE 9: Performance curve of BPNN.

Performance curve of BPNN is shown in Figure 9. As indicated in Figure 9, by increasing the number of iterations, the performance of the network will be improved. But when the number of iterations is large enough, an increase in the number of iterations will no longer reduce the error rate. Besides, we also found vehicle lateral trajectory is influenced by time headway and the prior lateral trajectory more obviously. It is reasonable, because most of lane-changing maneuvers are caused by a slow leading vehicle.

4.3. Model Validation Based on NGSIM Data. The Next Generation Simulation (NGSIM) data is used to test and verify our model. NGSIM data provides detailed vehicle trajectory data, wide-area detector data, and supporting data needed

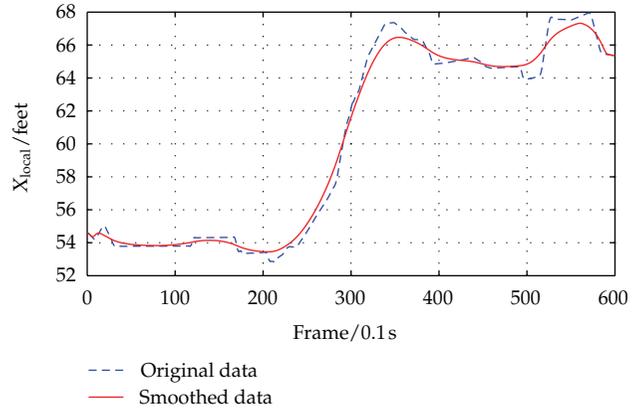


FIGURE 10: Smoothed lateral position.

for behavioral algorithm research. Considering observations in real traffic are always affected by measurement errors, we smooth the raw data which is used to test the model.

Taking the lateral position and vehicle velocity, for example, the smooth results are shown in Figures 10 and 11. The test results of different prediction time are shown in Figure 12. The green curve is lateral position with 2 s prediction time (MSE = 0.2273). The blue curve is lateral position with 1 s prediction time (MSE = 0.0184), which is more accurate.

The validation results demonstrate the effectiveness of BNs model around a prediction time of 1 s. And validations using a new sample of NGSIM also prove that our model can predict the future positions of a lane-changing vehicle under different path sections.

4.4. Comparison between BPNN and Elman Network. For the purpose of comparison, training time and accuracy are considered between the BPNN model and the Elman Network model. The training time is the time needed to train a neural network. The accuracy is measured by calculating the error for the testing data points. Elman Network differs from BPNN in that the first layer has a recurrent connection, shown in Figure 13. We employed the same transfer function, training algorithm, neural nodes, and training samples.

In Figures 14, 15, and 16, the comparison results with 1 s prediction time show that convergence rate is increased and the training time is reduced with the BPNN model. But the accuracy is increased with the Elman Network. The mean-squared error based on BPNN model is 0.0371, whereas mean-squared error based on Elman Network model is 0.0279.

Simulation results demonstrate that in this application BPNN is more advantageous to the training time and accuracy because of the simpler network structure. For the future research on real-time on-line prediction, both the training time and accuracy are important.

5. Conclusion

The main motivation for using a neural network is its ability to learn and incorporate the uncertainties from real driving data. This means, after learning from the driving behavior

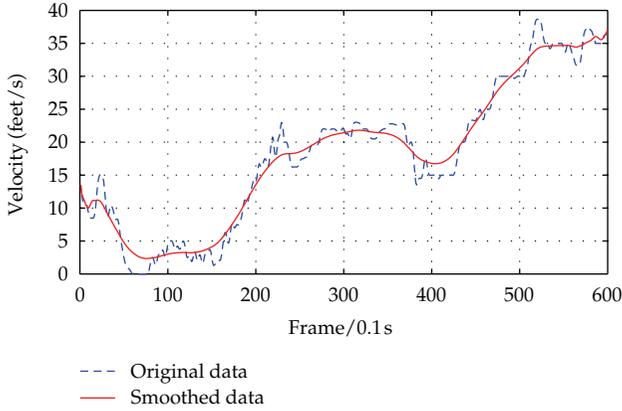


FIGURE 11: Smoothed vehicle velocity.

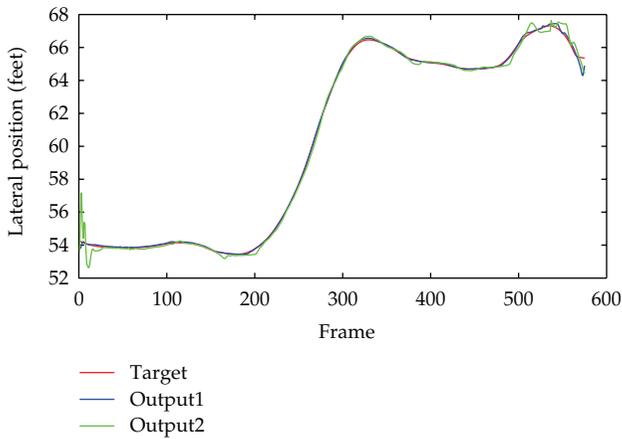


FIGURE 12: Predicted lateral trajectory of the vehicle with 1 s and 2 s prediction time.

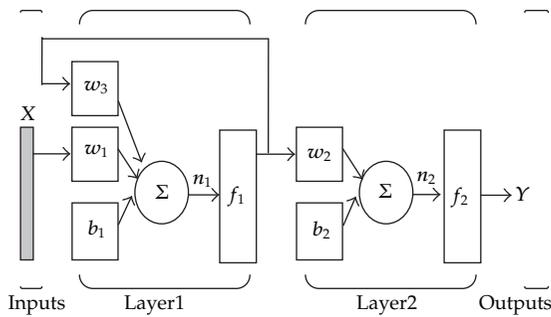


FIGURE 13: Elman Network with two hidden layers.

data, neural network could generate vehicles states to reproduce any style of driving. In this study, prediction was done by sequential inference through BPNN and Elman models using the collected driving data from a driving simulator and NGSIM. The BP neural network model did a better job predicting lane-changing trajectories under different path sections and generated reliable simulation results. Note that among the various data, the vehicle trajectory is influenced by time headway and the prior position more obviously.

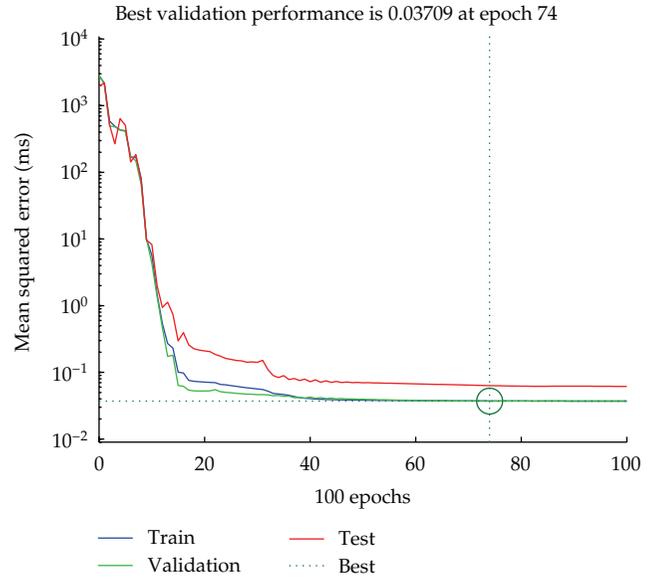


FIGURE 14: Performance curve of BPNN.

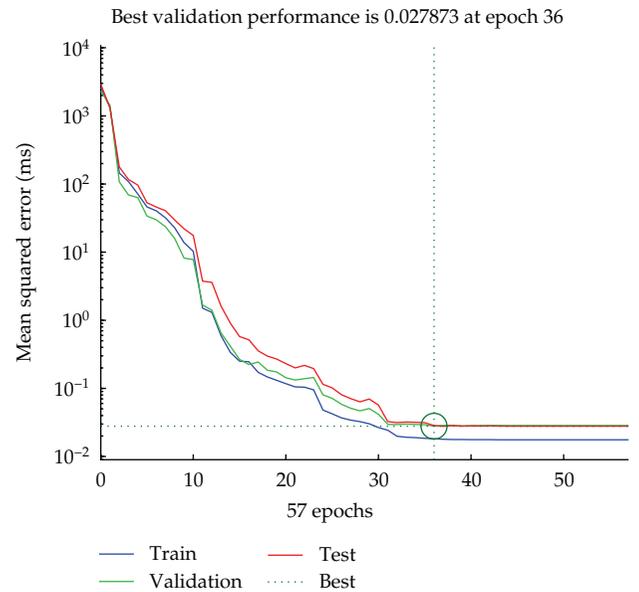


FIGURE 15: Performance curve of Elman Network.

We can use this network model as a basic model or an initial point to create more complex lane-changing models. For future research, perhaps the results of the simulation could be improved by using more inputs, such as the current states of the vehicles and the driver's reaction time. Reaction time could be increased when the driver is under stress or distracted. Therefore, future research should also consider how to incorporate mental workload in the driver behavior model and how to increase the length of the prediction window. One possible method to incorporate the mental workload is to measure the mental workload and define it as a part of the input to the neural network model. Meanwhile, for the future research, we attempt to develop a real-time,

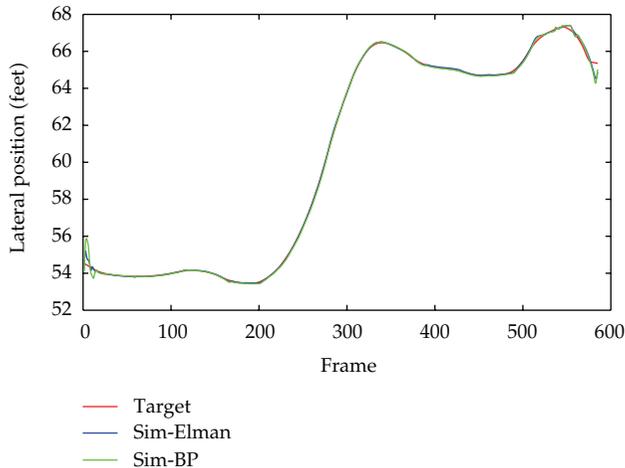


FIGURE 16: Predicted lateral trajectory of the vehicle based on BPNN and Elman Network with 1 s prediction time.

on-road lane-change detector that can anticipate the future driving state. This detector extracts signals from vehicle sensors and preprocesses them into feature vectors, which are then used for offline training and online inference. Finally, the simulation results could be used to assess the safety of lane-changing maneuvers and lay a necessary foundation for further development of the autonomous lane-changing and overtaking assistance systems.

Acknowledgments

This research was supported in part by National Nature Science Foundation of China under Grant 51010305084, 51110305060, 51210305046, and the Program of Introducing Talents of Discipline to Universities under Grant B12022. The authors also gratefully acknowledge the helpful comments and suggestions of the colleagues from Transportation Research Center of Beijing University of Technology, which have improved the paper.

References

- [1] W. H. Wang, Q. Cao, K. Ikeuchi, and H. Bubb, "Reliability and safety analysis methodology for identification of drivers' erroneous actions," *International Journal of Automotive Technology*, vol. 11, no. 6, pp. 873–881, 2010.
- [2] P. Varaiya, "Smart cars on smart roads: problems of control," *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 195–207, 1993.
- [3] M. Peden, R. Scurfield, D. Sleet et al., *World Report on Road Traffic Injury Prevention*, World Health Organization, Geneva, Switzerland, 2004.
- [4] S. E. Lee, E. C. B. Olsen, and W. W. Wienville, *A Comprehensive Examination of Naturalistic Lane-Changes*, Virginia Tech Transportation Institute I National Highway Transportation Safety Administration, Blacksburg, Va, USA, 2004.
- [5] T. Kumagai, Y. Sakaguchi, M. Okuwa et al., "Prediction of driving behavior through probabilistic inference," in *Proceedings of the International conference on Engineering Application of Neural Networks (EANN '03)*, pp. 117–123, Malaga, Spain.
- [6] S. Tezuka, H. Soma, and K. Tanifuji, "A study of driver behavior inference model at time of lane change using Bayesian Networks," in *Proceedings of the IEEE International Conference on Industrial Technology (ICIT '06)*, pp. 2308–2313, Mumbai, India, December 2006.
- [7] N. Kuge, T. Yamamura, and O. Shimoyama, *A Driver Behavior Recognition Method Based on a Driver Model framework*, SAE, Warrendale, Pa, USA, 1998.
- [8] A. Sathyanarayana, P. Boyraz, and J. H. L. Hansen, "Driver behavior analysis and route recognition by hidden Markov models," in *Proceedings of the IEEE International Conference on Vehicular Electronics and Safety (ICVES '08)*, pp. 276–281, Columbus, Ohio, USA, September 2008.
- [9] A. Pentland and L. Andrew, "Modeling and prediction of human behavior," *Neural Computation*, vol. 11, no. 1, pp. 229–242, 1999.
- [10] C. C. Macadam and G. E. Johnson, "Application of elementary neural networks and preview sensors for representing driver steering control behaviour," *Vehicle System Dynamics*, vol. 25, no. 1, pp. 3–30, 1996.
- [11] R. M. H. Cheng, J. W. Xiao, and S. LeQuoc, "Neuromorphic controller for AGV steering," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2057–2062, May 1992.
- [12] R. S. Tomar, S. Verma, and G. S. Tomar, "Prediction of lane change trajectories through neural network," in *Proceedings of the International Conference on Computational Intelligence and Communication Networks (CICN '10)*, pp. 249–253, Bhopal, India, November 2010.
- [13] W. Wang, H. Guo, H. Bubb, and K. Ikeuchi, "Numerical simulation and analysis procedure for model-based digital driving dependability in intelligent transport system," *KSCE Journal of Civil Engineering*, vol. 15, no. 5, pp. 891–898, 2011.
- [14] W. Wang, X. Jiang, S. Xia, and Q. Cao, "Incident tree model and incident tree analysis method for quantified risk assessment: an in-depth accident study in traffic operation," *Safety Science*, vol. 48, no. 10, pp. 1248–1262, 2010.
- [15] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research Part C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [16] W. H. Wang, C. X. Ding, G. D. Feng, and X. B. Jiang, "Simulation modelling of longitudinal safety spacing in inter-vehicles dynamics interactions," *Journal of Beijing Institute of Technology*, vol. 19, supplement 2, pp. 55–60, 2010.
- [17] A. H. H. A. Memar, P. Z. H. Bagher, and M. Keshmiri, "Mechanical design and motion planning of a modular reconfigurable robot," in *Proceedings of 11th International of Conference on Climbing and Walking Robots and the Support Technologies for Mo*, pp. 1090–1097, 2008.
- [18] T. Kohonen, "An introduction to neural computing," *Neural Networks*, vol. 1, no. 1, pp. 3–16, 1988.
- [19] E. G. Tsionas, P. G. Michaelides, and A. T. Vouldis, "Global approximations to cost and production functions using artificial neural networks," *International Journal of Computational Intelligence Systems*, vol. 2, no. 2, pp. 132–139, 2009.
- [20] E. Zio, "Neural networks simulation of the transport of contaminants in groundwater," *International Journal of Computational Intelligence Systems*, vol. 2, no. 3, pp. 267–276, 2009.

- [21] Gowrishankar and P. S. Satyanarayana, "Neural network based traffic prediction for wireless data networks," *International Journal of Computational Intelligence Systems*, vol. 1, no. 4, pp. 379–389, 2008.
- [22] B. Yegnanarayana, *Artificial Neural Networks*, PHI, New Delhi, India, 2005.
- [23] T. Kaya, E. Aktaş, I. Topçu, and B. Ülengin, "Modeling toothpaste brand choice: an empirical comparison of artificial neural networks and multinomial probit model," *International Journal of Computational Intelligence Systems*, vol. 3, no. 5, pp. 674–687, 2010.
- [24] M. R. K. Baumann, D. Rosier, and J. F. Krems, "Situation awareness and secondary task performance while driving," in *Proceedings of the 7th International Conference on Engineering Psychology and Cognitive Ergonomics (EPCE '07)*, vol. 4562 of *Lecture Notes in Computer Science*, pp. 256–263, Beijing, China, 2007.

Research Article

Dynamic Cognitive Self-Organized TDMA for Medium Access Control in Real-Time Vehicle to Vehicle Communications

Mario Manzano,¹ Felipe Espinosa,¹ Ángel M. Bravo-Santos,² Enrique Santiso,¹ Ignacio Bravo,¹ and David Garcia¹

¹ Electronics Department, Polytechnic School, University of Alcala, Campus Universitario s/n, 28871 Alcala de Henares, Spain

² Signal and Communications Theory Department, Higher Polytechnic School, Carlos III University, Avenida de la Universidad 30, 28911 Leganes, Spain

Correspondence should be addressed to Felipe Espinosa; espinosa@depeca.uah.es

Received 21 September 2012; Accepted 23 December 2012

Academic Editor: Wuhong Wang

Copyright © 2013 Mario Manzano 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 emergence of intelligent transport systems has brought out a new set of requirements on wireless communication. To cope with these requirements, several proposals are currently under discussion. In this highly mobile environment, the design of a prompt, efficient, flexible, and reliable medium access control, able to cover the specific constraints of the named real-time communications applications, is still unsolved. This paper presents the original proposal integrating Non-Cooperative Cognitive Time Division Multiple Access (NCC-TDMA) based on Cognitive Radio (CR) techniques to obtain a mechanism which complies with the requirements of real-time communications. Though the proposed MAC uses a slotted channel, it can be adapted to operate on the physical layer of different standards. The authors' analysis considers the IEEE WAVE and 802.11p as the standards of reference. The mechanism also offers other advantages, such as avoiding signalling and the adaptation capacity to channel conditions and interferences. The solution is applied to the problem of units merging a convoy. Comparison results between NCC-TDMA and Slotted-Aloha are included.

1. Introduction

The massive use of transportation vehicles in our society, and in particular of road transportation, has led to problems of congestion, pollution, and safety [1]. However, current advances in technology are steering the future of transport towards the intelligent transportation systems (ITS) and innovative cooperative transport alternatives [2, 3].

The first attempt to standardize the techniques related to ITS was the Digital Short Range Communication (DSRC) [4], developed in the United States in order to meet these communication needs. In the recent years, several organizations have led different programs and standardization efforts to cope with the specific requirements of the ITS [5]. Some of them are the industry consortia C2C-CC (Car to Car Communication Consortium) [6], the IEEE with the protocol suite WAVE (Wireless Access for the Vehicular Environment) [7] or the ISO (International Organization for Standardization) with

CALM (Continuous Air-Interface Long and Medium range), and the CVISs project (Cooperative Vehicle Infrastructure Systems) [8].

One of the common factors associated with these standardization activities is that the IEEE 802.11p technology is targeted to be the common V2V data link technology used for traffic safety applications [5].

As mentioned above, the IEEE proposal is the new protocol suite WAVE (Wireless Access in Vehicular Environments), which includes the description of the protocol IEEE 802.11p, belonging to the widely used 802.11 family of standards [8]. This protocol was introduced to solve the limitations inherent in nonuniversal standardization and based on the gained experience.

The aim of developing a universal standard such as WAVE, and therefore 802.11p, is to produce a standard which is generically applicable to the greatest possible number of intelligent transportation systems. Thus, the final system

design must incorporate specific improvements to the protocol to render the latter capable of complying with the particular constraints of each application.

The 802.11p specification focuses on the physical layer and on the medium access control (MAC) sublayer, taking as its basis the best features of other specifications of the 802.11 family, widely implemented in wireless access to internet services. The medium access control protocol contains some amendments to previous specifications, such as the use of service and control channels, and the Enhanced Distributed Channel Access (EDCA) protocol [9]. Whilst these modifications improve the standar performance, they do not guarantee a total compliance with the communication requirements of certain ITS applications [10].

In order to overcome these above-mentioned limitations of the IEEE 802.11p with regard to critical ITS applications, this paper presents a medium access control mechanism (NCC-TDMA) based on a self-organized TDMA scheme which does not require signalling. To achieve this, Cognitive Radio techniques are implemented.

The rest of the paper is organized as follows. In Section 2 the limitations of 802.11p are explained in detail; Section 3 includes a summary of other proposals found in literature; Section 4 is devoted to the formal approach to the explanation of the problem and Section 5 contains the theoretical description of the solution. Finally, Section 6 provides some simulation results and Section 7 is dedicated to discussion and conclusions.

2. Problem Description

The physical layer and MAC sublayer of the WAVE standard, currently being developed by the IEEE with the aim of meeting ITS requirements, are described in the IEEE 802.11p specification [11].

The technology associated with the 802.11 family offers high performance in terms of wireless access to local area network (LAN) environments. This technology provides higher transmission rates than other alternatives such as WiMAX or 3G and is relatively cheap and easy to install and maintain. These factors make the 802.11 family an excellent option for being used as a standard in vehicle environment communications. However, in order to be included in the WAVE standard, certain amendments were necessary [12].

The physical layer of IEEE 802.11p is basically the same as that of IEEE 802.11a, due in part to the fact that both take advantage of the same frequency range (5 GHz). The particular characteristics of vehicular environments, such as high speed nodes or rapidly changing network topology, require the inclusion of certain modifications in order to guarantee support of communication conditions in the field of ITS. IEEE 802.11p is based on applying Orthogonal Frequency Division Multiplexing (OFDM) with 10 MHz channels, in contrast to IEEE 802.11a which uses 20 MHz channels. This implies that the possible transfer rates are reduced by half, operating at 3, 4.5, 6, 9, 12, 18, 24, and 27 Mbps. In addition, greater constraints have been incorporated into the receiver design to mitigate the problem of adjacent channel interference [8].

As regards the medium access control sublayer, the key proposal is to simplify the connection and information exchange operations in local area wireless networks, providing effective use of adhoc communication between vehicles [8]. To this end, the infrastructure mode has been eliminated and the nodes exchange information by establishing networks in a new mode known as WAVE mode.

Using the WAVE mode eliminates the need to carry out prior network authentication and connection procedures, since in WAVE these functionalities are implemented by the upper layers of the system, as shown in Figure 1. This reduces the time and signalling required for a node to send information to network members.

On the other hand, as with the rest of the IEEE 802.11 family, the IEEE 802.11p uses Carrier Sensing Multiple Access/Collision Avoidance (CSMA/CA) as the medium access control mechanism. The only proposed improvement is the inclusion of Enhanced Distributed Channel Access (EDCA), already defined in the IEEE 802.11e amendment, whose aim is to provide the management of the quality of service. In this way the MAC mechanism of IEEE 802.11p allows to classify traffic according to the four different priorities defined in EDCA [13].

The resulting medium access control mechanism is therefore contention-based rather than deterministic, since it cannot guarantee medium access within a bounded time [10]. The improvement provided by EDCA does not guarantee deterministic access delay. As indicated in [13, 14], EDCA performance depends on the level of priority assigned to messages exchanged within the system, and the existence of a large number of high priority messages will generate an effectiveness reduction.

2.1. WAVE Multichannel Operation. Following the guidelines established from DSRC, WAVE standard includes a multichannel operation based on seven different channels. According to [15], six of them are called service channels (SCHs) and the seventh is the control channel (CCH), see Table 1. Although the EDCA strategy is applied to each of these channels, the control one is responsible for critical messages, or in terms of ITS, real-time communications. Moreover, for the control channel, the possibility of using a higher transmission power is provided in order to improve the likelihood of successfully carrying out a transmission on that channel.

In summary, the ability of WAVE to support real-time communications can be measured in the operation of the protocol on the control channel. Some of these applications are mentioned in the next section.

2.2. IEEE 802.11p Limitations. A fundamental requirement of real-time communication in the field of ITS is that the information contained in a transmission must be delivered at the appropriate time. In other words, a message must reach the receiver within a certain time and with a specified level of reliability (probability of error). Communication messages in real-time do not necessarily require high transmission rate or low delay, but it does require predictable behaviour, so that the message is received within the time limit and with

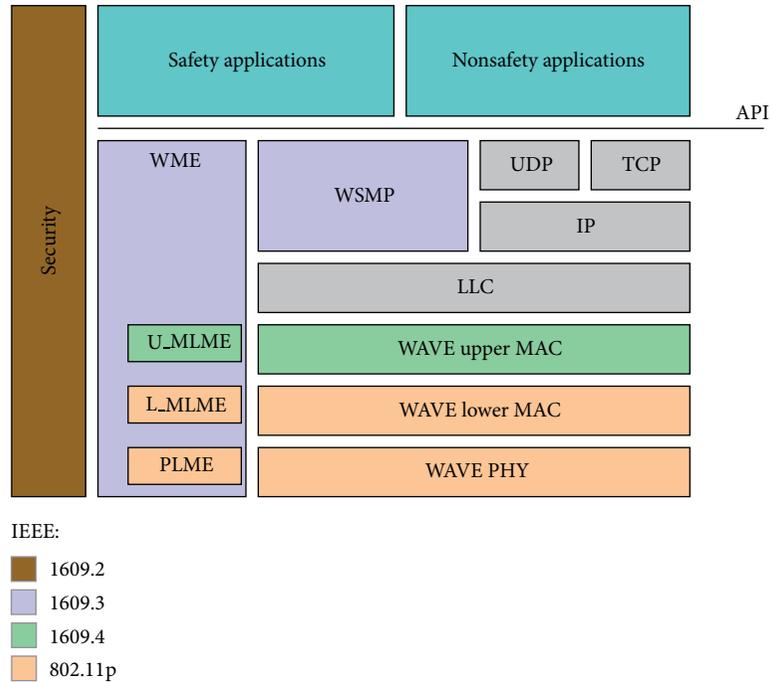


FIGURE 1: WAVE protocol stack.

the specified probability of error. Thus, real-time communication tasks are characterized by two important parameters: delivery time and reliability [10].

Cooperative guiding of autonomous vehicles is a clear real-time ITS application. A particular case of it is the guidance of transport units in convoy, where messages containing motion state must be delivered before global system stability is compromised [2]. This is the closest case to the system model analysed in this work.

Safety message access delay is defined as the average delay between the time at which the packet is generated and the time at which it is successfully received. According to [16], safety message access delay should be less than 200 milliseconds to allow proper driver reaction time to traffic warning signals.

The medium access control mechanism included in IEEE 802.11p is a contention-based mechanism capable of providing quality of service, but not in a deterministic way. Thus, it cannot guarantee delivery of information within an established time limit.

From the foregoing, it can be concluded that the improvement of the medium access control mechanism is required to support the special communication requirements of these ITS safety and control applications.

3. Related Work

Having identified the limitations of the IEEE 802.11p medium access control mechanism, in this section we will review the alternative solutions which have been proposed in the literature.

The most suitable medium access control mechanisms for implementing the specific functions required in ITS can be classified as follows [17]

- (i) Contention-based protocols: the main advantage of these protocols lies in the fact that the typical network reconfiguration present in vehicular environments or Vehicular Ad Hoc Networks (VANET) does not require reconfiguration in the protocol. However, due to the random nature of medium access produced by these kinds of mechanisms, access to the medium cannot be guaranteed within a bounded time limit. These are nondeterministic mechanisms, and examples include CSMA, CSMA/CA, Aloha, and Slotted-Aloha.
- (ii) Schedule-based or contention-free protocols: these mechanisms can provide bounded access delay, even in saturated environments, but they require reconfiguration during system operation in order to adapt to the ever-changing network topology. Examples of this are the token passing protocols (e.g., Token Ring), Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), and Code Division Multiple Access (CDMA).

One of the options included in this second type protocol is TDMA. In this scheme, the shared resource is divided into time intervals which are allocated to different users for transmission, thus avoiding transmission collisions.

In a dynamic TDMA scheme, in which network topology or members are variable, schedules for access to transmission time slots are calculated continuously to reflect network

TABLE 1: Multichannel operation in WAVE.

	Frequency [GHz]						
	5.855	5.865	5.875	5.885	5.895	5.905	5.915
Channel number	172	174	176	178	180	182	184
Channel type	Service channel	Service channel	Service channel	Control channel	Service channel	Service channel	Service channel
Application	Nonsafety	Nonsafety	Traffic efficiency	Critical safety	Critical safety	Traffic efficiency	Traffic efficiency
Radio range	C2C	Medium	Medium	All	Short	Short	Intersections
Tx power level	33 dBm	33 dBm	33 dBm	44.8 dBm	23 dBm	23 dBm	40 dBm

TABLE 2: Qualitative comparison of the analyzed protocols.

Type	CSMA/CA Contention	Slotted-Aloha Contention	RR-Aloha Scheduled	STDMA Scheduled	VeSOMAC Scheduled	NCC-TDMA Scheduled
Topology-independent	Yes	Yes	Yes	No	No	No
Signalling	No	No	Yes	Yes	Yes	No
Message exchange required	No	No	Yes	Yes	Yes	No
Possibility of collision	Yes	Yes	Yes	Yes	No	No
Time synchronization required	No	Yes	Yes	Yes	No	Yes
Reason for access schedule reconfiguration	N/A	Collision	New users, new transmissions.	Periodic reconfiguration	Collision with other system users	Detection of channel occupation, whether by another user or due to outside signal interference
Oriented at application	No	No	No	Yes: vehicle localization and movement	Yes: vehicle localization	No

changes. This schedule must be transmitted to members so that each one knows their allocated time slot. Its implementation involves time costs and requires signalling between system users in order to transmit the global schedule. Furthermore, in a centralized scheme, calculating time slot allocation schedules is carried out by a single central manager, responsible for establishing appropriate allocation within the system and transmitting this information to all the system users.

Another alternative, which is somewhat more complicated but also more suited to the special characteristics of vehicular environments, is the self-organized dynamic TDMA. In this case, there is no central unit responsible for calculating the slot allocation for the users, since allocation becomes the responsibility of the system users themselves. Consequently, the medium access control mechanism is able to contend with high levels of user mobility and network topology variations inherent in ITS.

Table 2 gives a summarized comparison of the main previous alternatives: CSMA/CA protocol included in IEEE 802.11p [7, 11], Slotted-Aloha proposals [18], RR-ALOHA [19, 20], Self-organizing Time division multiple access (STDMA [10]), and Vehicular Self Organising MAC protocol (VeSO-MAC [21]), together with the basic characteristics of the authors' proposal Non-Cooperative Cognitive Time Division Multiple Access (NCC-TDMA).

The mechanism incorporated into IEEE 802.11p, together with the EDCA modification enabling the provision of quality of service, is a nondeterministic mechanism. CSMA/CA

performance does not depend on network topology, since it does not need to be reconfigured when the user number or its location changes. It is a totally application-independent protocol and it does not require the use of signalling or time synchronization. The main disadvantage is that, since it does not avoid the possibility of collision, maximum delay is not bounded and situations could arise where node delivery continues indefinitely.

The Slotted-Aloha [18] is a contention-based protocol, although it also employs concepts from TDMA schemes. However, in the case of collision, selection of transmission times is based on a random process which means that maximum delay is not bounded for all cases. On the other hand, as with CSMA/CA, this mechanism does not require neither signalling nor reconfiguration when topology changes. This mechanism has been selected for comparison with NCC-TDMA in Section 6.

One protocol capable of achieving dynamic channel sharing is the well-known Reservation Aloha (R-Aloha). In R-Aloha [22] "trial and error" transmission is used to access an available slot. If the transmission is recognized as successful then the slots reserved for that terminal in subsequent frames cannot be accessed by other terminals until the channel is released. The correct operation of R-Aloha requires a central repeater through which the terminals receive all the transmitted signals and, more important, get the same slot status information, for example, busy, free, or collided.

In order to implement a Dynamic TDMA in a distributed way, the protocol named Reliable R-Aloha (RR-Aloha [19])

allows the same R-Aloha procedure to be applied to vehicular communications. In RR-Aloha each terminal periodically transmits the perceived status of the slots in the preceding period (*Frame Information*). This information can be elaborated to get a common slot status, as in R-Aloha [22], and is used by new terminals that reach the network to get access.

RR-Aloha [19] results in a nontopology dependent protocol and it is not oriented to an specific application. On the other hand, it requires continuous transmission of signaling (*Frame Information*) even in a stable situation. It also requires time synchronization.

STDMA [10] is a decentralized scheme where the network members themselves are responsible for sharing the communication channel and, due to the decentralized network topology, the synchronization among the nodes is done through a global navigation satellite system, such as GPS or Galileo. The algorithm is dependent on that all nodes in the network regularly send messages containing information about their own position. The STDMA algorithm will use this position information when choosing slots in the frame. It is a protocol independent of the network topology but it is oriented to vehicle localization and movement warning applications; it requires time synchronization and, above all, channel access schedules are periodically reconfigured, even when a situation of satisfactory balance has been reached.

VeSOMAC [21] is characterized by having bounded maximum delay, at the cost of requiring periodic signalling. As with STDMA [10], it is oriented to applications based on vehicle localization although in contrast, it does not require time synchronization, it eliminates the possibility of collisions, and schedules are only reconfigured when the system detects a conflict between users over the selected slot.

The solution proposed in this paper can be classified among the schedule-based or contention-free protocols. The starting point is a self-organized dynamic TDMA which, thanks to the application of Cognitive Radio techniques (see Section 5.1. Cognitive Radio), does not require signalling messages to be sent between users in order to transmit time slot allocation schedules. This is the main difference between NCC-TDMA and others proposals found in literature.

4. System Model

The proposal presented in this paper has been developed on a system model applicable to the guidance of autonomous vehicles in convoy. In one such application, vehicles are grouped into convoys, each headed by a leader vehicle. These leader vehicles transmit information of the selected route to the rest of the convoy members.

In addition, the control law of each follower demands knowledge of the motion state of the preceding unit. The spread of this information is critical, or in real-time as defined by ITS, as the loss of messages between vehicles could jeopardize the stability of the convoy chain [2]. On the other hand, in the problem of cooperative guidance, there is a possibility that vehicles can leave a convoy and decide to join a different one, which incorporates variability in their sizes.

From a generic point of view, it is a system in which users (vehicles of the convoy) should convey critical information

within a specified maximum deadline. The established communication system is variable with respect to the number of users (due to the ability of vehicles to join or leave each convoy) and with respect to their position (due to the mobility of the ITS vehicles).

The main difference between the work described in this paper and other ITS systems is that the number of users has been chosen within a range of lower values, closer to the convoy guided situations than other common ITS applications. Despite being a specific application in intelligent transport systems, the system model used to develop the proposal presented in this work has been generalized to make the description of the NCC-TDMA mechanism applicable to other ITS examples with real-time communication requirements.

4.1. Scenario. System users are member vehicles of an autonomous convoy that flows freely, ignoring in this paper the special conditions of start, stop, line change, and the generation of reference paths for the convoy on a real map. It is assumed that the vehicles of the transport scenario travel in convoy and they can separate from it (leaving the communication system) in the same way that other vehicles can join it (as new users).

Cooperation in the guidance system of autonomous vehicles in convoy is required in all tasks related to control as well as coordinated tracking of trajectories. However, from the point of view of CR, the system is not cooperative because the messages exchanged among users contain only information of the application, not related to the management of medium access control mechanism. This feature eliminates the need for signalling among users: vehicles never send information about the transmission slot they are using.

4.2. Channel. It is considered that communications among vehicles in a convoy are critical safe communications. Therefore, according to the definition of WAVE channels (see Table 1) only the control channel (CCH) is used.

The main requirement of real-time applications in ITS is that the exchange of messages must be made within a time limit or determined critical time T_c . In the system model, this time defines the duration of the TDMA frame. This frame is divided into different transmission slots whose duration is determined by the needs of the application.

It is assumed that (a) during each transmission slot, it is possible to transmit all the amount of information needed by each user to cover the system requirements, and (b) there are no external interferences on the communication process. Such a case should be treated by the upper layers of the protocol, which is beyond the scope of this work.

4.3. Communication. To perform the planned functions for the NCC-TDMA proposal, it is assumed that all users rely on a communication system compatible with the IEEE 802.11p standard, with the capacity to observe the medium (spectrum sensing), and with an auxiliary device able to provide time synchronization [23].

The nature of the system regarding the vehicles mobility and the possibility that a user join or separate from the

convoy makes it impossible to solve the medium access by implementing a fixed TDMA scheme. It is not possible to allocate a slot for an indefinite period for each vehicle as they can incorporate to or separate from the convoy during the system operation. If this fixed allocation is carried out, it would not allow more incorporations than the originally planned ones, or transmission slots would remain free (without exploitation) when units leave the convoy.

This is the same reason that forces the NCC-TDMA proposal to be a fully decentralized solution where each convoy unit automatically calculates the transmission slot within the TDMA frame.

4.4. Key Aspects. In summary, the model of the communication system under study is characterized by the following restrictions.

- (i) The shared medium is exploited on a TDMA scheme whose frame lasts T_c and is divided into r slots for n users.
- (ii) Each user must transmit its data within the limit T_c . To achieve that is necessary and sufficient the access to the medium during a transmission slot.
- (iii) Once the access to the medium is gained, during the transmission period on the selected slot, it is considered that the transmission is carried out correctly.
- (iv) Vehicles are equipped with a communication system capable of monitoring the environment and providing time synchronization.
- (v) For the mechanism of medium access control, there is no cooperation among users. They neither exchange signaling nor have explicit knowledge of transmission slots used by the rest.

5. NCC-TDMA Proposal

The Non-Cooperative Cognitive Time Division Multiple Access (NCC-TDMA) proposal is aimed at guaranteeing a bounded medium access delay to ITS applications which require real-time communication. Though the proposed MAC uses a slotted channel, it can be adapted to operate on the physical layer of different standards, either by replacing the medium access control mechanism or by complementing it.

This new mechanism of medium access control is based on two of the fundamental concepts of Cognitive Radio: exploitation of the free zones of the shared medium by observing it and learning the external conditions to facilitate adequate adaption to them.

5.1. Cognitive Radio. It was during the introduction and consolidation of a great number of wireless network services and applications that Cognitive Radio (CR) emerged as a promising technology for maximizing the use of the limited radio bandwidth [24].

Cognitive Radio is an intelligent wireless communication system, which is aware of its environment and uses the

methodology of understanding-by-building to learn from the environment and adapt to statistical variations in the input stimuli by effecting the corresponding changes in determined real-time parameters. The twofold aim of CR is to obtain a highly reliable communication whenever and wherever needed and to achieve an efficient utilization of the radio spectrum [25, 26].

To achieve an efficient utilization of the electromagnetic spectrum, one of the main ideas that CR exploits is the observation of the spectrum use. In a CR system, generally primary users are defined with nominal permission to use a particular frequency band, and secondary users are those that take advantage of CR to find the free radiofrequency bands between those occupied by primary users. These free zones are called spectrum holes, and searching for them, as well as using them, constitutes one of the main objectives of CR [27].

To this end, CR systems have receptors able to realize an observation of the medium "spectrum sensing" [23] and to detect its occupation by any kind of signals. The result of this observation of the occupation of the medium leads us to one of the great challenges of CR-based systems: the design of dynamic spectrum access techniques. These techniques, designed in accordance with the level of cooperation among users, range from noncooperative distributed schemes to completely centralized ones.

The NCC-TDMA proposal exploits the concept of consciousness in basing its operation on an observation of the medium occupation. Unlike conventional systems of CR, the observation of the medium does not suppose a sensing of the electromagnetic spectrum searching for free frequency bands (spectrum holes), but on an observation of the occupation of the transmission slots within the TDMA frame defined in the communication system. The other cognitive pillar of the proposal is related to the concepts of learning and adaptability. The idea exploited by the NCC-TDMA mechanism is to adapt a vector regarding the internal state of each user, to the information obtained from the observation of the occupancy of the TDMA frame. This is known as estimated allocation vector or eav , $eav \in R^r$ being r the number of slots.

5.1.1. Observation of the Occupancy of the Medium: Transmission Schemes. A system user, after selecting a slot in the TDMA frame by running the NCC-TDMA algorithm described in next section, can access the medium generating the following situations.

- (A) The selected slot is different from the slot used in the previous iteration. In this case, the user should check before transmitting whether the slot is free or not. Therefore, the total time slot is divided into
 - (i) a first part of observation or sensing of the medium: in which it checks if another user is occupying that time slot and
 - (ii) depending on the outcome of this observation it makes use of the second part of the selected slot as follows.

With the "LISTEN" action: the user listens to receive the information transmitted through the

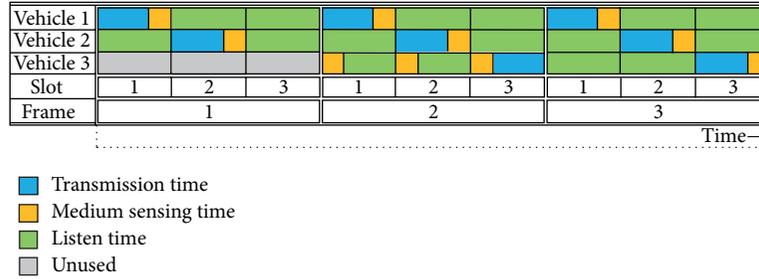


FIGURE 2: Transmission schemes.

channel. This situation leads to the “SENSING + LISTEN” scheme.

With the transmission action “TX”: the critical information of the user is transmitted. This situation corresponds to the “SENSING + TX” scheme.

- (B) The selected slot by the algorithm is the same employed by the user in the previous iteration. Therefore, the user is considered the owner of this slot. In this case the division of the slot is conducted as follows.
 - (i) A primary part of transmission “TX” in which the critical information of the system is transmitted.
 - (ii) A second part of observing the medium “SENSING” in which it is confirmed that there are no interfering transmitters (from outside of the system) occupying the same slot of the frame. This situation results in the “TX + SENSING” scheme.
- (C) Finally, the users access the medium to listen to the transmitted information in the rest of slots of the frame that were not selected by the NCC-TDMA algorithm.

Consequently, a mechanism is established in which the users that would like to transmit in new slots have to firstly check that the channel is free in these slots. The user owning the slot transmits its information in the first part of the slot, for which it will be determined that the slot is occupied by any other user intending to use that slot.

An example of the application of these transmission schemes is presented in Figure 2. It describes a simplified situation for a system of three vehicles; two of them are already operating within a convoy. Vehicle 1 transmits its critical information in the first slot of each frame, while vehicle 2 uses the second slot of each frame. Both vehicles apply the “TX + SENSING” scheme, such that after a continuous operation they transmit successfully in these slots and maintain the ownership of using them.

Vehicle number 3 appears in the system in the second frame intending to join the convoy. In this case, it is treated as a new user who should search for a free slot to carry out the

transmission of its critical information. In a first trial, it selects the first slot of the second frame; therefore it observes the occupancy in the first part of the slot. Once the slot occupied by the other vehicle is detected, the second part of the slot is employed to listen to the transmitted information. During that period, a new iteration of the NCC-TDMA algorithm is done leading to the selection of the second transmission slot for a new trial. With the same result, the occupied slot is found for the second vehicle. Finally, after a new iteration of the algorithm, the third slot is chosen and is found to be free. As a result, the second part of the slot is used to transmit the critical information (the “SENSING + TX” scheme is applied in this slot).

Finally, in the third TDMA frame, vehicle number 3 can be considered as owner of this slot; as a result it will utilize the “TX + SENSING” scheme. At this moment, the system would reach an equilibrium situation. Only the appearance of new vehicles, interferences, or leaving any vehicle from the convoy could unbalance the equilibrium of the system.

5.1.2. Learning and Adaptability: Estimated Allocation Vector.

The results obtained from the observation of the medium during the “SENSING” phases are saved in the eav of each vehicle. This vector has as many elements as slots in the TDMA frame, in this way each element of the vector corresponds with a slot of the frame. Each element contains scalar information about the estimation of finding that free slot. A value close to maximum indicates a high estimate to find the slot free, and a value close to minimum is associated with a high estimate to find the slot occupied.

During the operation of the system, the values of the vector eav are updated in terms of the result obtained with the observation of the medium as explained later, see Section 5.2. In this way, the different elements of the vector change such that the index of the maximum element of the vector determines the estimated free slot in each iteration. It is possible to say that eav saves the information collected at each observation and acts as a tool to choose the transmission slot in the TDMA frame.

This mechanism allows the system adapt to the changes of the environment. The changes in the occupation of the different slots of the TDMA frame are reflected in the values of the users’ eav with each observation of the medium. For example, the appearance of an interfering emitter or new vehicles joining the convoy will be detected during

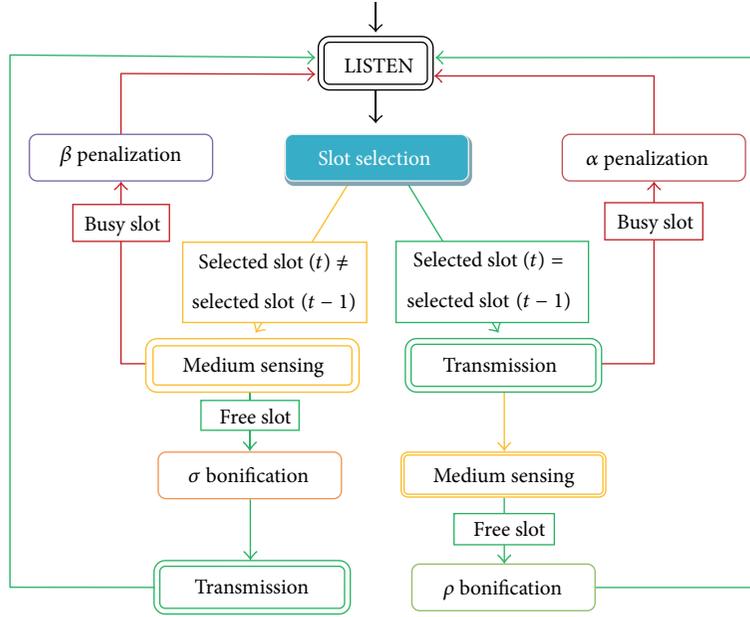


FIGURE 3: NCC-TDMA algorithm flow diagram.

the observation of the medium by each user, reducing the estimation of finding these slots free.

On the other hand, the dynamic change of users' eav causes a learning effect. The existence of interference or occupancy patterns of the TDMA frame is reflected in the medium term in the values of each user's eav. In this way, the users tend to avoid the slots that are found occupied historically.

The next section explains how the contents and the evaluation of the user's eav are parameterized.

5.2. NCC-TDMA Algorithm. The NCC-TDMA algorithm manages the operation of each user within the communication system, specifically in the medium access control. Essentially each user can carry out the next operations:

- (A) medium access: in order to receive data (LISTEN), observe the occupancy of the channel (SENSING), or transmit information (TX),
- (B) update its eav: starting from the result of an operation of observation of the medium (SENSING),
- (C) select a slot of the TDMA frame: when it is necessary to transmit data.

Figure 3 shows the flow chart of the NCC-TDMA algorithm. The starting point of the algorithm is the "LISTEN" state. In this situation, the user is listening to the medium waiting the data reception of the rest of vehicles in the system. Once the application determines the need to transmit information, the selection of the slot is carried out.

The slot is selected by obtaining the index of the maximum value of the user's eav. In the case of starting the system

operation, each user has an eav by default. For example, for a TDMA frame of 10 slots, it can be

$$\text{eav}_{\text{default}} = [10, 0, 7, 0, 6, 0, 3, 0, 0, 0]. \quad (1)$$

Moreover, it is necessary to know at which moment of the TDMA frame the user is found. For example, if during the execution of the algorithm, it enters into the function of slot selection at the start of the TDMA frame. The slot is determined on the complete eav, given that all the possible indices are associated in this moment with transmission slots in future time. However, if the block of slot selection is executed in the mid of the TDMA frame, only the values of the vector associated with slots in future time will be taken into account. Returning to the previous example, for a vehicle with eav, such as that of (1), executing the function of slot selection at the beginning of the frame, it will obtain the slot number 1 as a candidate (index of the maximum value of the vector). If this vehicle runs this function at an instant corresponding to the third slot, the eav on which the selection will be done agrees with

$$\begin{aligned} \text{eav}_{\text{partial}} &= [\text{eav}_{\text{default}4}, \text{eav}_{\text{default}5}, \dots, \text{eav}_{\text{default}9}, \text{eav}_{\text{default}10}] \\ &= [0, 6, 0, 3, 0, 0, 0] \end{aligned} \quad (2)$$

for which the selected slot would be number 5. As can be appreciated in (2), the partial eav coincides with the elements of (1) vector corresponding to the slots from 4 to 10, according to the instant at which the transmission slot is being calculated.

Once the slot has been selected, it is necessary to determine whether this slot is the same as in the previous iteration

of the algorithm. If so, this means that the user can be considered as the owner of this slot and the TX + SENSING scheme is used. On the contrary, the user should firstly be sure that the slot is free; therefore the first action to be done is the observation of the medium (SENSING).

In both cases, once the transmission has been carried out, it is returned to the initial state of listening. After the observation of the medium, the eav is updated through the bonuses or penalties application. All the bonuses and penalties consist of multiplying a parameter $(\alpha, \beta, \sigma, \rho)$ by the value of the element associated with the selected slot as follow.

Penalization α : given that $0 < \alpha < 1$, it weakens the selection of the associated slot in future iterations, it is applied when the slot considered in property is busy (the selected slot is the same as in the previous iteration).

Penalization β : given that $0 < \beta < 1$, the vector element is decreased reducing the probability of selecting that slot in the next iterations. It is applied when a new slot (which is not considered in property) is busy.

Giving a Bonus σ : given that $1 < \sigma$, the vector element increases. It takes place when the observation of the medium indicates that a new free slot has been found, and therefore, it is expected to remain free in future iterations. This slot is given a bonus to enhance its choice in the next cycles.

Giving a Bonus ρ : in the same mode like σ , $1 < \rho$. In this case, the vector element is given a bonus when the slot in which transmission took place is free and is considered in property.

In case of not transmitting in the selected slot (after applying α or β penalties), a new iteration of the algorithm is triggered. In case of having a successful transmission, after applying σ or ρ bonuses, the user remains in the LISTEN state until the next TDMA frame.

5.2.1. Characteristic Parameters of the NCC-TDMA Algorithm.

The NCC-TDMA algorithm is parameterized so that it is possible to control the behavior of the mechanism. That is, it is possible to modify the mode in which the observation on the medium occupation influences the development of the user's eav. A rapid evolution of the vector causes the users to change their slot often before minor variations of the medium. A slow evolution ensures that the users tend to stay in the same slots.

All parameters control or limit the evolution of the values of the eav components. The number of iterations required for convergence of the system to an equilibrium situation in the medium access depends on the primary and secondary parameters.

The primary parameters are those of bonuses (ρ y σ) and penalties (α y β) already described in the previous section.

On the other hand, the following secondary parameters or restrictions on eav are defined.

- (i) Maximum value eav_{\max} : limits the upper value of any element of the vector. Giving a bonus to an element is not applied if it has already reached the maximum value.

- (ii) Minimum value eav_{\min} : it is assumed that the minimum value is zero. Hence, the values of the elements of the vector vary between 0 and eav_{\max} .

- (iii) Constant total sum eav_{sum} : the sum of all elements of eav remains constant during the iterations of the algorithm. This implies that when applying a bonus on an element, the increment should be subtracted from other nonnull elements of the vector. Similarly, the decrement after applying a penalty should be distributed among other elements of the vector until the total sum is eav_{sum} .

- (iv) Number of nonnull elements eav_{NonZero} : a fixed number of vector elements whose value is not zero must be maintained. Thus, there are always candidate elements of the vector to be selected during the iteration of the algorithm.

For a better understanding of the operation of the algorithm and the function of each of the defined parameters, different examples of applications are included in the results section.

5.3. Equilibrium. Given that the present work is limited to the situation in which the maximum number of users transmitting at the same TDMA frame has to be less than or equal to the number of slots in the TDMA frame, possible solutions for media access, or equilibrium situations, can be described mathematically as detailed below.

As already indicated, the eav of each user contains the estimated information of occupancy of the slots of the TDMA frame. This vector has as many elements as there are slots in the TDMA frame, so that each vector element is associated with each one of the slots in the frame. According to this information, the algorithm selects a slot to access the medium in each iteration.

Similarly, you can define an allocation vector with as many elements as slots in the TDMA frame so that each element of the vector is associated to each slot of the frame. Each element will have a value of 1 for a slot occupied by the user and 0 for the remaining values. By combining all these user allocation vectors, the system allocation matrix is obtained.

For example, in a system with 4 users and 10 transmission slots, the allocation vector for user 1 occupying the third transmission slot is

$$a_1 = [0010000000]. \quad (3)$$

Thus, the allocation matrix takes the following form:

$$A = \begin{pmatrix} 0010000000 \\ 0100000000 \\ 0001000000 \\ 1000000000 \end{pmatrix}. \quad (4)$$

The matrix shown in (4) corresponds to a system in which users 1, 2, 3, and 4 transmit in slots 3, 2, 4, and 1, respectively. In other words, the matrix reflects an organized TDMA system in which each user is allocated during an exclusive transmission slot.

When this approach is applied to a system of n users and r slots (where $r \geq n$), resolving transmission schedule is reduced to finding how to order the matrix in such a way that the unit values of each user's allocation vector are located in different matrix columns. This situation is described by what are known as permutation matrices.

A permutation matrix is a square matrix where all its $n \times n$ elements are equal to 0, except one, which should be equal to 1, in each row and column. According to this definition, $n!$ different permutation matrices are possible.

If this definition is applied to the problem under study, the outcome is a situation where, in an organized system of n users and r slots, each of the users transmits using a slot which is not occupied by any other system user.

The strict definition of a permutation matrix is limited to a square matrix corresponding to a system with the same number of available slots as users. This situation is the most restrictive for the medium access control mechanism described in this section, but it is not the only one since a system where the number of available transmission slots is greater than the number of users can also be applied. Thus, in this paper we extend the definition of a permutation matrix to include any matrix of $n \times r$ elements (where $r \geq n$) in which each of the rows has only one value equal to 1, located in a column in which there is no other value equal to 1. According to this extended definition, the matrix in (4) is a permutation matrix.

In the case of n users in a system with r slots, the number of possible matrix permutations is given by the following expression:

$$P(r, n) = \frac{r!}{(r-n)!}, \quad \forall r \geq n. \quad (5)$$

Taking the description of the solution to the problem one step further, the aim is to reorder the system allocation matrix so that it becomes a permutation matrix.

6. Results

In order to quantify the NCC-TDMA mechanism performance, a simulation experiment has been designed taking into account the special scenario of autonomous vehicles guidance and platooning manoeuvring. For such scenario, a reduced number of users have to be considered, in a different way than typical ITS experiments where lots of users are involved. Results shown in this paper are a new updated version of the results included in the authors' previous work [28].

The performed test considers the initial state of a group of transport units trying to form a convoy. Given a set of users with a number of transmission slots available, an initial situation was proposed in which each of the users had randomly selected a slot. The random selection of each user initial state was performed in order to start the test with a conflict situation between two or more users. In that way the algorithms under study are forced to find an equilibrium solution without conflicts among users. The number of algorithm iterations (number of time slots required to find

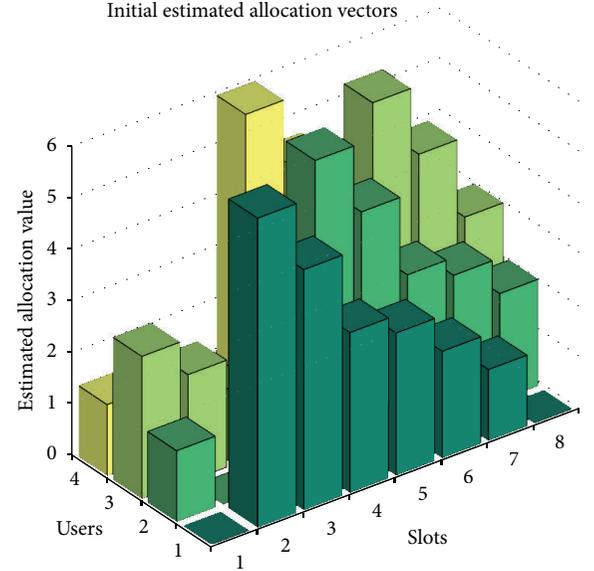


FIGURE 4: Initial estimated allocation vectors of the example with 4 users and 8 slots.

TABLE 3: Characteristic parameters for the experiment with NCC-TDMA.

eav_{max}	eav_{min}	eav_{sum}	$eav_{NonZero}$	ρ	σ	α	β
6	0	20	6	1.2	1.2	0.2	0.2

the equilibrium) has been counted to measure the algorithm performance. In order to find the equilibrium, the algorithm has to reorder the transmission schedule so that each user could select a slot which had not been selected by another.

With the aim of simplifying the experiment, the application example considers 4 users and 8 slots. Firstly, a graphic illustration of how the NCC-TDMA algorithm operates is provided. Figure 4 shows the randomly generated eav 's which values were obtained for a simulation with 4 users and 8 slots. These initial vectors were generated according to the parameters shown in Table 3. The slots selected by each user, at the initial iteration, are shown in Figure 5 (maximum values of each vector).

Focusing on the algorithm efficiency to reorder the schedule, when conflict arose among users, the TX + SENSING scheme is assigned to low user identifiers and the SENSING + TX scheme in other cases.

Since initial strategy selection was random, 2 vehicles were detected as being in conflict at the start of the assay. These were vehicles 2 and 4, whose maximum values of its estimated allocation vectors were located in slot 4. To solve the conflict, NCC-TDMA waits until the initial time of the conflict slot; at this moment the algorithm of the user in conflict changes its selection by updating its estimated allocation vector by bonus and penalties. The new slot selection is performed until the equilibrium is reached. For the basic test described here to show the mechanism behaviour, user 4 reallocates its selection to use slot number 5 after 5 iterations, as shown in Figures 6 and 7.

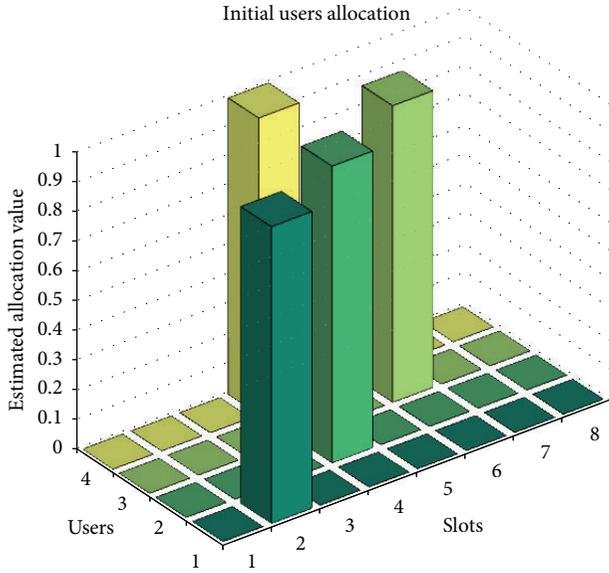


FIGURE 5: Initial allocation for the example under test.

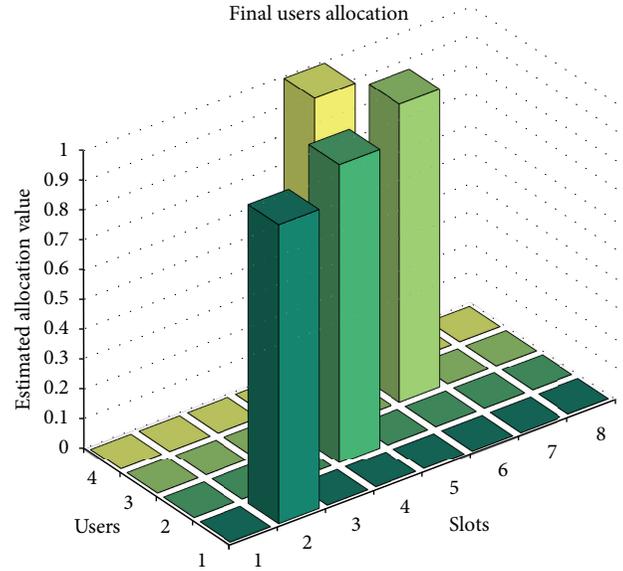


FIGURE 7: Final allocation for the example under test.

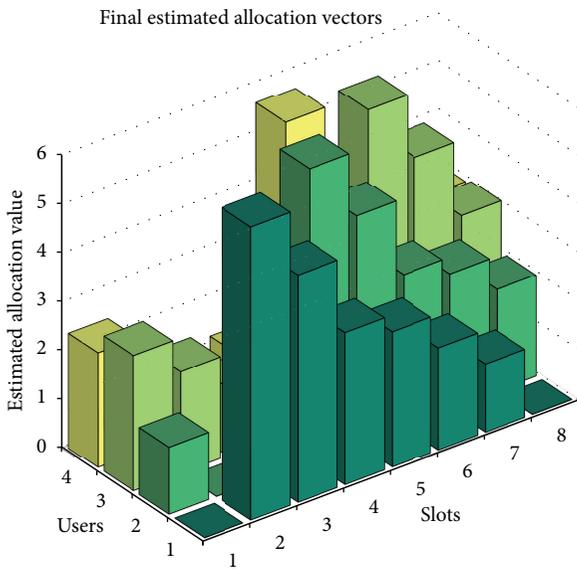


FIGURE 6: Final estimated allocation vectors of the example with 4 users and 8 slots.

In all the cases, following successful transmission, the eav value associated with the corresponding slot was awarded an incentive to reinforce the same decision in future iterations. In this way, once equilibrium had been reached, the eav values for each user's slot selection were reinforced maintaining the equilibrium state.

6.1. Comparison with Slotted-Aloha. With the objective of comparing the authors' proposal with another alternative medium access control protocol, the same simulation experiment was carried out using the Slotted-Aloha mechanism [18].

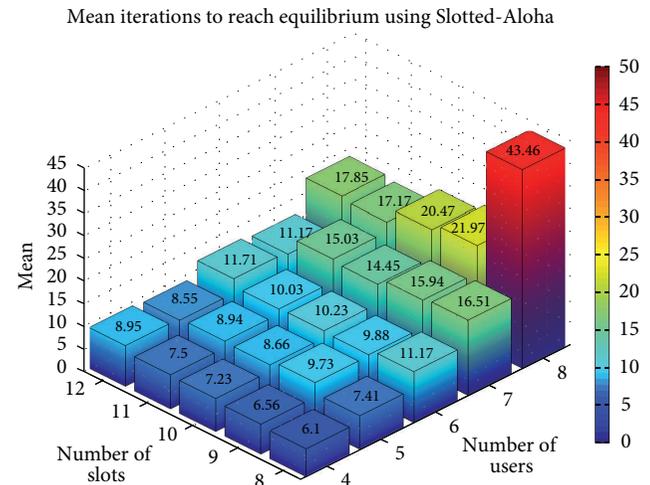


FIGURE 8: Results of the repeated experiment using Slotted-Aloha.

The Slotted-Aloha protocol was selected for two reasons. Firstly, because it is one of the most widely studied protocols in the literature, and secondly because we considered it more instructive to compare the NCC-TDMA with a generic MAC mechanism which is not oriented to the application and does not use signalling, as it is the case of STDMA [10] and VeSOMAC [21].

In the Slotted-Aloha MAC mechanism [18], medium access is obtained using TDMA time slots. When a user needs to transmit, this is attempted at the beginning of the following time slot. If this slot is occupied, a random number is generated to determine the number of time slots the user must wait for attempting a new transmission. In case of collisions, the involved users must wait for a random number of slots, implying a certain level of probability that the collision will be repeated.

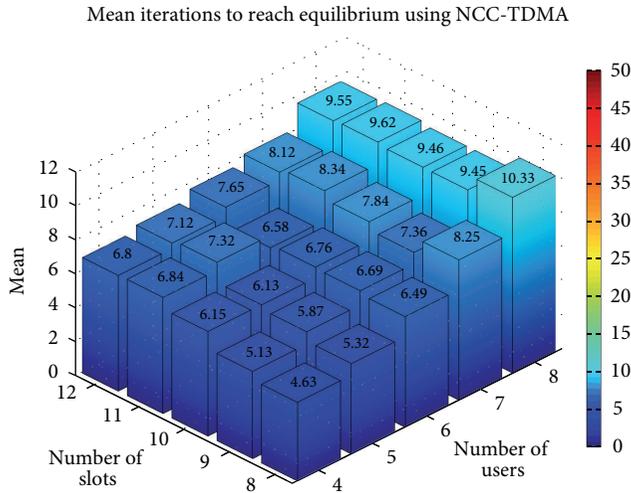


FIGURE 9: Results of the repeated experiment using NCC-TDMA.

The results shown in Figures 8 and 9 correspond to the mean number of iterations necessary for each mechanism to achieve the equilibrium state where each user has a free slot, coming from a randomly generated access schedule.

The mean value was obtained after 100 repetitions and for simulations using from 4 to 8 users and from 8 to 12 slots. The characteristic parameters shown in Table 3 were applied to NCC-TDMA. The most restrictive simulation was the one involving 8 users and 8 slots. As can be observed, for the same number of slots, the greater number of users the greater number of iterations needed, due to the possibility of conflict among users increases. The number of iterations is evaluated as the number of time slots until the equilibrium is reached. In fact, this increment in the number of iterations is caused by the position of the conflict, not for the algorithm performance requirements. In the most restrictive case, the NCC-TDMA mechanism solves the medium access order of users applying less iterations than the Slotted-Aloha mechanism. The mean number of iterations for NCC-TDMA is 10.33 and the mean number of iterations for Slotted-Aloha is 43.46.

7. Discussion and Conclusion

This work describes a medium access control mechanism complying with real-time requirements for ITS applications. Special mention deserves V2V case of platooning guidance and manoeuvring where the loss of critical information commits the global stability of the convoy formation.

The main contribution of the NCC-TDMA proposal is a schedule-based mechanism, which, on one hand, eliminates the possibility of medium access collision and, on the other hand, does not require signalling. Thanks to Cognitive Radio techniques as spectrum sensing, learning, and adaptability, it is a generic protocol capable of detecting conflicts among system users or due to outside signal interference.

As in other systems based on self-organized TDMA, each user is responsible for determining the medium access schedule employed by the system. Furthermore, with the

NCC-TDMA, users have the necessary information without needing any more than the information they themselves can obtain from observation of the channel. Consequently, they are capable of detecting the occupation of a slot, whether this is occupied by another user or by an interfering signal.

The result is a dynamic mechanism which adapts to environmental conditions, however, variable these may be. Moreover, the mechanism is capable of learning from these conditions, for instance, avoiding certain transmission slots which are sporadically occupied by interfering signals.

For guidance of autonomous vehicles in convoy, the medium access controls is of vital relevance in split and merge manoeuvres. In these cases, the mechanism of medium access control must deal with new vehicles joining the convoy, keeping a satisfactory transmission for member vehicles. Finally, it must be highlighted the fact that the NCC-TDMA mechanism is capable of acting as a stand-alone medium access control protocol, or as a complement to the protocol incorporated into the WAVE standard. However, this first version does require time synchronization for each user.

To summarize, it can be concluded that the medium access control mechanism described by authors is characterized by the following.

- (i) It is based on a self-organized dynamic TDMA scheme.
- (ii) It does not require the exchange of signalling among users.
- (iii) It does not require compulsory exchange of frames among users if a user does not have any application information to transmit.
- (iv) Once a stable situation has been reached, the mechanism remains in this state until it detects a new event, which may include channel occupation by another user in the system, and channel occupation by an outside interfering signal.
- (v) The system adapts to environment variations and is capable of learning from repeated situations.

However, there are some situations in which the current version of NCC-TDMA is not able to cope with by its own. The case of two or more users exploring a free slot at the same time and finding it as free will provoke a conflict in the next iteration as both users will transmit in the same slot. Same problem will happen according to the well-known hidden terminal problem. For both cases, authors are working on NCC-TDMA review where these problems are going to be tackled in higher layers of the protocol stack where the NCC-TDMA offers alternatives to update the estimated allocation vector. On top of that, two main work lines have been identified as future work.

- (i) The NCC-TDMA algorithm will be tested against disturbances and uncertainties over the specific scenario of V2V communication in generic ITS environments.
- (ii) The NCC-TDMA algorithm efficiency will be compared with other alternatives for V2V communication that uses signalling as RR-Aloha.

Acknowledgments

This work has been supported by the Spanish Ministry of Science and Innovation through the VISNU (Ref. TIN2009-08984) and the ESPIRA (Ref. 2010–2013) Projects.

References

- [1] J. Jakubiak and Y. Koucheryavy, "State of the art and research challenges for VANETs," in *Proceedings of the 5th IEEE Consumer Communications and Networking Conference (CCNC '08)*, pp. 912–916, January 2008.
- [2] C. Santos, F. Espinosa, D. Pizarro, F. Valdés, E. Santiso, and I. Díaz, "Fuzzy decentralized control for guidance of a convoy of robots in non-linear trajectories," in *Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation (ETFA '10)*, pp. 1–8, 2010.
- [3] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: a fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.
- [4] Y. L. Morgan, "Notes on DSRC & WAVE standards suite: its architecture, design, and characteristics," *IEEE Communications Surveys and Tutorials*, vol. 12, no. 4, pp. 504–518, 2010.
- [5] G. Karagiannis, O. Altintas, E. Ekici et al., "Vehicular networking: a survey and tutorial on requirements, architectures, challenges, standards and solutions," *IEEE Communications Surveys and Tutorials*, vol. 13, no. 4, pp. 584–616, 2011.
- [6] CAR 2 CAR Communication Consortium Manifesto, 2007, http://elib.dlr.de/48380/1/C2C-CC_manifesto_v1.1.pdf.
- [7] D. Jiang and L. Delgrossi, "IEEE 802.11p: towards an international standard for wireless access in vehicular environments," in *Proceedings of the 67th IEEE Vehicular Technology Conference (VTC '08)*, pp. 2036–2040, May 2008.
- [8] T. Ernst, V. Nebehaj, and R. Søråsen, "CVIS: CALM proof of concept preliminary results," in *Proceedings of the 9th International Conference on Intelligent Transport Systems Telecommunications (ITST '09)*, pp. 80–85, October 2009.
- [9] N. Ferreira, J. A. Fonseca, and J. S. Gomes, "On the ad cy of 802.11p MAC protocols to support safety services in ITS," in *Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation*, pp. 1189–1192, 2008.
- [10] K. Bilstrup, E. Uhlemann, E. G. Ström, and U. Bilstrup, "On the ability of the 802.11p MAC method and STDMA to support real-time vehicle-to-vehicle communication," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 902414, pp. 1687–1472, 2009.
- [11] IEEE 802.11p, "Standard for Information technology—telecommunications and information exchange between systems—local and metropolitan area networks—specific requirements—part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 6: Wireless Access in Vehicular Environment".
- [12] R. A. Saeed, A. B. HjNaemat, A. B. Aris, I. M. Khamis, and M. K. Bin-Awang, "Evaluation of the IEEE 802.11p-based TDMA MAC method for road side-to-vehicle communications," *International Journal of Network and Mobile Technologies*, vol. 1, no. 2, pp. 81–87, 2010.
- [13] S. Sehrawat, R. P. Bora, and D. Harihar, "Performance analysis of QoS supported by Enhanced Distributed Channel Access (EDCA) mechanism in IEEE 802.11e," *IAENG International Journal of Computer Science*, vol. 33, article 1, 2007.
- [14] S. Eichler, "Performance evaluation of the IEEE 802.11p WAVE communication standard," in *Proceedings of the 66th IEEE Vehicular Technology Conference (VTC '07)*, pp. 2199–2203, 2007.
- [15] S. A. Mohammad, A. Rasheed, and A. Qayyum, "VANET architectures and protocols stacks: a survey," in *Communication Technologies for Vehicles*, vol. 6596 of *Lecture Notes in Computer Science*, pp. 95–105, Springer, 2011.
- [16] K. Fawaz, A. Ghandour, M. Olleik, and H. Artail, "Improving reliability of safety applications in vehicle ad hoc networks through the implementation of a cognitive network," in *Proceedings of the 17th IEEE International Conference on Telecommunications (ICT '10)*, pp. 798–805, April 2010.
- [17] S. Kumar, V. S. Raghavan, and J. Deng, "Medium access control protocols for ad hoc wireless networks: a survey," *Ad Hoc Networks*, vol. 4, no. 3, pp. 326–358, 2006.
- [18] L. Roberts, "Aloha packet system with and without slots and capture," *ACM SIGCOMM Computer Communication Review*, vol. 5, no. 2, pp. 28–42, 1975.
- [19] F. Borgonovo, A. Capone, M. Cesana, and L. Fratta, "ADHOC MAC: new MAC architecture for ad hoc networks providing efficient and reliable point-to-point and broadcast services," *Wireless Networks*, vol. 10, no. 4, pp. 359–366, 2004.
- [20] F. Borgonovo, A. Capone, M. Cesana, and L. Fratta, "RR-ALOHA, a Reliable R-ALOHA broadcast channel for ad-hoc inter-vehicle communication networks," Dipartimento Elettrotecnica e Informazione, Politecnico di Milano, 2002medhoc.pdf 2002medhoc.pdf.
- [21] F. Yu and S. Biswas, "Self-configuring TDMA protocols for enhancing vehicle safety with DSRC based vehicle-to-vehicle communications," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1526–1537, 2007.
- [22] W. Crowther, R. Rettberg, D. Waldem, S. Ornstein, and F. Heart, "A system for broadcast communications: reservation ALOHA," in *Proceedings of the 6th Hawaii International Conference on System Sciences*, January 1973.
- [23] D. Cabric, S. M. Mishra, and R. W. Broerdsen, "Implementation issues in spectrum sensing for cognitive radios," in *Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers*, pp. 772–776, November 2004.
- [24] E. Hossain and V. K. Bhargava, *Cognitive Wireless Communication Networks*, Springer, 2007.
- [25] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.
- [26] J. Mitola, *Cognitive Radio Architecture: The Engineering Foundations of Radio XML*, John Wiley & Sons, 2006.
- [27] M. Di Felice, K. R. Chowdhury, and L. Bononi, "Analyzing the potential of cooperative cognitive radio technology on inter-vehicle communication," in *Proceedings of the IFIP Wireless Days (WD '10)*, October 2010.
- [28] M. Manzano, F. Espinosa, A. M. Bravo, D. Garcia, A. Gardel, and I. Bravo, "Medium access control based on a non cooperative cognitive radio for platooning communications," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 408–413, 2012.

Research Article

Multiobjective Model for Emergency Resources Allocation

Zhaosheng Yang,^{1,2} Huxing Zhou,¹ Xueying Gao,¹ and Songnan Liu¹

¹ College of Traffic, Jilin University, Changchun 130012, China

² State Key Laboratory of Automobile Dynamic Simulation, Jilin University, Changchun 130012, China

Correspondence should be addressed to Huxing Zhou; zhouhuxing@gmail.com

Received 7 August 2012; Revised 21 November 2012; Accepted 18 December 2012

Academic Editor: Wuhong Wang

Copyright © 2013 Zhaosheng 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.

Emergency resources allocation is essential to the emergency rescue effectiveness, and it has become a research focus for emergency rescue. This paper proposes a multiresource dynamic allocation model of emergency rescues and corresponding solving method. The object of the proposed model is to maximize the overall emergency rescue effectiveness of allocated resources and minimize the allocating costs of resources. The model considers the dynamic nature that the casualties of trapped victims change over time. At last, a numerical example is presented to test the model and its algorithm.

1. Introduction

Emergency resources allocation is the material foundation of the emergency rescue supplies and can greatly improve the supply efficiency of the emergency resources. In this paper, the emergency resources allocation studies primarily aim at facing the multihazard disaster threat during the emergency regional preparedness. A multistage resources allocation optimization model is formulated which is more dynamic based on the temporal and spatial variation features of the emergency demand.

Alsalloum and Rand studied the minimum quantity and the location optimization of the emergency relief ambulance, showing how to choose the right location for the ambulance to maximize the expected demand coverage while minimizing the number of vehicles with the goal programming method. Rajagopalan et al. proposed a model to meet the dynamic demands of the ambulance during a certain period of time [1, 2]. Yi and Özdamar came up with an integrated location distribution model for coordinating logistics support and evacuation operations in disaster response activities, which is an LRP problem [3]. Chang et al. formulated the flood emergency logistics problem with uncertainty as two stochastic programming models, the grouping and classifying model and the location-allocation model, which allow the determination of a rescue resource distribution system for urban flood disasters [4]. Rajagopalan et al. formulated

the dynamic available coverage location (DACL) model to determine the minimum number of ambulances and their locations for each time cluster in which significant changes in demand pattern occur while meeting coverage requirements with a predetermined reliability and used the tabu search algorithm to obtain its globally optimal solutions [2]. In order to strengthen natural disaster prevention mechanism, Rawls and Turnquist developed an emergency response planning tool to obtain the emergency preparedness resources locations and inventory levels under uncertain situation [5]. Mete and Zabinsky proposed a stochastic optimization approach for the storage and distribution problem of medical supplies to be used for disaster management under a wide variety of possible disaster types and magnitudes [6].

2. Multiresource Dynamic Allocation Model for Emergency Rescue

The rationality of rescue resource allocation directly affects the effectiveness of rescue. Strategic decision of rescue resource allocation is influenced by the variety, scale, cost, and various factors. Overladen resource reserve would cause massive waste of money and resource, while too little reserve would seriously reduce rescue effectiveness.

2.1. Problem Description. The rescue resource allocation discussed in this paper is about dealing with the allocation for

frequent disasters in regular resource spot. Since the threat come from disasters, the rescue demand and resources characteristics are dynamically changeable, the rescue resource allocation should also be dynamically changeable along with the situation of the disaster. Furthermore, given that frequent disasters happen with strong regular patterns, the rescue demands in different periods of disaster can be estimated according to historic disaster records; thus, the rescue resource could also be rationally allocated. Scientific rescue resource allocation should be based on the following considerations.

- (1) The importance of resource. The importance of rescue resource refers to the utility of resource in rescue [7]. It is mainly reflected in the replacement of resource and the effect on rescue. Hereinafter, the fungible resources are seen as the same to simplify the problem without affecting resource allocation. So the extent that resource has effect on disasters is used to reflect the importance of resource. To reflect the importance of various types of resources more directly, we could hereby quantify the extent of effect on disasters brought by every types of resources, and then the importance of resources could be represented by a range of figures, for example 1–10 grades.
- (2) The timeliness of resource. The rescue resource is valuable only when it is delivered to the demanders [8]. The amount of survivors reduces with the loss of time, and the quantity of resources which could have effect in rescue also decreases. When the arrival of rescue resource has exceeded the time limit of rescue resource, the arrival of resources is meaningless.
- (3) The scarcity degree of resource. Resource scarcity degree is the degree of difficulty of collect or purchase. Rare resources, especially in emergent need, should be stored; otherwise, when needed, collecting the necessary resources would take too much time to miss the best rescue time and cause serious casualties. For the resources which could be easily purchased and collected, the storage should be rationally reduced to save storage cost without causing bad effects on rescue.
- (4) The cost of resource. The purpose of rescue resource storage is for emergency, but the cost should not be ignored. The scale of rescue resource is usually massive, and the investment from the government is also numerous. By allocating the scale of resource storage reasonably to reduce the cost of purchasing and storage would highly ease the fiscal tension of local rescue department.

2.2. Model Formulation. For easily describing and understanding the model, the model construction should satisfy the 3 premise assumptions as the following.

- (1) Because of the strong discipline of frequent disasters, we assume that the type, range, and degree of possible

disasters and the expected value of types and quantity of resources in different periods from different demanding spots are expectable.

- (2) Resource spots and demanding spots are all connected, with time of arrival known between every two spots.
- (3) Each resource spot delivers each type of resource to every demanding spot at most.

Multiresource dynamic allocation model for emergency rescues is a multigoal optimization model with two goals, maximizing rescue effectiveness and minimizing allocating cost.

- (1) Maximum of rescue effectiveness. The purpose of rescue resource allocation is to deal with emergencies, so the maximum of rescue effectiveness is the first goal and the maximum of the importance of all the resources could reflect this goal. On the other hand, according to the dynamic change of the threat of the disasters, the rescue period could be divided into several time intervals, and the model seeks the maximum of rescue resource effectiveness during the whole projection period. The so-called “maximum of rescue resource effectiveness” here concerns not only the effectiveness of the resource allocation, but also the maximum of the whole rescue project with the consideration of the whole rescue demands.

The effectiveness of rescue resource has important relationship with the scarcity degree of resource. The timeliness of resource is reflected by the amount of resources which is still of value with the loss of time and also matters the casualties of the victims. The scarcity degree of resource could be seen as the length of time needed for preparation. The resource with higher scarcity need longer time to be transferred and be produced from other place. So the time for preparation could be used to represent the degree of scarcity of resource. The so-called “time for preparation” means the time needed for the arrival of rescue resource at the demanding spot after disaster happens. The first goal of the model is as the following:

$$\text{Obj1 } f_1 = \max \sum_{k=1}^o \sum_{l=1}^p \left(u_l \sum_{i=1}^n \sum_{j=1}^m x_{ijkl} f_{jkl}(t_{ij}) + u_l \left(d_{jkl} - \sum_{i=1}^n \sum_{j=1}^m x_{ijkl} \right) f_{jkl}(t_{\text{rep}j}) \right). \quad (1)$$

In (1), i represents the number of resource spot; j represents the number of demand resource spot; k represents the number of response time; l represents the number of resources type; u_l represents the utility value of l type resources per unit; x_{ijkl} is the amount of l type resources delivered from point i to point j during the time period k ; d_{jkl} represents the demand resources of l type that point j needed during the time period k ; t_{ij} is the transfer time from point i to

TABLE 1: Conditions of the threat of disasters that each region faced at all stages.

Analytical stages		I		II		III	
Types of disasters		A	B	C	D		
Demand spots	1	0.7	0.5	0.7	0.4		
Disasters	2	0.5	0.8	0.3	0.8		
Occurrence	3	0.2	0.7	0.8	0.6		
Probability	4	0.8	0.2	0.2	0.6		

point j ; t_{repj} is the preparation time of j type rescue resources; the rescue time-effect function of the l type resources needed by demand spot j at moment t in period k is given by $f_{jkl}(t)$.

Through (1), we can find that the section before the plus sign is the emergency rescue utility value of allocated resources, while the section after the plus sign is the emergency rescue utility value of unallocated resources which need temporary preparation. This objective can reflect the total effectiveness of the whole emergency resources allocation proposal in the planning period.

- (2) The lowest-cost allocation of resources. The cost of emergency response resources are divided into three parts: purchasing cost, inventory cost, and opportunity loss cost, c_{al} and c_{bl} , respectively, represent purchase cost and inventory cost. Opportunity loss cost is due to a shortage of resources in disasters, which results in that resources which purchase cost will be higher than usual. Temporarily purchasing the unreserved emergency resources will cause a certain loss. Consequently, it will increase the cost of the emergency rescue work, which is the focus of this study. We use o_{pi} to represent the unit opportunity loss cost of the l type resources. Besides, most resource spots have a certain economic functions, so if a certain disaster did not happen that the corresponding resources can be sold to reduce the cost of the resources allocation. Assume that the sale of a resource per unit will earn p_{rl} profit and the possibility of the q type disaster happening on demand spot j during period k is α_{jkq} . We use N_{jkl} to represent the demand of the l type resources of the point j during the period k . The objective of the lowest total cost of the allocation resources is given as the following:

$$\text{Obj2} \quad f_2 = \min \sum_{j=1}^m \sum_{k=1}^o \sum_{l=1}^p \left(\sum_{i=1}^n \left((c_{al} + c_{bl}) \cdot x_{ijkl} - p_{rl} \cdot \sum_{q=1}^v (x_{ijk1} \cdot (1 - \alpha_{jkq})) \right) + o_{pl} \cdot \left(N_{jkl} - \sum_{i=1}^n x_{ijkl} \right) \right). \quad (2)$$

The temporary emergency resources location model formulated should satisfy the following constraints.

- (1) The total amount of available stored resources at different stages allocated cannot exceed its reserve capacity. Here is the mathematical expression of the following constraint equation (3), N_{ik} is the reserve capacity of the resource spot i :

$$\sum_{j=1}^m \sum_{l=1}^p x_{ijkl} \leq N_i. \quad (3)$$

- (2) The total resources that supply all kinds of resources spot cannot exceed the actual total demand. It is given by (4):

$$\sum_{i=1}^n x_{ijk1} \leq N_{jk1}. \quad (4)$$

- (3) The amount of the resources allocated to each resources spot should be a positive number, as shown in (5):

$$x_{ijk1} \in \{0, R^+\}. \quad (5)$$

The proposed dynamic emergency rescues allocation model is a typical multiobjective programming model. Similar to above-mentioned optimization location model, the requirements of both objectives should be considered in an integrative way. The model is solved by the ideal point method (IPM) in order to achieve a noninferior solution with the result of each subobjective close to the optimum value. The result derived from this method can be further improved by utilizing the idea of the unit cost utility method, as demonstrated in (6):

$$\text{Obj}^{**} \quad \max \frac{f_1}{f_2}. \quad (6)$$

The details of the solving steps are as follows.

Step 1. Solve the model composed of (1) and (3)–(5) for an optimized value (here we go for the minimum value with a negated object), denoted as f_1^* , go to Step 2.

Step 2. Solve the model composed of (2)–(5) except (1) for an optimized value, denoted as f_2^* , go to Step 3.

Step 3. Solve the model composed of (7), (3)–(5) for an optimized value, denoted as f_3^* , go to Step 4.

$$\text{Obj} \quad \min f_3 = \sqrt{(-f_1 - f_1^*)^2 + (f_2 - f_2^*)^2}. \quad (7)$$

Step 4. Construct the objective function and constrains, as illustrated in (8) and (9), (8) is set as the objective for

TABLE 2: Quantities of the demand of resources that different demand spots respond to various types of disasters.

Types of disasters	A		B		C		D
	1	2	3	4	5	6	
Types of resources	1	2	3	4	5	6	7
Effectiveness of resource units	5	8	6	5	9	7	
All types of resources required for demand spots (m ³)	1	1500	2000	1200	1200	1200	800
	2	1200	1500	2500	500	400	2400
	3	600	800	2400	2000	1800	1800
	4	2000	2500	500	300	300	1800

TABLE 3: Scales of resources spots.

Resources spots	1	2	3
Scales (m ³)	5000	5000	6000

TABLE 4: Shortest time of transportation between resource and demand spots (h).

Demand spots	Resources spots		
	1	2	3
1	35	25	20
2	30	50	28
3	45	30	56
4	30	40	30

TABLE 5: Time of preparation that each demand spot needs all types of emergency resources (h).

Resources	Demand spots			
	1	2	3	4
1	35	35	40	40
2	48	48	48	48
3	30	25	36	30
4	40	20	20	50
5	30	30	40	50
6	40	25	20	30

TABLE 6: Cost associated with all types of resources.

Resource	1	2	3	4	5	6
c_{ai} (¥/m ³)	200	120	240	160	160	100
c_{bi} (¥/m ³)	10	8	10	8	10	8
p_{ri} (¥/m ³)	160	100	200	20	40	80
o_{pl} (¥/m ³)	30	0	40	30	40	0

optimization and (2)–(5) and (9) are used as constrains, the resulting solution is the optimized solution:

$$\text{Obj } \max f_4 = \frac{f_1}{f_2} \tag{8}$$

$$f_1 = f_3^* \tag{9}$$

Solutions are carried out by the software LINGO.

TABLE 7: Algorithm solution obtained from ideal point effectiveness of unit cost method (m³).

Analytical stages	Resources	Resource spots		
		1	2	3
I	1	0	600	1500
	2	1963	800	4037
	3	0	2181	1200
II	4	167	0	1333
	5	169	1800	1331
III	6	681	1800	4319

TABLE 8: Solutions obtained by ideal points (m³).

Analytical stages	Resources	Resource spots		
		1	2	3
I	1	2128	600	2500
	2	2500	800	3500
	3	0	2400	1200
II	4	159	0	1341
	5	158	1800	1342
III	6	900	1800	4100

3. Experimental Result

We proposed a numerical example to verify the emergency rescue resource allocation model. Assume that there are four demand spots in the analytical region, and the conditions of the region threatened by the disaster on average per year can be roughly divided into three stages during the planning period. The types of disasters and the quantities and types of required emergency resources in each stage are shown in Tables 1 and 2, respectively. There are three fixed resource spots in the region, and the scale of each resource spot is shown in Table 3. The time of transportation between each resource and demand spot is shown in Table 4, and the preparation time of all types of emergency resources each demand spot needs is show in Table 5. The cost of purchase and maintenance of all kinds of resources and the income of resources sale are shown in Table 6, the total investment of the allocation of resources does not exceed of ¥3.5 million. The functions of emergency timeliness of corresponding

TABLE 9: Solutions obtained by the maximum effectiveness of single target.

Analytical stages	Resources	Resource spots		
		1	2	3
I	1	2200	600	2500
	2	2500	800	3500
	3	0	2400	1200
II	4	0	0	1500
	5	0	1800	1500
III	6	0	1800	5000

TABLE 10: Comparison of indicators related to algorithm solution obtained from three methods.

Solutions	Indicators		
	Emergency effectiveness (the first target)	Total cost of configuration (million ¥)	Cost effectiveness of configuration
Ideal point effectiveness of unit cost	173719	325	536
Deal point	173719	368	472
The maximum effectiveness of single target	174877	371	472

resources about different demand spots respond to various types of disasters are as the following:

$$\begin{aligned}
 q_A(t) &= \exp\left(-0.1\left(\frac{t}{10}\right)\right), \\
 q_A(t) &= \exp\left(-0.2\left(\frac{t}{24}\right)^{1.5}\right), \\
 q_B(t) &= \exp\left(-0.3\left(\frac{t}{32^2}\right)\right), \\
 q_C(t) &= \exp\left(-0.8\left(\frac{t}{48}\right)^2\right), \\
 q_D(t) &= \exp\left(-0.1\left(\frac{t}{18^2}\right)\right), \\
 q_D(t) &= \exp\left(-0.2\left(\frac{t}{54}\right)^{2.5}\right).
 \end{aligned}
 \tag{10}$$

The optimized solution based on the algorithm is shown in Table 7. The solution of Step 4 without the algorithm above, namely, only use of the ideal point is shown in Table 8. Regardless of cost, that is, remove the second objective to obtain the solution which is shown in Table 9. The comparison of indicators related to algorithm solution obtained from three methods is shown in Table 10.

From Table 10, it can be seen that IPM may take the first and second objectives into account to some extent. The total value of effectiveness of IPM declines compared with the model which only considers the first objective, while the configuration costs are reduced, the effect is not obvious.

The cost and the emergency effectiveness are only reduced by 0.8% and 0.7%, respectively, but by the proposed algorithm, the cost of resources allocation is reduced by 12.4% under the condition that the emergency effectiveness only is

reduced by 0.7%. The quality of the optimization is greatly improved.

4. Conclusions

Based on the actual demand of the emergency rescue resources allocation, this paper proposes multiemergency rescue dynamic allocation model. The model considers the dynamic nature that the casualties of trapped victims change over time. The model for the allocation of rescue resources aims at maximizing the overall emergency rescue effectiveness of rescue proposal of the allocated and the unallocated resources at each stage during the planning period, and it also allows for the cost of allocation resources. Considering the purchasing cost, inventory cost, and opportunity loss cost during the cost objectives construction, we propose the method that combines ideal point method and unit cost utility method to solve the allocation model by utilizing the LINGO software. At the end of this paper, we design the numerical example to test the model and its algorithm.

Conflict of Interests

The authors claim that there is not any conflict of interests in their research.

Acknowledgment

Zhaosheng Yang express his sincere appreciation to the Chinese National High Technology Research and Development Program Committee no. 2009AA11Z208 and Project 20121100 supported by Graduate Innovation Fund of Jilin University for the financial support of this paper.

References

- [1] O. I. Alsalloum and G. K. Rand, "Extensions to emergency vehicle location models," *Computers and Operations Research*, vol. 33, no. 9, pp. 2725–2743, 2006.
- [2] H. K. Rajagopalan, C. Saydam, and J. Xiao, "A multiperiod set covering location model for dynamic redeployment of ambulances," *Computers and Operations Research*, vol. 35, no. 3, pp. 814–826, 2008.
- [3] W. Yi and L. Özdamar, "A dynamic logistics coordination model for evacuation and support in disaster response activities," *European Journal of Operational Research*, vol. 179, no. 3, pp. 1177–1193, 2007.
- [4] M. S. Chang, Y. L. Tseng, and J. W. Chen, "A scenario planning approach for the flood emergency logistics preparation problem under uncertainty," *Transportation Research E*, vol. 43, no. 6, pp. 737–754, 2007.
- [5] C. G. Rawls and M. A. Turnquist, "Pre-positioning of emergency supplies for disaster response," *Transportation Research B*, vol. 44, no. 4, pp. 521–534, 2010.
- [6] H. O. Mete and Z. B. Zabinsky, "Stochastic optimization of medical supply location and distribution in disaster management," *International Journal of Production Economics*, vol. 126, no. 1, pp. 76–84, 2010.
- [7] W. Wang, F. Hou, H. Tan, and H. Bubb, "A framework for function allocations in intelligent driver interface design for comfort and safety," *International Journal of Computational Intelligence Systems*, vol. 3, no. 5, pp. 531–541, 2010.
- [8] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: a fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.

Research Article

Study on Vehicle Track Model in Road Curved Section Based on Vehicle Dynamic Characteristics

**Ren Yuan-Yuan,^{1,2} Zhao Hong-Wei,¹
Li Xian-Sheng,² and Zheng Xue-Lian²**

¹ College of Computer Science and Technology, Jilin University, Changchun 130022, China

² College of Traffic, Jilin University, Changchun 130022, China

Correspondence should be addressed to Li Xian-Sheng, lixiansheng123@yahoo.com.cn

Received 3 August 2012; Revised 19 November 2012; Accepted 20 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Ren Yuan-Yuan 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.

Plenty of experiments and data analysis of vehicle track type in road curved section show that the deviation and the crossing characteristics of vehicle track paths are directly related to the driving stability and security. In this connection, the concept of driving trajectory in curved section was proposed, six track types were classified and defined, and furthermore their characteristic features were determined. Most importantly, considering curve geometry and vehicle dynamic characteristics, each trajectory model was established, respectively, and the optimum driving trajectory models were finally determined based on the crucial factors of vehicle yaw rate, which was also the most important factor that impacts vehicle's handling stability. Through it all, MATLAB was used to simulate and verify the correctness of models. Finally, this paper comes to the conclusion that normal trajectory and cutting trajectory are the optimum driving trajectories.

1. Introduction

Horizontal curves have been recognized as a significant safety issue for many years, and statistical analysis shows that the accidents in road curved sections are mostly head-on and lost control turning crashes. And the accident rate and accident severity in moderate curves are at their highest levels, which are 55% and 58%, respectively [1]. Great importance is therefore attached to data collections of driving behavior and research into factors that influence accident occurrence. Usually, the driving behavior within the curve areas is approached through the speed behavior, since speeds are fundamental for the design of road alignment. And an analysis of crashes associated with driving behaviors through curves, using the New Zealand Ministry of Transport's Crash Analysis System (CAS) database, generally supported this relationship between speed and increasing crash risk [1].

But one of the most difficult problems within the curve areas is the distinction between conscious and unconscious behaviors. Conscious behaviors includes traffic offenses and other driving processes which are not adapted to local or time conditions, and unconscious or unintentional behaviors are always conducted by lacking information. It is almost impossible to make the distinction between mention above with data collections in curves that are based only upon speed behavior.

Previous research findings on accidents and driving attempts revealed that, for some forms of curves, steering parallel to the axis of highway is difficult. That probably leads to uncertainties and steering corrections, which entail for their part increased centrifugal acceleration values. As early as 1980s, According to AGVS's earlier investigations into visual guidance for night driving, Federal Traffic Safety Working Group worked on the different paths of vehicles along curves in his early study, and most of which were inconsistent with tracks based on road design [2]. And other researchers, such as Glennon and Weaver in Tennessee [3], Prof. Xiao-Duan and Dean in Louisiana [4], Spacek in Switzerland [5–8], and so forth, studied the transverse distances of vehicles from the edge of the pavement of curves, and also come to the conclusion that tracks along curves are the major effect factors in leading to accident in curves.

In theory of driving trajectory, As far back as the 1970s in United States, some researches about drivers desires-related design consistency problems have been started, until 90s the self-explaining roads (SER) was proposed [9]. In the 2007, DVM simulates the vehicle's steering, braking, and acceleration processes through the driver's perception, cognition, control behavior, and so on to evaluate the speed and trajectory of the vehicle [10].

In the aspect of driving behavior theoretical models, From 1938, Gibson and Crooks proposed the theory of areas of vehicle traffic analysis (field-analysis) to now; as many as 20 kinds of driving behavior theoretical models were put forward by scholars from various countries, such as Prof. Guan et al. [11, 12], Xiao-Dong et al. [13], Wang et al. [14, 15], and so forth. In 2011, the research team of Prof. Wuhong Wang went deeply into discussing the desired trajectory and established the desired trajectory models, respectively. Combining with geometer parameter of China's mountain roads and driving characteristics [16]. Although the models were verified by simulation results, it is too idealized to be applied in practical production.

Against this background, there are three aims in this study: firstly, it is essential to obtain the observed data and determine typical patterns of track behavior at curves. Secondly, each model of track behavior should be set up. And thirdly, it would arrive at an optimum track behavior according to contrasting with the yaw stability of vehicles under the condition of different track behaviors. Finally, the proposed attempt will be made to provide theoretical support and develop suitable countermeasures to the reasonable optimization of widen curves, design of alignment, and the management of counter flow conflicts.

2. Typical Patterns and Mutability Characteristics of Track Behavior in Curves

2.1. Concept of the Vehicle Trajectory

The vehicle trajectories (paths or tracks) could be classified into two types: the expected trajectory and driving trajectory.

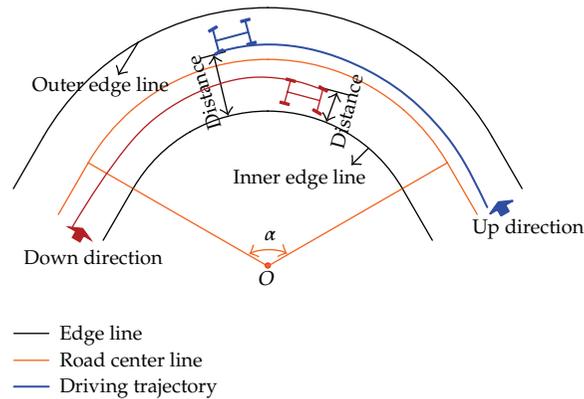


Figure 1: Schematic diagram of driving direction of vehicle trajectory.

Correspondingly, the driving trajectory along the curve (driving trajectory for short) is a result of expected trajectory with numerous adjustments by drivers based on the current road traffic condition, which is usually subjected to road factors, velocity factors, drivers factors themselves, and vehicle factors. Therefore, the mechanism of vehicle trajectory formation is very complicated.

In this paper, the vehicle trajectory related to the design of highway alignment and driving safety has the following characteristics:

- (1) it must be investigated under the condition of free volume;
- (2) it must be the observation trajectory;
- (3) it must be the real trajectory operated by drivers;
- (4) it must be related to the vehicle type selected.

Therefore, it can indicate the adaptability of drivers to road traffic condition, and its deviation can indicate the inconsistent between expectation of drivers and highway alignment design.

2.2. Definition of Objects

To avoid ambiguous statements, involved key objects in this paper were defined as follows:

- (i) Curved section: mainly refers to curved sections of two-way highway road, especially in rural, without physical separation in the middle, namely, mainly curved road sections of roads are of secondary level or below.
- (ii) Vehicle trajectory: when the vehicle runs into the curved section, the route passed by the left front wheel is defined as vehicle path or trajectory.
- (iii) Driving direction: the direction driving along the outer side of curved section is defined as up direction, while that driving along the inner side of curved section is defined as down direction, which can be seen in Figure 1.

Figure 2 shows the principle at one cross-section. The lane width in the direction of driving in the cross-sections at the main curve points and at the limits of the two adjacent

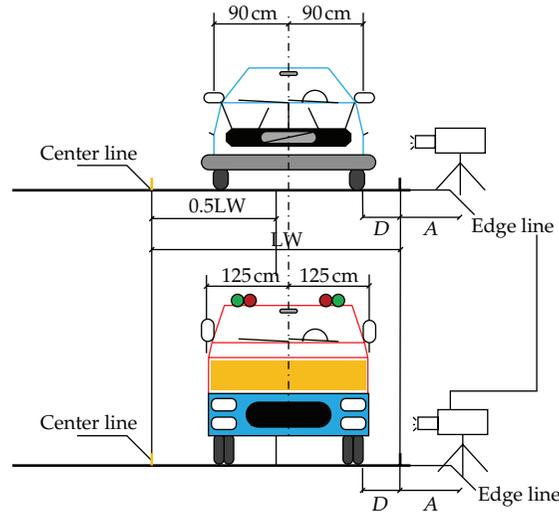


Figure 2: Measurement schematic diagram of vehicle trajectory. *A: distance of camera to the road edge line. *D*: distance of vehicle left wheel to the road edge line.

straight sections proved to be suitable references. At these points, the program checks the classification criteria.

For each cross-section, a bracket is given, within which the axis of the pass-through vehicle is to occur. A track path is assigned to a certain type only if it passes within the given brackets in all seven cross-sections.

2.3. Typical Patterns of Vehicle Track Behavior at Curves

One of the main purposes of this investigation was to define suitable track types that could be observed. Usually, the typical patterns of track behaviors can be defined as “basic trajectory” and “extreme trajectory” according to the deviation degree of driving trajectory. Basic trajectory usually refers to the paths of vehicles which is always consistent with the ideal trajectory based on road design, and extreme trajectory usually refers to which deviate seriously. According to the observation data and on the basis of test measurements, the following six track types were defined in this paper, and the experimental data and analytical assumptions can be found in reference [17].

- (i) Ideal behavior: this corresponds to a symmetrical track path formed by the vehicle left front wheel within a narrow area along the center of the road. Such idealized track path is assumed in the design standards.
- (ii) Normal behavior: this likewise indicates a symmetrical track path formed by vehicle left front wheel along the center of road, but within a somewhat broader area than with the ideal behavior and with slight cutting to the road center line.
- (iii) Cutting behavior: this is a track path with strong cutting to the road center line within the area of the circular arc (conscious driving process to balance centrifugal acceleration).

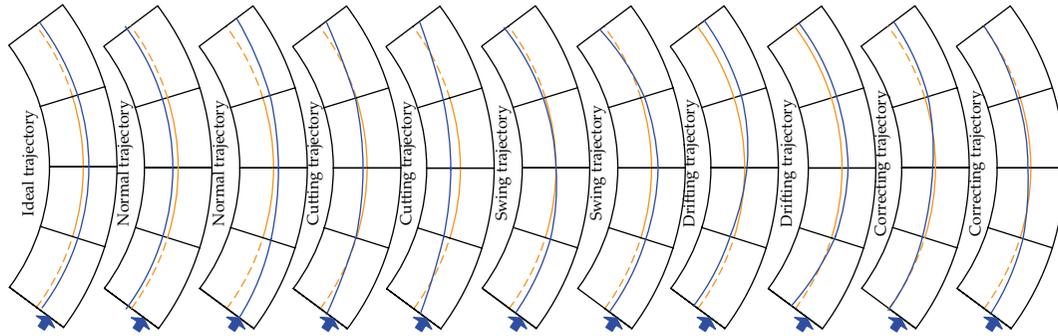


Figure 3: Sketches of track types (example for an up direction).

Table 1: Distribution probability of track types (%) at investigated curves.

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
P	0.01	0.10	0.11	0.04	0.21	0.12	0.05	0.07	0.09	0.04	0.16

- (iv) Drifting behavior: this is asymmetrical track path between the beginning and end of the curve with a pronounced tendency to drive on the down direction at the beginning of the curve and an increasing drift to the up direction the end of the curve at left-hand curves, analogous at right-hand curves.
- (v) Swinging behavior: this is asymmetrical track path between the beginning and end of the curve with a pronounced tendency to drive on the up direction at the beginning of the curve and an increasing drift to the down direction the end of the curve at left-hand curves, analogous at right-hand curves.
- (vi) Correcting behavior: this is a track path with increased drifting toward the inside/ outside of the curve and subsequent correction of the steering angle in the second half of the curve. It is assumed that this track type is a kind of unconscious driving behavior, due to underestimation of the curvature and/or the length of the curve.

Track paths that could not be assigned to those defined types are summarized in a separate group called “remaining” track paths.

The six paths of the track types described above are illustrated in Figure 3 by the example of an up direction. And Table 1 gives a distribution probability of the track types per measurement direction in the curves investigated.

3. Modeling of Driving Trajectory

3.1. Primes Supposition and Modeling Thought

Track path is influenced by many factors in actual. To reveal the formation mechanism of driving trajectory and simplify analysis process, assumptions are made as follows:

- (i) the study object is dual two-lane road bend section under small volume traffic, which without physical isolation;
- (ii) ignore the influence of vertical curve and transverse ultra high, road section is ideal horizontal curve;

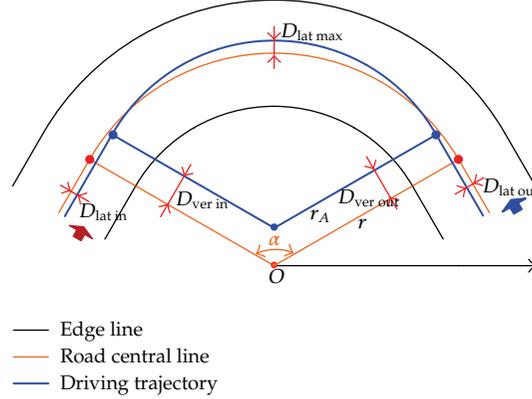


Figure 4: Schematic diagram coordinate system of vehicle trajectory.

- (iii) suppose ideal bend section is composed by straight-line section and circular curve section, transition curve is ignored;
- (iv) suppose vehicle drives with constant steering angle and speed, vehicle's longitudinal acceleration is ignored;
- (v) driver expect to pass the bend section with small curvature, and the steering angle of driving trajectory is the same with that of road circular curve;
- (vi) driving trajectory's deviation from the road centerline satisfies the maximum deviation limitation.

Based on the above assumptions, vehicle driving trajectory while driving on road curved section can be expressed by circular curve. To describe vehicle driving trajectory conveniently, polar coordinate system is constructed as Figure 4. The center of circular curve of road centerline is pole O ; horizontal ray passes pole and points to right is polar axis ox ; polar angle rotates in counterclockwise direction is positive. In this polar coordinate system, radius of circular curve of road centerline is r , radius of vehicle driving trajectory is r_A , radius of vehicle expected trajectory is r_E , and steering angle of road curved section is α .

Under most conditions, the center of vehicle driving trajectory does not coincide with that of road centerline. To describe driving trajectory conveniently, make the center of vehicle driving trajectory as relative pole to establish polar coordinates of driving trajectory.

Known by actual observation that there are lateral and longitudinal offsets between the endpoints of vehicle driving trajectory and that of road centerline. Longitudinal offset includes advance, parallel and lag; lateral offset includes medial offset, lateral offset and zero offset. The number of cross-point of driving trajectory and road centerline may be 0, 1 or 2. The definition of lateral and longitudinal offsets is as follows:

- (1) $D_{lat\ max}$: the maximum lateral offset between driving trajectory and road centerline.
- (2) $D_{lat\ in}$: lateral offset between the start point of driving trajectory and that of road centerline.
- (3) $D_{lat\ out}$: lateral offset between terminal point of driving trajectory and that of road centerline.
- (4) $D_{ver\ in}$: longitudinal offset between start point of driving trajectory and that of road centerline.

- (5) $D_{\text{ver out}}$: longitudinal offset between terminal point of driving trajectory and that of road centerline.

3.2. Characteristic Analysis and Model Establishment of Basic Trajectory

As driving trajectory parallels to road center or not, basic trajectory can be divided to ideal trajectory and normal trajectory.

3.2.1. Ideal Trajectory

Ideal trajectory is presented in Figure 5. Vehicle always drives along the left side of road centerline; driving trajectory parallels to road centerline; endpoints of driving trajectory and those of road centerline do not have longitudinal offset, compared to the curvature of road centerline, that of driving trajectory increases.

Radius of ideal trajectory can be expressed by

$$r_A = r + D_{\text{lat max}}. \quad (3.1)$$

And polar coordinates of ideal trajectory can be written by

$$f(\rho, \theta) = \begin{cases} \rho = r + D_{\text{lat max}}, \\ \theta = \left[\frac{\pi - \alpha}{2}, \frac{\pi + \alpha}{2} \right]. \end{cases} \quad (3.2)$$

Coordinate of relative pole O' which coincides with pole O is $(0, 0)$.

3.2.2. Normal Trajectory

Normal trajectory is presented in Figure 6. Vehicle always drives along the right side of road centerline; driving trajectory does not parallel to road centerline and there is no cross-point; compared to road centerline, driving trajectory's start point advances and terminal point lags, its curvature decreases.

Radius of normal trajectory can be expressed by

$$r_A = \frac{r + D_{\text{lat in}} - D_{\text{lat max}} \cos(\alpha/2)}{1 - \cos(\alpha/2)}, \quad (3.3)$$

where

$$\begin{aligned} D_{\text{ver in}} &= D_{\text{ver out}}, & D_{\text{lat in}} &= D_{\text{lat out}}, \\ D_{\text{lat max}} &\in [0, D_{\text{lat in}}], \\ D_{\text{ver in}} &= (D_{\text{lat in}} - D_{\text{lat max}}) \text{ctg} \frac{\alpha}{4}. \end{aligned} \quad (3.4)$$

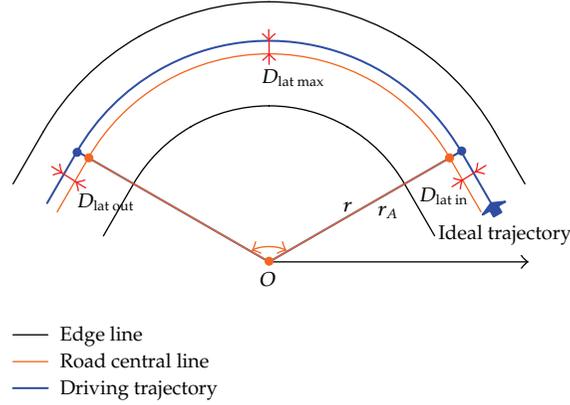


Figure 5: Characteristic analysis diagram of ideal trajectory.

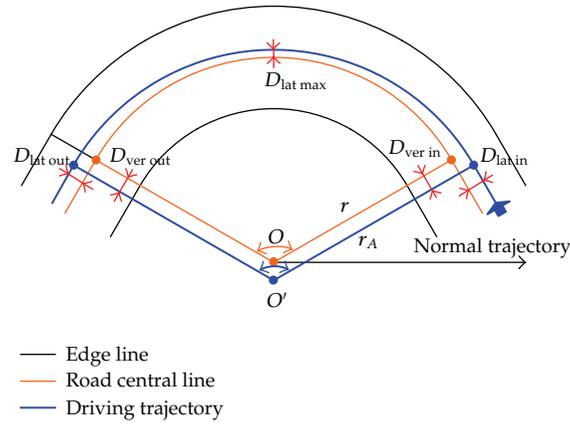


Figure 6: Characteristic analysis diagram of normal trajectory.

And polar coordinates of normal trajectory can be written by

$$f(\rho, \theta) = \begin{cases} \rho = r + \frac{D_{\text{lat in}} - D_{\text{lat max}} \cos(\alpha/2)}{1 - \cos(\alpha/2)}, \\ \theta = \left[\frac{\pi - \alpha}{2}, \frac{\pi + \alpha}{2} \right]. \end{cases} \quad (3.5)$$

Coordinate of relative pole O' is defined as:

$$\left(\frac{D_{\text{lat max}} - D_{\text{lat in}}}{\cos(\alpha/2) - 1}, \frac{3\pi}{2} \right). \quad (3.6)$$

3.3. Characteristic Analysis and Model Establishment of Extreme Trajectory

According to driver's consciousness, extreme trajectory can be divided to conscious extreme trajectory and unconscious extreme trajectory.

Conscious extreme trajectory is the one that driver deliberately chooses to achieve some purposes. At this time, driver has a certain capability and psychological preparation to bear risk. Cutting trajectory is one of the popular conscious extreme trajectories, is the resulting of driver deliberately deviate road design trajectory, and occupies the opposite direction lane to reduce lateral acceleration or keep high speed. This kind of trajectory usually appears on small radius curved section with small steering angle.

Unconscious extreme trajectory is the resulting as driver underestimates the curvature of road curved section and chooses this trajectory unconsciously. While driver discovered the trend to this trajectory, the general response is to turn abruptly to correct vehicle driving trajectory. Swinging and drifting trajectories are the popular ones of unconscious extreme trajectories. Although driver's purpose is to reduce vehicle's lateral acceleration, but large amount of survey found that the two trajectories will lead to lateral acceleration's sudden increases in partial part. High driving speed and braking operation are apt to cause vehicle out of control.

3.3.1. Cutting Trajectory

Cutting trajectory is presented in Figure 7. Vehicle drives along the right side of road center when it enters and pulls out the curved section, drives along the left side of road centerline near the middle of curved section. Compared to road centerline, cutting trajectory's start point advances, terminal point lags, and curvature decreases.

Radius of cutting trajectory can be expressed by

$$r_A = r + \frac{D_{\text{lat max}} \cdot \cos(\alpha/2) + D_{\text{lat in}}}{1 - \cos(\alpha/2)}, \quad (3.7)$$

where

$$\begin{aligned} D_{\text{lat in}} &\in [0, D_{\text{lat max}}], \\ D_{\text{ver in}} &= (D_{\text{lat max}} + D_{\text{lat in}}) \text{ctg} \frac{\alpha}{4}. \end{aligned} \quad (3.8)$$

And coordinate of cutting trajectory can be written by

$$f(\rho, \theta) = \begin{cases} \rho = r + \frac{D_{\text{lat max}} \cdot \cos(\alpha/2) + D_{\text{lat in}}}{1 - \cos(\alpha/2)}, \\ \theta = \left[\frac{\pi - \alpha}{2}, \frac{\pi + \alpha}{2} \right]. \end{cases} \quad (3.9)$$

Coordinate of relative pole O' is defined as

$$\left(\frac{D_{\text{lat max}} + D_{\text{lat in}}}{1 - \cos(\alpha/2)}, \frac{3\pi}{2} \right). \quad (3.10)$$

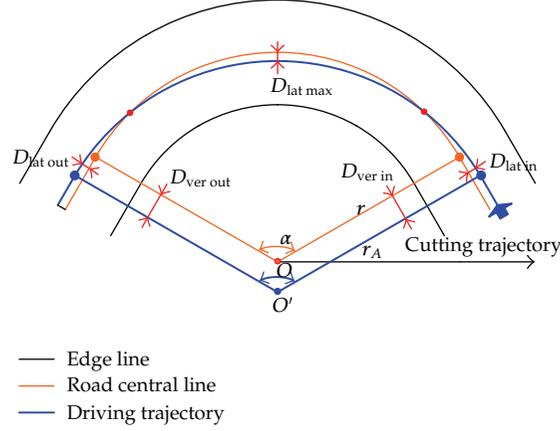


Figure 7: Characteristic analysis diagram of cutting trajectory.

3.3.2. Swinging Trajectory

Swinging trajectory is shown in Figure 8. Vehicle drives along the right side of road center when it enters the curved section and along the left side of road centerline near the middle of curved section. Near the middle of curved section, vehicle driving trajectory locates in uncertain position. Compared to road centerline, cutting trajectory's start point advances, terminal point lags, and curvature decreases.

Radius of swinging trajectory can be expressed by

$$r_A = r + \frac{D_{\text{lat in}}^2 + D_{\text{ver in}}^2 - D_{\text{lat out}}^2 - D_{\text{ver out}}^2}{2(D_{\text{lat in}} + D_{\text{lat out}})}, \quad (3.11)$$

where

$$\begin{aligned} D_{\text{lat in}} &\leq D_{\text{lat max}}, \\ D_{\text{lat out}} &\leq D_{\text{lat max}}, \\ D_{\text{ver in}} &= \frac{D_{\text{lat max}} + D_{\text{lat in}}}{1 - \cos((\alpha/2) + \angle 1)} \sin\left(\frac{\alpha}{2} + \angle 1\right), \\ D_{\text{ver out}} &= \frac{D_{\text{lat max}} - D_{\text{lat out}}}{1 - \cos((\alpha/2) - \angle 1)} \sin\left(\frac{\alpha}{2} - \angle 1\right), \\ \angle 1 &\leq \min\left\{\frac{\pi - \alpha}{2}, \frac{\alpha}{2}\right\}. \end{aligned} \quad (3.12)$$

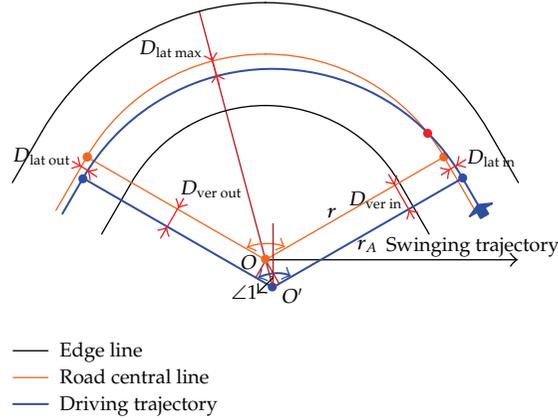


Figure 8: Characteristic analysis diagram of swinging trajectory.

And polar coordinates of swinging trajectory can be written by

$$f(\rho, \theta) = \begin{cases} \rho = r + \frac{D_{\text{lat in}}^2 + D_{\text{ver in}}^2 - D_{\text{lat out}}^2 - D_{\text{ver out}}^2}{2(D_{\text{lat in}} + D_{\text{lat out}})}, \\ \theta = \left[\frac{\pi - \alpha}{2}, \frac{\pi + \alpha}{2} \right]. \end{cases} \quad (3.13)$$

Coordinate of relative pole O' is defined as

$$\left(\frac{D_{\text{ver in}}}{\sin((\alpha/2) + \angle 1)}, \frac{3\pi}{2} + \alpha \right). \quad (3.14)$$

3.3.3. Drifting Trajectory

Drifting trajectory is shown in Figure 9. Vehicle drives along the left side of road center when it enters the curved section and along the right side of road centerline near the middle of curved section. Near the middle of curved section, vehicle driving trajectory locates in uncertain position. Compared to road centerline, curvature of drifting trajectory decreases.

Radius of drifting trajectory can be expressed by

$$r_A = r + \frac{D_{\text{ver in}}^2 - D_{\text{lat in}}^2 + D_{\text{lat out}}^2 - D_{\text{ver out}}^2}{2(D_{\text{ver in}} + D_{\text{lat out}})}, \quad (3.15)$$

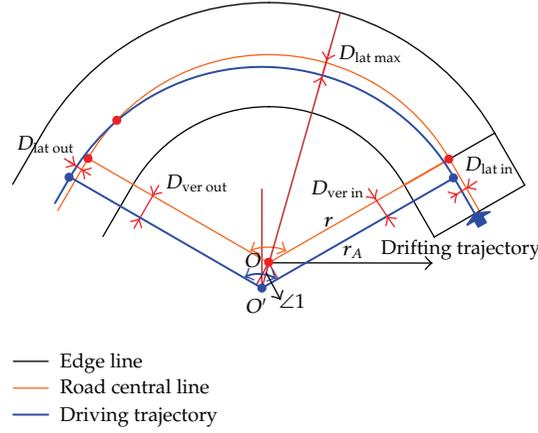


Figure 9: Characteristic analysis diagram of drifting trajectory.

where

$$\begin{aligned}
 D_{\text{lat in}} &\leq D_{\text{lat max}}, \\
 D_{\text{lat out}} &\leq D_{\text{lat max}}, \\
 D_{\text{ver in}} &= \frac{D_{\text{lat in}} - D_{\text{lat max}}}{\cos((\alpha/2) - \zeta 1) - 1} \sin\left(\frac{\alpha}{2} - \zeta 1\right), \\
 D_{\text{ver out}} &= \frac{D_{\text{lat max}} + D_{\text{lat out}}}{1 - \cos((\alpha/2) + \zeta 1)} \sin\left(\frac{\alpha}{2} + \zeta 1\right), \\
 \zeta 1 &\leq \min\left\{\frac{\pi - \alpha}{2}, \frac{\alpha}{2}\right\}.
 \end{aligned} \tag{3.16}$$

And polar coordinates of drifting trajectory can be written by

$$f(\rho, \theta) = \begin{cases} \rho = r + \frac{D_{\text{ver in}}^2 - D_{\text{lat in}}^2 + D_{\text{lat out}}^2 - D_{\text{ver out}}^2}{2(D_{\text{ver in}} + D_{\text{lat out}})}, \\ \theta = \left[\frac{\pi - \alpha}{2}, \frac{\pi + \alpha}{2}\right]. \end{cases} \tag{3.17}$$

Coordinate of relative pole O' is defined as

$$\left(\frac{D_{\text{lat in}} - D_{\text{lat max}}}{\cos((\alpha/2) - \zeta 1) - 1}, \frac{3\pi}{2} - \alpha\right). \tag{3.18}$$

3.3.4. Correcting Trajectory

Correcting trajectory is shown in Figure 10. Vehicle drives near the road centerline when it enters and pulls out the curved section. There are no cross-point between driving trajectory

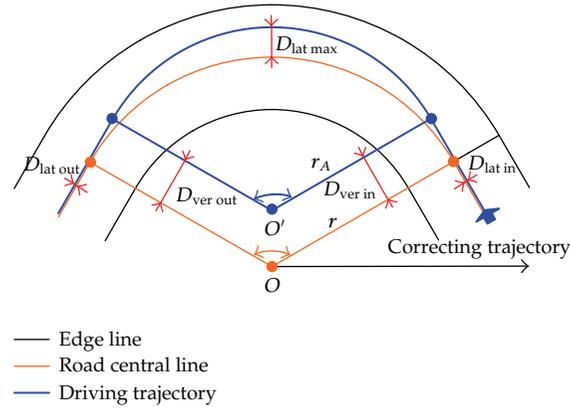


Figure 10: Characteristic analysis diagram of correcting trajectory.

and road centerline. Compared to road centerline, correcting trajectory's start point lags and terminal point advances, and curvature increases.

Radius of correcting trajectory can be expressed by

$$r_A = r - \frac{D_{\text{lat max}} \cos(\alpha/2) - D_{\text{lat in}}}{1 - \cos(\alpha/2)}, \quad (3.19)$$

where

$$D_{\text{ver in}} = \frac{D_{\text{lat max}} - D_{\text{lat in}}}{1 - \cos(\alpha/2)} \sin \frac{\alpha}{2}. \quad (3.20)$$

And polar coordinates of correcting trajectory can be written by:

$$f(\rho, \theta) = \begin{cases} \rho = r - \frac{D_{\text{lat max}} \cos(\alpha/2) - D_{\text{lat in}}}{1 - \cos(\alpha/2)}, \\ \theta = \left[\frac{\pi - \alpha}{2}, \frac{\pi + \alpha}{2} \right]. \end{cases} \quad (3.21)$$

Coordinate of relative pole O' is defined as

$$\left(\frac{D_{\text{lat max}} - D_{\text{lat in}}}{1 - \cos(\alpha/2)}, \frac{\pi}{2} \right). \quad (3.22)$$

4. Simulation and Verification

The six developed models are applied to simulate driving track behaviors at curves in order to study its rationality. The simulation diagrams with different corner (120° and 60°) can be shown in Figures 11 and 12 the parameters in models are varied between limits except radius. It can be seen that the models are verified by digital simulation results.

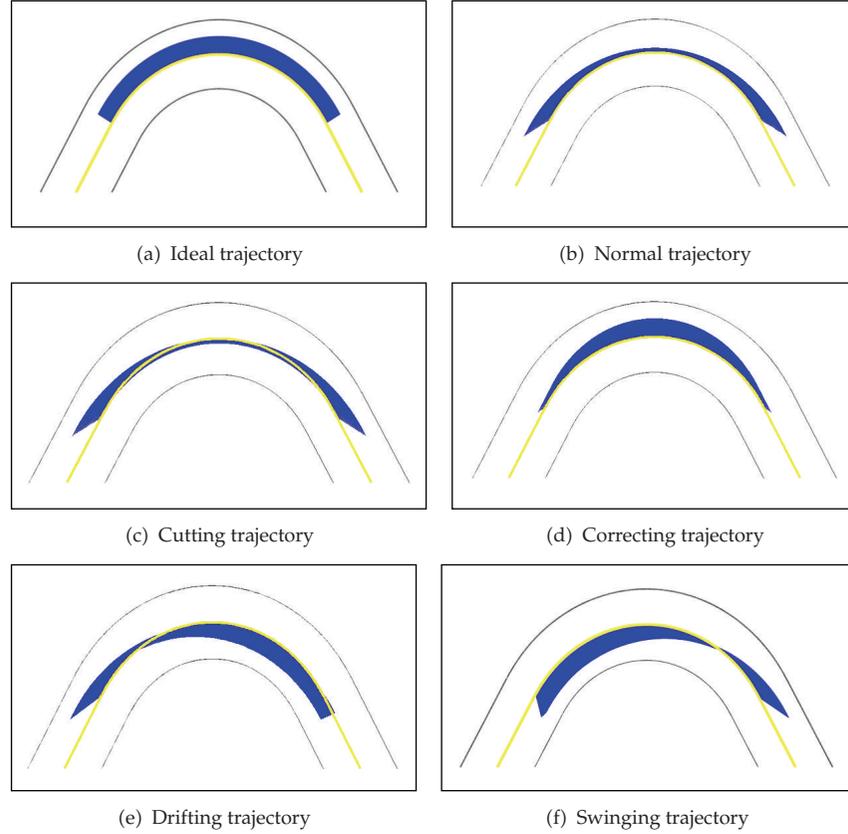


Figure 11: Simulated diagram with the corner angel being 120°.

5. Optimized Trajectory and Vehicle Yaw Stability

2-degree-of-freedom vehicle dynamic model is used to analyze vehicle yaw rate under different driving trajectories, which is presented as follows:

$$\begin{aligned}
 mV(\dot{\beta} + r) &= -2k_f \left(\beta + \frac{l_f}{V}r - \delta \right) - 2k_r \left(\beta - \frac{l_r}{V}r \right), \\
 I_z \dot{r} &= -2k_f \left(\beta + \frac{l_f}{V}r - \delta \right) l_f + 2k_r \left(\beta - \frac{l_r}{V}r \right) l_r,
 \end{aligned} \tag{5.1}$$

where m is vehicle total mass, V is driving speed, β is sideslip angle, r is yaw rate. k_f , k_r are front and rear tires' cornering stiffness, respectively. l_f , l_r are front and rear track, respectively. δ is steer angle. I_z is yaw moment of inertia of total mass.

Matlab is used to simulate vehicle time varying characteristics numerically. Vehicle yaw rates under normal and ideal trajectory are presented in Figure 13.

As seen from Figure 13, vehicle yaw rate under normal trajectory is 11.11% smaller than that under ideal trajectory, which indicates vehicle steering stability under normal trajectory is better than that under ideal trajectory.

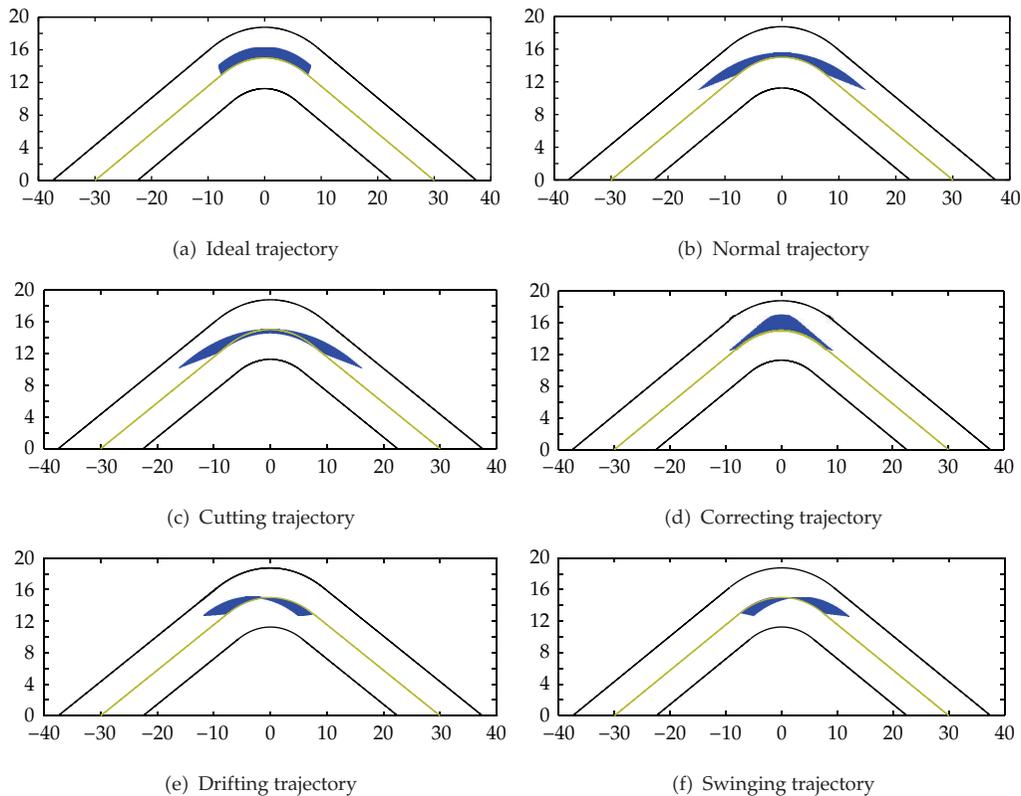


Figure 12: Simulated diagram with the corner angel being 60° .

Vehicle yaw rates under four kinds of extreme trajectories are presented in Figure 14. As seen from Figure 14, yaw rate under swing trajectory is the biggest and yaw rate under cutting trajectory is the smallest, which indicates that vehicle has the best steering stability under cutting trajectories while it has the worst steering stability under swing trajectory.

According to the analysis above, compared to other driving trajectories, normal and cutting trajectory are the safer driving trajectories for vehicle passing by the curves.

6. Conclusions

In contrast to the usual descriptions of driving behavior in curve areas in terms of speeds, this paper investigated track behavior. To this point, a classification of the driving processes according to the type of the track paths along curves was developed and the trajectory models were established, respectively, combining with curve geometry of two-way highway road. At last, 2-degree-of-freedom vehicle dynamic model was used to analyze vehicle yaw rate under different type of trajectories, considering the sight distance, normal trajectory and cutting trajectory were determined as the optimum driving trajectories. The study findings have great importance to providing theoretical support and developing suitable countermeasures to the reasonable optimization of widen curves, design of alignment, and the management of counter flow conflicts.

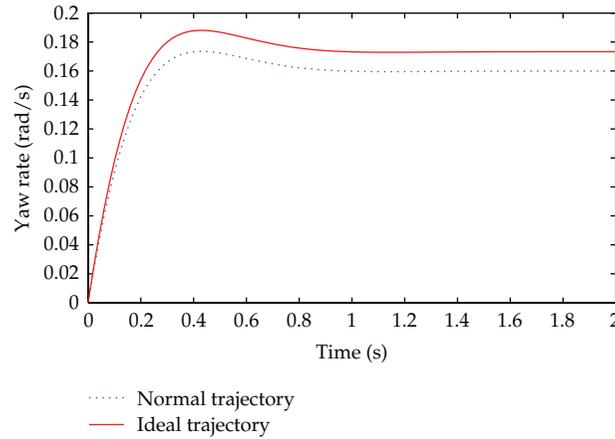


Figure 13: Vehicle yaw rates under two kinds of basic trajectory.

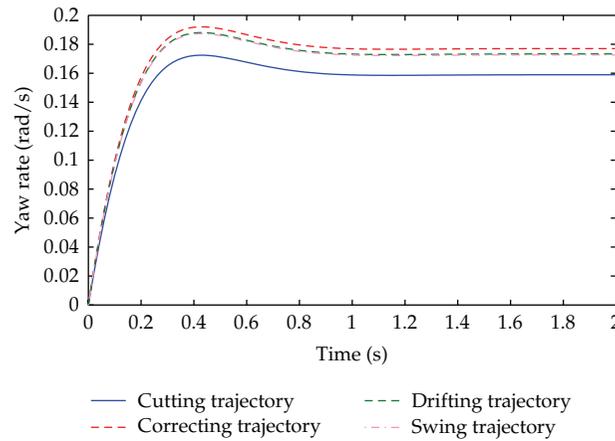


Figure 14: Vehicle yaw rates under four kinds of extreme trajectories.

Since we ignored the transition curves in deriving the vehicle track models for simplification purpose, thus, what the vehicle track models really are when the transition curves are considered will be conducted in a future study.

Acknowledgment

This study is supported by the National Natural Science Foundation of China (nos. 50978114 and 51208225).

References

- [1] S. G. Charlton and J. J. D. Pont, "Curved speed management," Research Report 323, Land Transport New Zealand, Wellington, New Zealand, 2007.
- [2] Arbeitsgruppe Verkehrssicherheit (AGVS) and Federal Traffic Safety Working Group, *Traffic Safety by Night, Traffic-Engineering-Related Improvements to the Highway Infrastructure*, EJPD, Berne, Switzerland, 1980.

- [3] J. C. Glennon and G. D. Weaver, "The relationship of vehicle paths to highway curve design," Research Report 134-6, Texas Highway Department and the Federal Highway Administration, 1971.
- [4] S. Xiao-Duan and T. Dean, "Impact of edge lines on safety of rural two-lane highways," Tech. Rep. 414, Civil Engineering Department University of Louisiana at Lafayette, Lafayette, La, USA, 2005.
- [5] U. Scheifele and P. Spacek, *Measuring Poles: A Measuring Device for Surveying Driving Behavior on Highways*, IVT-ETH Zurich and Planitronic Zurich, supported by the Swiss Road Safety Fund, Zurich, Switzerland, 1992.
- [6] P. Spacek, "Driving behavior and accident occurrence in curves, driving behavior in curve areas," Research Project 16/84, IVT-ETH, Swiss Federal Highways Office, Zurich, Switzerland, 1998.
- [7] P. Spacek, *Environmental Impact of Traffic, Section: Road Safety of Traffic Facilities*, Lecture Notes, IVT-ETH, Zurich, Switzerland, 2003.
- [8] P. Spacek, "Track behavior in curve areas: attempt at typology," *Journal of Transportation Engineering*, vol. 131, no. 9, pp. 669–676, 2005.
- [9] Federal Highway Administration, *Development of a Driver Vehicle Module for the Interactive Highway Safety Design Model*, United States Department of Transportation, Washington, DC, USA, 2007.
- [10] M. Althoff, O. Stursberg, and M. Buss, "Safety assessment of driving behavior in multi-lane traffic for autonomous vehicles," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 893–900, June 2009.
- [11] X. Guan, Z. Gao, and K. Guo, "Driver fuzzy decision model of vehicle preview course and simulation under typical road conditions," *Automotive Engineering*, vol. 23, no. 1, pp. 13–20, 2001 (Chinese).
- [12] G. Zhen-Hai, G. Xin, and G. Kong-Hui, "Driver fuzzy decision model of vehicle preview course," *Natureal Science Journal of Jilin University of Technology*, vol. 30, no. 1, pp. 7–10, 2000 (Chinese).
- [13] P. Xiao-Dong, D. Zhi-Gang, J. Hong et al., "Experiment research on relationship between the variation of drivers' heart rate and systolic blood pressure and a lignment of mountainous highway," *Chinese Journal of Ergonomics*, vol. 12, no. 2, pp. 16–18, 2006 (Chinese).
- [14] W. H. Wang, C. X. Ding, G. D. Feng, and X. B. Jiang, "Simulation modelling of longitudinal safety spacing in inter-vehicles dynamics interactions," *Journal of Beijing Institute of Technology*, vol. 19, supplement 2, pp. 55–60, 2010.
- [15] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: A fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.
- [16] W. Wuhong, G. Weiwei, M. Yan et al., "Model-based simulation of driver expectation in mountainous road using various control strategies," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 1187–1194, 2011.
- [17] R. Yuan-Yuan, L. Xian-Sheng, R. You, G. Rachel, and G. Wei-Wei, "Study on driving dangerous area in road curved section based on vehicle track characteristics," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1237–1245, 2011.

Research Article

Traffic Congestion Evaluation and Signal Control Optimization Based on Wireless Sensor Networks: Model and Algorithms

Wei Zhang, Guozhen Tan, Nan Ding, and Guangyuan Wang

School of Computer Science and Technology, Dalian University of Technology, Dalian 116023, China

Correspondence should be addressed to Guozhen Tan, gztan@dlut.edu.cn

Received 15 June 2012; Accepted 14 November 2012

Academic Editor: Geert Wets

Copyright © 2012 Wei 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.

This paper presents the model and algorithms for traffic flow data monitoring and optimal traffic light control based on wireless sensor networks. Given the scenario that sensor nodes are sparsely deployed along the segments between signalized intersections, an analytical model is built using continuum traffic equation and develops the method to estimate traffic parameter with the scattered sensor data. Based on the traffic data and principle of traffic congestion formation, we introduce the congestion factor which can be used to evaluate the real-time traffic congestion status along the segment and to predict the subcritical state of traffic jams. The result is expected to support the timing phase optimization of traffic light control for the purpose of avoiding traffic congestion before its formation. We simulate the traffic monitoring based on the *Mobile Century* dataset and analyze the performance of traffic light control on VISSIM platform when congestion factor is introduced into the signal timing optimization model. The simulation result shows that this method can improve the spatial-temporal resolution of traffic data monitoring and evaluate traffic congestion status with high precision. It is helpful to remarkably alleviate urban traffic congestion and decrease the average traffic delays and maximum queue length.

1. Introduction

The traffic crowds seen in intersection of urban road networks are highly influential in both developed and developing nations worldwide [1]. Urban residents are suffering poor transport facilities, and meanwhile the considerable financial loss caused by traffic becomes a large and growing burden on the nation's economy, including costs of productivity losses from traffic delays, traffic accidents, vehicular collisions associated with traffic jams, higher emission, environmental pollution, and more. The idea that the improvements to transport infrastructure are the efficient way has been central to transport economic analysis, but in fact

this problem cannot be resolved with better roads [2–4]. Intelligent transportation systems (ITS) have been proven to be a scientific and efficient solution [5]. Comprehensive utilization of information technology, transportation engineering and behavioral sciences to reveal the principle of urban traffic, measuring the traffic flow in real time, and try to route vehicles around them to avoid traffic congestion before its formation promotes a prospective solution to resolve the urban traffic problem from the root [5–7].

Nowadays, in an information-rich era, the traditional traffic surveillance and control methods are confronted with great challenges [8, 9]. How to get meaningful information from large amounts of sensor data to support transportation applications becomes more and more significant [6, 10]. Modern traffic control and guidance systems are always networked in large scale which need real time, traffic data with higher spatial-temporal resolution that challenges the traditional traffic monitoring technologies such as inductive loop, video camera, microwave radar, infrared detector, UAV, satellite image, and GPS [11]. The state-of-the-art, intelligent, and networked sensors are emerging as a novel network paradigm of primary relevance, which provides an appealing alternative to traditional traffic surveillance approaches in near future [12], especially for proactively gathering monitoring information in urban environments under the grand prospective of cyber physical systems [13, 14]. Wireless sensors have many distinctive advantages such as low cost, small size, wireless communication, and distributed computation. Over the last decade, many researchers have endeavored to study traffic monitoring with novel technologies, and recent research shows that the tracking and identification of vehicles with wireless sensor networks for the purpose of traffic surveillance and control are widespread applications [15–19].

Traffic research currently still cannot fully express the intrinsic principle of traffic congestion formation and predict under which conditions traffic jam may suddenly occur. In the essentials, urban traffic is a typical self-driven many-particle huge system which is far from equilibrium state, where the traffic flow is a complicated nonlinear dynamic process, and the traffic congestion is the spatial-temporal conglomeration of traffic volume in finite time and space. In 2009, Flynn et al. have conducted some theoretical work to model traffic congestion with macroscopic traffic flow theory and obtained some basic results in congestion prediction [20], which are regarded as a creative solution of traffic equations proposed in 1950s and reported as a great step towards answering the key question that is how can the occurrence of traffic congestion be avoided. Based on current research, the congestion status of traffic flow is expected to be evaluated in real time and higher precision to support traffic control.

Traffic light control at urban intersection can be modeled as a multiobjective optimization problem (MOP). In UTCS (Urban Traffic Control System) such as SCOOT/SCATS/REHODES system, it always employs single loop sensor or double loops as vehicle detector deployed at upstream of the signalized intersections. Generally, in current traffic control strategies, optimization objectives include stop of vehicle, average delay, travel time, queuing length, traffic volume, vehicle speed, and even exhaust emission [21]. The traditional traffic detection is *Eulerian sensing* which collects data at predefined locations [22], and the sensors cannot be deployed in large amount as compared to the huge scale of urban road networks for sake of budget restriction and maintenance cost; as a result the data such as vehicle stops and delays of individual's vehicle is difficult to be achieved accurately. In the essentials, comparing to existing criteria mentioned above, the traffic congestion is a directly relevant factor and the root reason. Introducing a method to evaluate the degree of traffic congestion and proposing into the optimization model of traffic light control promote a feasible approach to improve traffic control performance.

In this paper, we studied the intrinsic space-time properties of actual traffic flow at the intersection and near segments and build an observation system to estimate and collect traffic parameters based on sparsely deployed wireless sensor networks. We are interested in understanding how to evaluate and express the degree of traffic congestion quantitatively and what the performance for traffic signal control would be if we take into account the traffic congestion factor as one of the objectives in timing optimization.

The rest of the paper is organized as follows. The current studies on traffic surveillance with wireless sensor networks are briefly reviewed in Section 2. The observation model based on traffic flow theory and traffic flow parameters estimation algorithm based on wireless sensor networks are described in detail in Section 3. The traffic congestion evaluation model and congestion factor based signal phrase optimization algorithms are discussed in Section 4. The performance is analyzed based on simulation and experimental results in Section 5. Finally, a conclusion and future works are given in Section 6.

2. Related Works and Problem Statement

Several research works on traffic monitoring with wireless sensor networks have been carried out in recent years. Most of them have focused on individual vehicle and point data detection, where the traffic spatial-temporal property is not an issue in these circumstances. Pravin et al. creatively applied the magnetic sensor networks to vehicle detection and classification in Berkeley PATH program from 2006 and obtained high precision beyond 95% [12, 23]. In 2008, UC Berkeley launched a pilot traffic-monitoring system named *Mobile Century* (successor project is known as *Mobile Millennium*) to collect traffic data based on the GPS sensor equipped in cellular phones [22]. They found that 2–5% point data provided by mobile sensors is sufficient to provide information for traffic light control, and their conclusion motivates the research to collect traffic data and control traffic flow via sparsely deployed sensor networks in this paper. Hull et al. studied the travel time estimation with Wi-Fi equipped mobile sensor networks [24, 25]. Bacon et al. developed an effective data compress and collection method based on sensor networks using the weekly spatial-temporal pattern of traffic flow data in TIME project [26]. But in current research there are some important aspects out of consideration. (1) Few considerations are given to the intrinsic space-time properties and operation regularity of actual traffic flow and traffic congestion formation. (2) How to evaluate traffic congestion quantitatively with sufficient precision and real-time performance and be introduced as an objective to support control optimization in traffic light control? (3) How to combine traffic surveillance sensor networks with traffic control system to analyze future traffic conditions under current timing strategies and try to avoid traffic congestion before its formation.

The discipline of transportation science has expanded significantly in recent decades, and particularly traffic flow theory plays a great role in intelligent transportation systems [27–29]. The typical models include LWR continuum model [30] and Payne-Whitham higher model [31]. From the physical view of traffic flow, the spatiotemporal behavior is the fundamental propriety in nature. In previous work, the vast majority of inductive techniques were focused on state-space methodology that forecasts short-term traffic flow based on historical data with relatively small number of measurement locations [32–34]. Limited amount of work has been performed using space-time model [35], and the resolution or precision is insufficient for the purpose of traffic light control. In 2008, Sugiyama et al. explained the formation process of traffic congestion by experimental observations [36], and

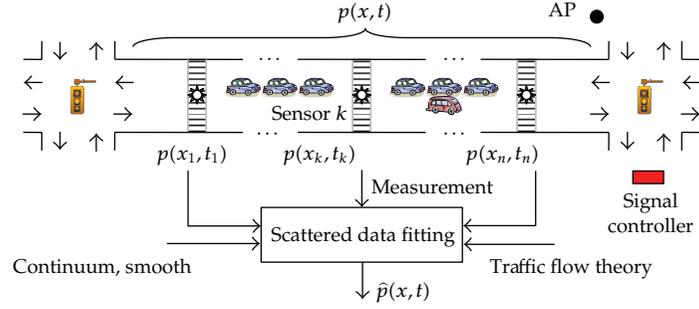


Figure 1: Deployment of wireless sensor networks for urban traffic surveillance.

based on this, Flynn et al. built a congestion model to explain and predict traffic congestion with macroscopic traffic flow theory in 2009 [20], which is regarded as a solution of traffic flow equations and a great step towards answering the key question that how can the occurrence of traffic congestion be avoided.

The goal of this paper is to estimate traffic parameters based on sparsely deployed sensor networks, evaluate the degree of traffic congestion, and obtain a quantitative factor to express the spatiotemporal properties of traffic flow in real time. Based on this, introduce the congestion factor to the optimization model of traffic light control. In this paper we use *Lagrangian* detection [37]. Not only detect point data via imperfect binary proximity sensor network [38], but also try to estimate the time-space properties along the road segment based on scattered sensor measurements. The deployment of sensor networks is shown in Figure 1, where $p(x, t)$ denotes traffic data such as velocity and density. Based on this, the congestion status and evaluation criteria can be studied from the comprehensive scope. The sensor network is expected to monitor real-time traffic data, to predict the subcritical state, and to control traffic signal to avoid the traffic jams before its formation.

The urban road network can be modeled as a directed graph consisting of vehicles $v \in V$ and edges $e \in E$. Let L_e be the length of edge e . The spatial and temporal variables are road segment $x \in [0, L_e]$ and time $t \in [0, +\infty)$, respectively. On a given road segment x_e and time t , the traffic flow speed $u(x, t)$ and density $\rho(x, t)$ are distributed parameter system in time and space. While vehicle labeled by $i \in N$ travels along the road segment with trajectory $x_i(t)$, the sensor measurements $u(x_i(t), t)$ and $\rho(x_i(t), t)$ are discrete and instant values, as shown in (2.1), and here k is the sensor node number. The problem of traffic flow information monitoring can be transformed to estimate traffic parameters in given spatial and temporal variables t with these discrete values (Nomenclature and symbols are available in Table 1):

$$U_t = (u_1, \dots, u_k)^T, \quad P_t = (\rho_1, \dots, \rho_k)^T. \quad (2.1)$$

3. Traffic Monitoring and Data Estimation

In this section, we firstly describe the intrinsic characteristic of traffic flow and then propose a method to estimate traffic parameters based on scattered data collected by sparsely deployed sensor networks.

Table 1: Nomenclature and symbols.

$x \in [0, L_e]$	Location in road segment
$u(x, t)$	Traffic flow speed
$x_i(t)$	Vehicle trajectory
$\hat{P}(x, t)$	Estimated traffic data
\tilde{u}	Equilibrium speed
$p(\rho)$	Traffic pressure
$S(k)$	Sensor readings at time k
$t_{\text{up}}, t_{\text{down}}$	Time signals exceed threshold
$\Delta t, \Delta x$	Temporal-spatial scales
e_k^m	Error from sensor k of vehicle m
$\eta = (s - xt)/\tau$	Self-similar variable
g_i^l, g_i^u	Min/max green time
$J_m(k)$	Cost function on lane m
$q_{\text{out}}^j(k)$	Outflow in phase j
$d_i(k)$	Demand flow in phase j
S_{ni}^g	Saturation flow for green
$\xi_{ni}(k)$	Existing phase state
$t \in [0, +\infty)$	Observation time
$\rho(x, t)$	Traffic density
$p(x, t)$	Traffic data
ρ_M	Maximum traffic density
u_f	Free speed on empty road
$s(x, t)$	Flow production rate
$h(k)$	Vehicle detection threshold
$d(k)$	Detection flag
u_k^m	Speed of vehicle m at sensor k
\hat{E}_k	Mean square error (MSE)
$C_{cf}^i(t)$	Congestion factor of lane i
G_i	Effective green time
$q_{\text{in}}^j(k)$	Inflow in phase j
$q_s^i(k)$	Arrival traffic flow at stop line
$s_j(k)$	Exit flow in phase j
S_{nj}^y	Saturation flow for yellow
$l_{ni}(k)$	Queue length in phase i

3.1. Continuum Traffic Flow Theory and Theoretical Models

The continuum model is excellent to describe the macroscopic traffic properties such as traffic congestion state. In 1955, Lighthill and Whitham introduced the continuum model (LWR model) [30] based on fluid dynamics, which builds the continuous function between traffic density and speed to capture the characteristics such as traffic congestion formation. In 1971, Payne introduced dynamics equations based on the continuum model and proposed the second-order model (Payne-Whitham model) [31]. Consider the

Payne-Whitham model defined by (3.1) (conservation of mass) and the acceleration equation, written in nonconservative form as (3.2):

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} = s(x, t), \quad (3.1)$$

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + \frac{1}{\rho} \frac{\partial p}{\partial x} = \frac{1}{\tau} (\tilde{u} - u), \quad (3.2)$$

where x and t denote the space and time, respectively, $u(x, t)$ and $\rho(x, t)$ are the traffic flow speed and density at the point x and time t , respectively, ρ is traffic density in unit of vehicles/length, τ is delay, and p is traffic pressure which is inspired from gas dynamics and typically assumed to be a smooth increasing function of the density only, that is, $p = p(\rho)$. The parameter \tilde{u} denotes the equilibrium speed that drivers try to adjust under a given traffic density ρ , which is a decreasing function of the density $\tilde{u} = \tilde{u}(\rho)$ with $0 < \tilde{u}(0) = u_f < \infty$ and $\tilde{u}(\rho_M) = 0$. Here u_f is the desired speed on empty road, and ρ_M is the maximum traffic density at which vehicles are bumper-to-bumper in the traffic jam. In MIT model of self-sustained nonlinear traffic waves, the relationship between \tilde{u} and ρ denotes as the following. Here u_f denotes free flow speed, and ρ_M is the traffic flow density in congestion state:

$$\tilde{u}(\rho) = \tilde{u}_0 \left(1 - \frac{\rho}{\rho_M}\right)^n, \quad u(\rho) = u_f - \frac{u_f}{\rho_M} \rho. \quad (3.3)$$

In (3.1), the $s(x, t)$ is flow production rate, and for road segment with no ramp $s(x, t) = 0$, for entrance ramp $s(x, t) < 0$, and for exit ramp $s(x, t) > 0$. Assume the velocity of vehicle traveling from the given intersection during the green light interval is $v_x(t)$, and the intervals of green light phase are T ; thus the flow production rate can be denoted as follows:

$$s(x, t) = \int_0^T v_x(t) dt. \quad (3.4)$$

Based on the exact LWR solver developed by Berkeley [39], we can obtain the solutions of traffic equations with given initial parameters. That means the operation status and future parameters of the traffic flow can be predicted and analyzed on a system scale.

3.2. Signal Processing for Traffic Data Estimation Based on Sensor Networks

In this paper, we employ high sensitive magnetic sensor, as shown in Figure 2(a), to detect vehicles. Given that the detection radius is R , sensor node detects travelling vehicle with the ATDA algorithm developed by UC Berkeley [12], which detects vehicle presence based on an adaptive threshold, and estimates vehicle velocity with the time difference of up/down thresholds and the lateral offset [12, 23], as shown in Figure 2(b).

Where D is sensor separation, $s(t)$ is the raw data, which will be sampled as sensor readings in discrete values $s(k)$ and transformed to $a(k)$ after noise filtering. $h(k)$ is the threshold at detection interval k , and $d(k)$ is the corresponding detection flag. The instantaneous velocity can be estimated by (3.5). Here time t_{up} and t_{down} are the moments

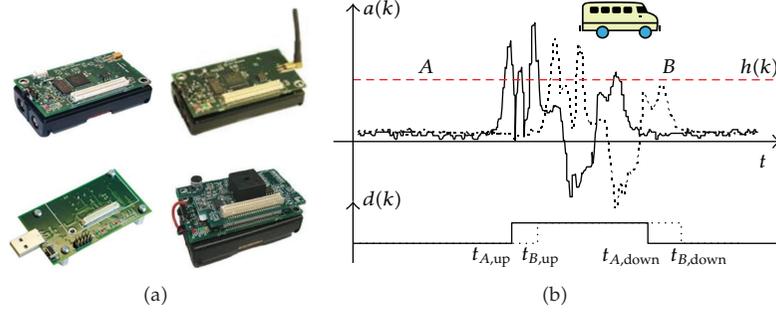


Figure 2: (a) Magnetic sensor node and gateway. (b) Presence and velocity detection based on ATDA.

when magnetic disturbance signals exceed the threshold continuously with count N and M , respectively:

$$\hat{v}_{mk} = \text{avg} \left(\frac{D_{AB}}{t_{B,up} - t_{A,up}}, \frac{D_{AB}}{t_{B,down} - t_{A,down}} \right). \quad (3.5)$$

In actual applications, for sake of cost, the sensor node number is expected as few as possible [40], so there need a tradeoff between sensor number and measurement precision. In this paper we try to improve the traffic detection exactness based on the spatial and temporal relations of sampled data. The main idea is to estimate the lost traffic information based on the limited sensor readings with traffic flow model and numerical interpolation. Assuming the temporal-spatial scales are Δt and Δx , the vehicle trajectory r and observation time t are dispersed into L and T sections, respectively. Consequently the two-dimensional $x - t$ domain is transformed to a grid mesh, as shown in Figure 3, which can be denoted by (3.6) for an arbitrary location and detection time. Where (x_i, t_j) is grid point and the h and k are spatiotemporal scales that can be denoted as $h \equiv \Delta x$ and $k \equiv \Delta t$,

$$x_i = ih, \quad t_j = jk, \quad i \in [0, L], \quad j \in [0, T]. \quad (3.6)$$

For sensor reading $u(x_i, t_j)$ in grid cell $g(i, j)$ may be considered as a detection unit on location $[i, i + 1] \cdot \Delta x$, and there is a single sensor node which takes effect in time interval $[j, j + 1] \cdot \Delta t$. To take into account the disconnected vehicle queue under unsaturated state, here the sensed traffic flow speed is defined as the average velocity of all vehicles that pass the detection point in predefined interval. In actual applications, the traffic data is typically collected in 20 s, 30 s, 1 min, or 2 mins.

The sensor network is sparsely deployed, and the total number of sensor node is K . We denote by v_{mk} the actual speed of the m th vehicle travelling from the k th sensor in the detection grid $g(i, j)$, \hat{v}_{mk} is the estimated speed calculated from sensor measures, u_k is the average speed in detection grid, m and m' are the first and last vehicles in detection interval,

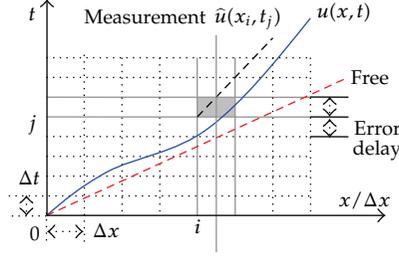


Figure 3: The detection grid in $x-t$ space.

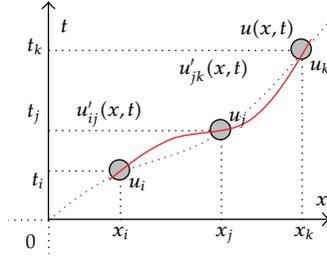


Figure 4: Scattered data fitting with proximity points.

respectively, and $u(x, t)$ is the theoretical speed based on the continuous traffic flow model. The actual and estimated traffic flow speed can be denoted by the following equations:

$$u_k = \frac{1}{m' - m} \sum_{i=m}^{m'} v_{ik}, \quad \hat{u}_k = \frac{1}{m' - m} \sum_{i=m}^{m'} \hat{v}_{ik}. \quad (3.7)$$

Assume that we have trajectories of a certain number of vehicles M in an observation interval. If the scale is small enough, it could be inferred that the traffic flow speed keeps unchanged in the unit grid. And consequently the partial differential equations (3.1)–(3.4) can be rewritten in an approximated way, such as (3.8). Here the subscripts i and j indicate space and time, respectively:

$$\left[\frac{\partial u}{\partial x} \right]_i^j = \frac{u_i^j - u_{i-1}^j}{h}. \quad (3.8)$$

With the scattered measurements as boundary initial values, the traffic data can be estimated by numerical interpolation based on the approximated traffic equations, as shown in Figure 4. For instance of traffic flow speed detection, denote by \hat{u}_k^m and u_k^m the estimated and actual velocities of m th ($m \in [1, M]$) vehicle on sensor k ($k \in [1, K]$), respectively. The estimation error is e_k^m , which can be formulated as

$$e_k^m = \hat{u}_k^m - u_k^m. \quad (3.9)$$

There are many evaluation criteria for error optimization; we use the same objective function as that in [41], which has the expression of (3.10) as follows. Here \hat{E} is the objective function, and \hat{E}_k is the mean square error (MSE) of traffic parameter estimation for all M vehicles on sensor k . And the purpose of optimal estimate algorithm is to minimize the total MSEs of all sensors:

$$\hat{E} = \frac{\sum_{k=1}^K \sum_{m=1}^M (e_k^m)^2}{M} = \sum_{k=1}^K \hat{E}_k \quad \text{where } \hat{E}_k = \frac{\sum_{m=1}^M (e_k^m)^2}{M}. \quad (3.10)$$

Assume K point data $\hat{u}(x_i, t_i)$ is obtained in detection area $g(i, j)$, and $u(x_i, t_i)$ is the corresponding value given by traffic equations. The noise root-mean-square error σ_{rms} between model outputs and measured data can be denoted as (3.11), which is a two-dimensional random field, and we assume it is unbiased:

$$\frac{1}{K} \sum_{i=1}^K \left[\frac{\hat{u}(x_i, t_i) - u(x_i, t_i)}{\hat{u}(x_i, t_i)} \right]^2 = \sigma_{\text{rms}}^2. \quad (3.11)$$

The velocity change in real traffic flow $u(x, t)$ is continuous. To eliminate noise, we introduce the smoothing factor with the minimum sum of squares of the second derivative, as shown in (3.12). Where Ω denotes two-dimensional square detection area,

$$\omega_{\min} = \min \int_{\Omega} \sum_x \sum_t \left(\frac{\partial^2 u(x, t)}{\partial x \partial t} \right)^2 d\Omega. \quad (3.12)$$

The traffic data estimation can be transformed to a two-dimensional data fitting problem with time-space constraints based on scattered measurements. To solve the conditional extremum problem based on (3.11) and (3.12), we can use the similar method in [42] based on Lagrange multiplier and finite elements method.

4. Congestion Factor Based Signal Optimization

In this section, we focus on traffic congestion evaluation and signal optimization. Based on traffic flow theory, the traffic flow near signalized intersections and connecting links can be modeled as entrance and exit ramps. The traffic light control algorithm will generate a shock wave at the stop line of the lanes, from the beginning of red signal phase, which will affect the traffic state in future. We introduce congestion factor to evaluate the degree of traffic congestion, and cost function to represent the influence of current timing phase on next phase. The result is helpful to optimize signal control.

4.1. Traffic Congestion Evaluation and Congestion Factor

The traffic congestion without external disturbance is an unsolved mystery. Knowing that traffic on a certain road is congested is actually not very helpful to traffic control system, and the information about how congested it is and the process it formed is more useful. There

is much novel research about traffic congestion prediction and evaluation in last decades [43, 44]. Flynn et al. studied these phenomena and introduced the traffic congestion model named *Jamitons* [20], in which the traffic congestion is modeled as traveling wave. Based on the traffic model described in (3.1)-(3.2), the traffic congestion can be expressed and denoted in a theoretical way. Assuming the speed of traveling wave is s , with introducing the self-similar variable defined by $\eta = (s - xt)/\tau$, the traffic equations in Section 3.1 can be rewritten, and (4.1) holds:

$$\frac{du}{d\eta} = \frac{(u - s)(\tilde{u} - u)}{(u - s)^2 - c^2}, \quad (4.1)$$

where s is the speed of the traveling shock wave, and traffic flow density and speed can be denoted as function of μ , $\rho = \rho(\eta)$, $u = u(\eta)$. The subcritical state can be predicted by (4.1), where $c = \sqrt{p\rho} > 0$ denotes the subcritical condition. To solve these equations, we select the shallow water equations [45] denoted as (4.2) to simplify the problem:

$$p = \frac{1}{2}\beta\rho^2. \quad (4.2)$$

Applying this assumption to (4.1) and the LWR model denoted by (3.1) and (3.2), (4.1) can be rewritten as (4.3). Here m is a constant denoting the mass flux of vehicles in the wave frame of reference:

$$\frac{du}{d\eta} = \frac{(u - s)\{\tilde{u}_0(1 - (m/\rho_M(u - s))) - u\}}{(u - s)^2 - (\beta m/(u - s))}. \quad (4.3)$$

The subcritical condition is therefore denoted as (4.4). If this equation is satisfied, the traffic congestion is inevitable to occur. The density will reach ρ_M immediately when traffic conditions exceed the subcritical state:

$$u_c = s + (\beta m)^{1/3}. \quad (4.4)$$

The road can be regarded as share resource for vehicle and traffic flow link, and according to Jain's fairness index for shared computer systems, the quantitative congestion factor can be defined based on the traffic congestion model, as (4.5). Here i indicates the lane number, x is the locations coordinate with origin starting from stop line, and the traffic density is sampled in n discrete values with fixed frequency. The congestion factor indicates the general congestion state on the whole road segment, which is a number between 0 and 1, and larger value means more crowded:

$$C_{cf}^i(t) = \frac{(\sum_{m=1}^n \rho(x_m))^2}{n \sum_{m=1}^n (\rho(x_m))^2}. \quad (4.5)$$

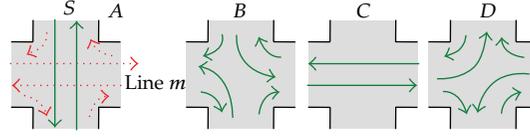


Figure 5: Four phases of traffic control.

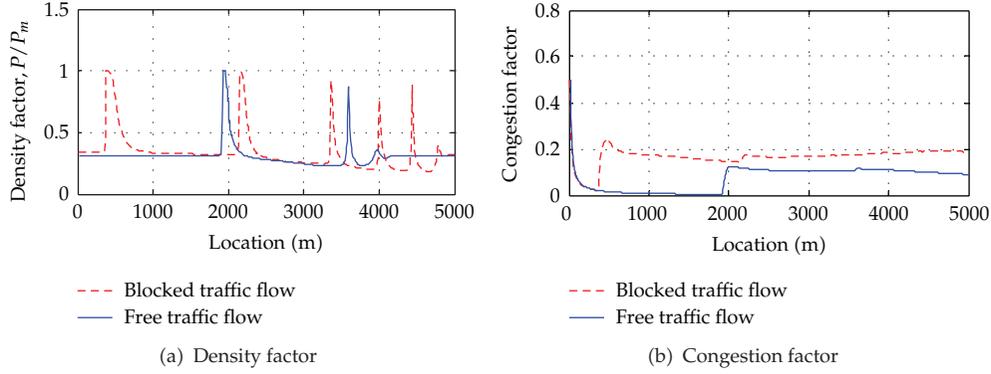


Figure 6: Traffic congestion factor at observation point x .

Considering an intersection with four phases numbered A , B , C , and D , as shown in Figure 5, the phase timing can be denoted as (4.6). Here g_i^l and g_i^u represent the minimum and maximum green times, respectively, and G_i is the effective green time of phase i :

$$G = \{G_A, G_B, G_C, G_D\}, \quad G_i \in [g_i^l, g_i^u]. \quad (4.6)$$

Under the scenario of traffic flow stops by red signal, for instance of lane m during signal phase i , the traffic flow from west to east will be blocked from the beginning of phase A , and the interval is G_A . The corresponding cost function on lane m is denoted as (4.7). Here ΔT is timing adjustment step length, and $C_{cf}^m(k)$ and $C_{cf}^m(k)$ represent congestion factor on lane m of traffic flow under blocking status by signal and normal condition with green light, respectively. The normal condition can be simulated based on (3.1) and (3.2) with initial values detected by sensor networks at time t , where $s(t) \equiv 0$. And traffic parameters can be predicted by resolving the traffic equations:

$$J_m(k) = \sum_{i=0}^K |C_{cf}^m(k) - C_{cf}^m(k)|, \quad k \in [0, K], \quad K = \frac{G_A}{\Delta T}. \quad (4.7)$$

With the Matlab implementation of an exact LWR solver [39], we can build a virtual simulator of traffic flow scheduling to analyze the traffic equations, congestion factor, and cost function in a theoretical way, based on given initial conditions. For traffic flow of a straight lane, consider two scenarios that traffic flow runs continuously and blocked by red signal at time t , the congestion factor and cost function can be simulated. The result is shown in Figure 6.

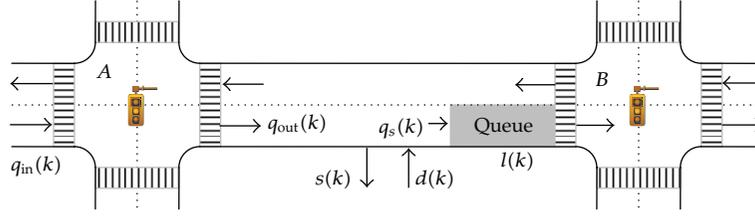


Figure 7: Urban intersection and road link model for traffic signal control.

4.2. The Multiobjective Optimization Model for Signal Control

The problem of traffic timing optimization for an urban intersection in a crowded city has been previously approached in much research [46, 47], and the existing traffic signal optimization formulations usually do not take traffic flow models in consideration. The variables on a signalized intersection and connecting links of phase j are shown in Figure 7. We define $q_{in}^j(k)$ and $q_{out}^j(k)$ to be the inflow and outflow, respectively, and define $d_j(k)$ and $s_j(k)$ to be the demand flow and exit flow during the phase j in an interval $[k\Delta T, (k+1)\Delta T]$, where ΔT is the timing adjustment step, and k is a discrete index. Define S_{nj}^g and S_{nj}^y as the saturation flow for green and yellow times of phase j at intersection n . $u_{ni}^k(k)$ indicates the signal, and $u_{ni}^k(k) = 0$ means green light and $u_{ni}^k(k) = 1$ means red light. To simplify the problem we just optimize the phase timing, with assumption that phase order is kept unchanged, four phases, as shown in Figure 5, transfer in the presupposed order A, B, C , and D .

Based on the dynamics of traffic flow, the control objective of the dynamic model is to minimize the total delay and traffic congestion factor. To minimize,

$$\text{Delay } TD = \Delta T \sum_{n=1}^N \sum_{i \in I_n} \sum_{k=1}^K l_{ni}(k), \quad (4.8)$$

$$\text{Congestion factor } CF = \sum_{m=1}^M \sum_{k=1}^K C_{cf}^m(k), \quad (4.9)$$

$$\text{Cost function } J = \sum_{m=1}^M \sum_{k=1}^K J_m(k). \quad (4.10)$$

With constraints subject to

$$\begin{aligned} g_i^l &\leq G_i \leq g_i^u, \\ l_{ni}(k) &\geq 0, \quad k \in K; \quad l_{ni}(k) \geq_{ni}(k-1) + (q_s^i(k) - q_{out}^i(k))\Delta T, \\ q_{in}^j(k) &= \sum_i b_{ij} q_{out}^i(k), \\ q_{out}^i(k) &= (1 - u_{ni}(k)) \left[S_{ni}^g (1 - \xi_{ni}(k)) + S_{ni}^y \times \xi_{ni}(k) \right] + S_{ni}^g \times \xi_{ni}(k) \times u_{ni}(k). \end{aligned} \quad (4.11)$$

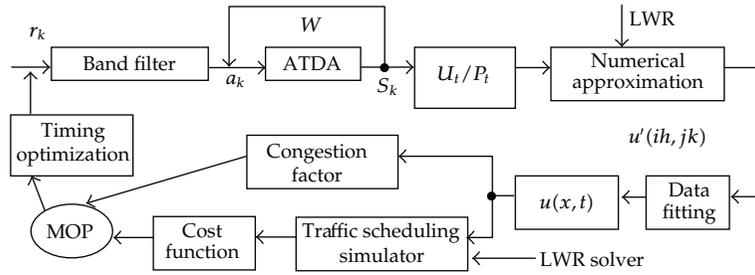


Figure 8: Flow diagram of traffic flow detection and adaptive control model based on sensor network.

For a given time window T , based on constraints of (4.10), the timing problem can be separated into h ($1 \leq h \leq T/g^l - 1$) subproblems. We can solve these h problems and obtain h noninferiority set of optimal solutions and then merge them to get a new noninferiority set of optimal solutions, which is the solution of the original problem. In this paper we use MOPSO-CD (*Multiobjective Particle Swarm Optimization Algorithm* using crowding distance) to find the optimal timing.

4.3. Traffic Flow Detection and Control Algorithms

Based on the above model and computational method, the overall block diagram of traffic data detection and control algorithm is shown in Figure 8. It employs magnetic sensor and detects magnetic signature based on ATDA. The individual vehicle data is collected in time window W , and traffic flow speed is monitored at regular intervals. The scattered point data U_t, P_t contains all sensor readings that will be used to approximate the traffic equation and numerical approximation $w(ih, jk)$ obtained. Finally we can get the traffic data $u(x, t)$ and $\rho(x, t)$, which is expected to provide data to traffic control and evaluate traffic congestion.

The traffic congestion state can be evaluated based on (3.9), and we can obtain the congestion factor in every segment near the intersection. At the same time, a cost function in next control phase can be calculated with a traffic scheduling simulator which is based on traffic equations and LWR solver. When we give priority to different possible directions and block traffic flow on other directions, the overall delay cost from alternative timing strategy will be taken into consideration before making the final signal, and the optimal timing can be obtained by solving a MOP. Finally, the traffic controller will choose the optimal timing scheme. This process operates in a circulation and in an adaptive way.

5. Simulation Result and Performance Analysis

The model and algorithms are simulated based on VISSIM platform. The traffic flow data is generated with the Mobile Century field test dataset [22, 48] and LWR solver [39]. VISSIM is a microscope, time interval, and driving behavior based traffic simulation tool kit. It supports external signal control strategies by providing API with DLL. The simulation tool will invoke the *Calculate* interface with presupposed frequency. And user can obtain the signal control related data in this interface.

With the DLL and COM interfaces, we designed a software/hardware in the loop simulation platform based on VISSIM, as shown in Figure 9.

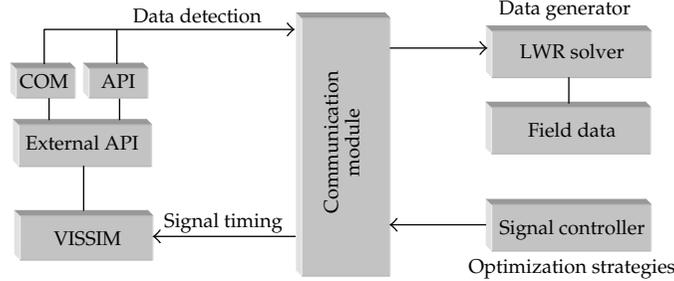


Figure 9: Software/hardware in the loop simulation based on VISSIM.

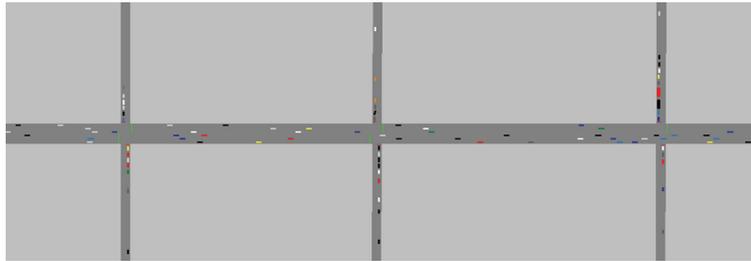


Figure 10: Traffic networks for timing optimization simulation.

The traffic data for simulation is based on Mobile Century dataset. Traffic data near three intersections is used to simulate traffic data collection and timing phase optimization. The traffic network is shown in Figure 10.

We select a fixed coordinate without sensor and try to estimate traffic parameters with the method proposed in this paper based on proximity sensor readings. The estimation precision under different smooth factor ω is shown in Figure 11. The performance is better when compared to traffic prediction based on BP neural network.

In the control simulation, we analyzed the performance by two scenarios: control with delay constraint only and combining delay with traffic congestion factor together as the optimization objective, and compare the performance with fixed time control. On the same traffic flow dataset, the performance is illustrated in Figure 12. The criteria include average delay and the maximum queue length. The result shows that congestion factor based control optimization can increase the performance with lower average waiting time and shorter queue length.

6. Conclusion and Future Research

In this paper we study the traffic flow congestion evaluation and congestion factor based control method using sparsely deployed wireless sensor network. Taking into consideration the traffic flow intrinsic properties and traffic congestion model, try to obtain optimal phase timing with as few sensor node as possible. The main idea is to study the congestion and its influence on future traffic flow, combine traffic equations with the optimization function, to obtain the numerical solution of the traffic equations via approximate method, and finally to refine traffic sensor data based on data fitting. The model and algorithms are simulated based

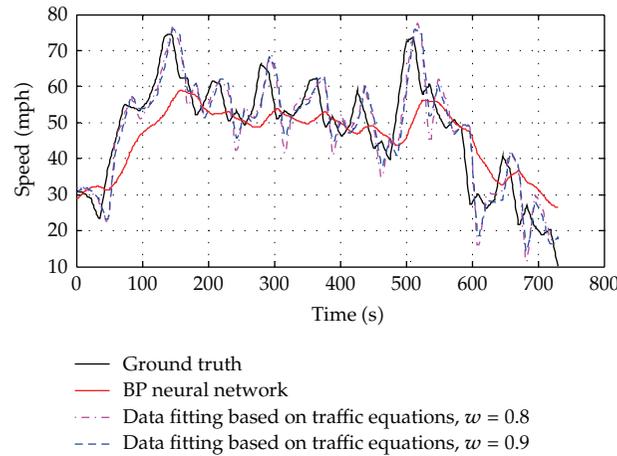


Figure 11: Performance of traffic data estimation based on traffic equations.

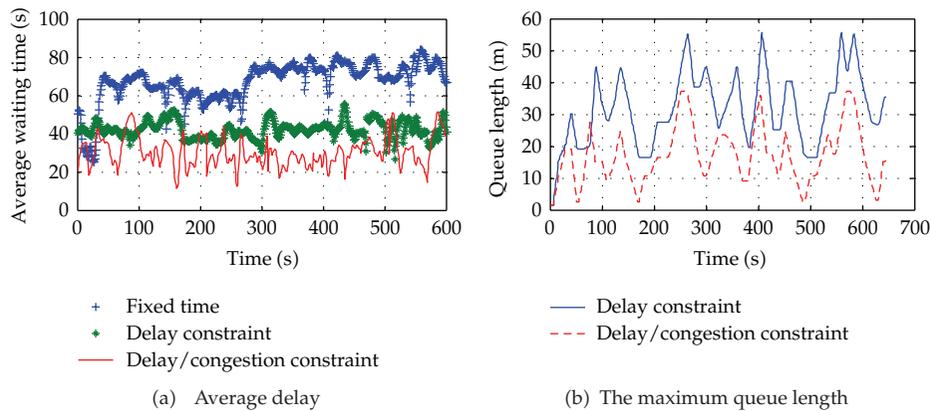


Figure 12: Performance analysis of traffic control based on congestion factor.

on VISSIM platform and *Mobile Century* dataset. The result shows better performance, and it is helpful to decrease average delay and the maximum queue length at the intersection.

Current research is limited to single intersection and simple segments with continuous traffic flow. Future research should focus on complex segments and even road network, such as ramp, long road with multi-intersections. And the traffic control strategy, road capability, and dynamics caused by incidents need to be taken into consideration in actual applications. Furthermore, complex traffic flow pattern simulation and traffic control strategies on a networked scale among multi-intersections and arbitrary connecting segments in road network are also an important aspect in next step.

Acknowledgments

This work was supported in part by the National High Technology Research and Development 863 Program of China under Grant no. 2012AA111902, the National Key Technology R&D Program of China under Grant no. 2011BAK02B02, the National Natural

Science Foundation of China under Grant no. 60873256, and the Fundamental Research Funds for the Central Universities under Grant no. DUT12JS01.

References

- [1] IBM, "Frustration rising: IBM 2011 Commuter Pain Survey," 2011, <http://www.ibm.com/us/en/>.
- [2] D. Metz, "The myth of travel time saving," *Transport Reviews*, vol. 28, no. 3, pp. 321–336, 2008.
- [3] G. Orosz, R. Eddie Wilson, and G. Stefan, "Traffic jams: dynamics and control," *Philosophical Transactions of the Royal Society A*, vol. 368, no. 1928, pp. 4455–4479, 2010.
- [4] E. Naone, "GPS data on Beijing cabs reveals the cause of traffic jams," *MIT Technology Review*, 2011.
- [5] G. Dimitrakopoulos and P. Demestichas, "Intelligent transportation systems: systems based on cognitive networking principles and management functionality," *IEEE Vehicular Technology Magazine*, vol. 5, no. 1, pp. 77–84, 2010.
- [6] A. Faro, D. Giordano, and C. Spampinato, "Integrating location tracking, traffic monitoring and semantics in a layered ITS architecture," *IET Intelligent Transport Systems*, vol. 5, no. 3, pp. 197–206, 2011.
- [7] J. P. Zhang, F. Y. Wang, and K. F. Zhang, "Data-driven intelligent transportation systems: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.
- [8] R. M. Murray, K. J. Åström, S. P. Boyd, R. W. Brockett, and G. Stein, "Future directions in control in an information-rich world," *IEEE Control Systems Magazine*, vol. 23, no. 2, pp. 20–33, 2003.
- [9] P. R. Kumar, *The Third Generation of Control Systems*, Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, Ill, USA, 2009.
- [10] J. W. C. Van Lint, A. J. Valkenberg, A. J. Van Binsbergen, and A. Bigazzi, "Advanced traffic monitoring for sustainable traffic management: experiences and results of five years of collaborative research in the Netherlands," *IET Intelligent Transport Systems*, vol. 4, no. 4, pp. 387–400, 2010.
- [11] A. K. Lawrence, K. M. Milton, and R. P. G. David, *Traffic Detector Handbook*, vol. 1-2, Federal Highway Administration, U.S. Department of Transportation, 2006.
- [12] S. Y. Cheung and V. Pravin, "Traffic surveillance by wireless sensor networks," Final Report, Department of Electrical Engineering and Computer Science, University of California, Berkeley, Calif, USA, 2007.
- [13] L. Atzori, A. Iera, and G. Morabito, "The Internet of things: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
- [14] F. Qu, F. Y. Wang, and L. Yang, "Intelligent transportation spaces: vehicles, traffic, communications, and beyond," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 136–142, 2010.
- [15] D. Estrin, D. Culler, K. Pister, and G. Sukhatme, "Connecting the physical world with pervasive networks," *IEEE Pervasive Computing*, vol. 1, no. 1, pp. 59–69, 2002.
- [16] M. Welsh, "Sensor networks for the sciences," *Communications of the ACM*, vol. 53, no. 11, pp. 36–39, 2010.
- [17] M. Tubaishat, Z. Peng, Q. Qi, and S. Yi, "Wireless sensor networks in intelligent transportation systems," *Wireless Communications and Mobile Computing*, vol. 9, no. 3, pp. 287–302, 2009.
- [18] D. Tacconi, D. Miorandi, I. Carreras, F. Chiti, and R. Fantacci, "Using wireless sensor networks to support intelligent transportation systems," *Ad Hoc Networks*, vol. 8, no. 5, pp. 462–473, 2010.
- [19] U. Lee and M. Gerla, "A survey of urban vehicular sensing platforms," *Computer Networks*, vol. 54, no. 4, pp. 527–544, 2010.
- [20] M. R. Flynn, A. R. Kasimov, J. C. Nave, R. R. Rosales, and B. Seibold, "Self-sustained nonlinear waves in traffic flow," *Physical Review E*, vol. 79, no. 5, pp. 61–74, 2009.
- [21] K. P. Li, *Guidelines for Traffic Signals (RiLSA)*, China Architecture & Building Press, Beijing, China, 2006.
- [22] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment," *Transportation Research Part C*, vol. 18, no. 4, pp. 568–583, 2010.
- [23] A. Haoui, R. Kavalier, and P. Varaiya, "Wireless magnetic sensors for traffic surveillance," *Transportation Research Part C*, vol. 16, no. 3, pp. 294–306, 2008.
- [24] B. Hull, V. Bychkovsky, Y. Zhang et al., "CarTel: a distributed mobile sensor computing system," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys '06)*, pp. 125–138, New York, NY, USA, November 2006.

- [25] A. Thiagarajan, L. Ravindranath, K. LaCurts et al., "VTrack: accurate, energy-aware road traffic delay estimation using mobile phones," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys '09)*, pp. 85–98, Berkeley, Calif, USA, November 2009.
- [26] J. Bacon, A. R. Beresford, D. Evans et al., "TIME: an open platform for capturing, processing and delivering transport-related data," in *Proceedings of the 5th IEEE Consumer Communications and Networking Conference (CCNC '08)*, pp. 687–691, Las Vegas, Nev, USA, January 2008.
- [27] D. C. Gazis, "The origins of traffic theory," *Operations Research*, vol. 50, no. 1, pp. 69–77, 2002.
- [28] G. F. Newell, "Memoirs on highway traffic flow theory in the 1950s," *Operations Research*, vol. 50, no. 1, pp. 173–178, 2002.
- [29] H. Lieu, "Traffic flow theory: a state-of-the-art report. Committee on Traffic Flow Theory and Characteristics, FHWA," Department of Transportation, 2001.
- [30] M. J. Lighthill and G. B. Whitham, "On kinematic waves II: a theory of traffic on long crowded roads," *Proceedings of the Royal Society of London A*, vol. 229, no. 1178, pp. 317–345, 1955.
- [31] H. J. Payne, "Models of freeway traffic and control," *Mathematical Models of Public System*, vol. 1, pp. 51–61, 1971.
- [32] S. Sun, G. Yu, and C. Zhang, "Short-term traffic flow forecasting using sampling Markov Chain method with incomplete data," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 437–441, June 2004.
- [33] W. Zheng, D. H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.
- [34] Y. Zhang and Z. Ye, "Short-term traffic flow forecasting using fuzzy logic system methods," *Journal of Intelligent Transportation Systems*, vol. 12, no. 3, pp. 102–112, 2008.
- [35] Y. Kamarianakis and P. Prastacos, "Space-time modeling of traffic flow," *Computers and Geosciences*, vol. 31, no. 2, pp. 119–133, 2005.
- [36] Y. Sugiyamal, M. Fukui, M. Kikuchi et al., "Traffic jams without bottlenecks-experimental evidence for the physical mechanism of the formation of a jam," *New Journal of Physics*, vol. 10, no. 3, pp. 1–7, 2008.
- [37] D. B. Work, O. P. Tossavainen, Q. Jacobson, and A. M. Bayen, "Lagrangian sensing: traffic estimation with mobile devices," in *Proceedings of the American Control Conference (ACC '09)*, pp. 1536–1543, June 2009.
- [38] N. Shrivastava, R. Mudumbai, U. Madhow, and S. Suri, "Target tracking with binary proximity sensors," *ACM Transactions on Sensor Networks*, vol. 5, no. 4, pp. 1–30, 2009.
- [39] P. E. Mazare, C.G. Claudel, and A.M. Bayen, "Analytical and grid-free solutions to the Lighthill-Whitham-Richards traffic flow model," Department of Civil and Environmental Engineering, University of California at Berkeley, Berkeley Calif, USA, 2011.
- [40] M. Gentili and P. B. Mirchandani, "Locating sensors on traffic networks: models, challenges and research opportunities," *Transportation Research Part C*, vol. 24, pp. 227–255, 2012.
- [41] X. Jeff Ban, L. Chu, R. Herring, and J. D. Margulici, "Sequential modeling framework for optimal sensor placement for multiple intelligent transportation system applications," *Journal of Transportation Engineering*, vol. 137, no. 2, pp. 112–120, 2010.
- [42] Z. Y. Cai and M. Z. Li, "A smoothing-finite element method for surface reconstruction from arbitrary scattered data," *Journal of Software*, vol. 14, no. 4, pp. 838–844, 2003.
- [43] S. F. Hasan, N. H. Siddique, and S. Chakraborty, "Extended MULE concept for traffic congestion monitoring," *Wireless Personal Communications*, vol. 63, pp. 65–82, 2010.
- [44] J. Bacon, A. Bejan, D. Evans et al., "Using real-time road traffic data to evaluate congestion," *Lecture Notes in Computer Science*, vol. 6875, pp. 93–117, 2011.
- [45] R. F. Dressler, "Mathematical solution of the problem of roll waves in inclined channel flow," *Communications on Pure and Applied Mathematics*, vol. 2, no. 2-3, pp. 149–194, 1949.
- [46] M. Dotoli, M. P. Fanti, and C. Meloni, "A signal timing plan formulation for urban traffic control," *Control Engineering Practice*, vol. 14, no. 11, pp. 1297–1311, 2006.
- [47] W. M. Wey, "Model formulation and solution algorithm of traffic signal control in an urban network," *Computers, Environment and Urban Systems*, vol. 24, no. 4, pp. 355–377, 2000.
- [48] Mobile Century dataset, Department of Electrical Engineering and Computer Science, University of California, Berkeley, Calif, USA, 2008, <http://traffic.berkeley.edu/>.

Research Article

Electrical Drive Radiated Emissions Estimation in Terms of Input Control Using Extreme Learning Machines

**A. Wefky, F. Espinosa, L. de Santiago, P. Revenga,
J. L. Lázaro, and M. Martínez**

Electronics Department, Polytechnics School, University of Alcalá, Campus Universitario 28871, Alcalá de Henares, Spain

Correspondence should be addressed to F. Espinosa, espinosa@depeca.uah.es

Received 21 September 2012; Revised 9 November 2012; Accepted 9 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 A. Wefky et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the increase of electrical/electronic equipment integration complexity, the electromagnetic compatibility (EMC) becomes one of the key points to be respected in order to meet the constructor standard conformity. Electrical drives are known sources of electromagnetic interferences due to the motor as well as the related power electronics. They are the principal radiated emissions source in automotive applications. This paper shows that there is a direct relationship between the input control voltage and the corresponding level of radiated emissions. It also introduces a novel model using artificial intelligence techniques for estimating the radiated emissions of a DC-motor-based electrical drive in terms of its input voltage. Details of the training and testing of the developed extreme learning machine (ELM) are described. Good agreement between the electrical drive behavior and the developed model is observed.

1. Introduction

Because of the integration density increase and the assembly of electric/electronic circuits gathered in a confined space such as electrical vehicles, certain unwanted effects caused by EMC and electromagnetic interference (EMI) become unavoidable [1–3]. Electrical drives are the main source of electromagnetic emissions of electrical vehicles, including both electrical motor and the required power electronics. Indeed, the addition of equipment operating in fast switching is susceptible to cause harmful conducted and/or radiated interference that can spread through the onboard network [4]. These coupling effects may damage the electrical module in which they were integrated as can be found in digital and mixed-signal electronic circuits such as the RF devices and microcontrollers integrated in the command devices [5].

To predict the various EMI effects generated by electrical motors, different models of machine windings [6] and magnets [7] using lumped elements [8] were proposed. The influences of an innovative mechanism of collector-brush contact on the EMC model of DC motors have been investigated in [9].

ELMs are one of the most powerful emergent tools that have been widely used in recent years in various fields [10] such as security assessment [11], data privacy [12], EEG and seizure detection [13], image quality assessment [14], implementation with FPGAs [15], face recognition [16], and human action recognition [17]. Function approximation is one of the basic learning tasks that an ELM can accomplish [18]. This paper exploits this capability to estimate the drive's radiated emissions in terms of its input voltage.

The Electronics Department at the University of Alcalá (UAH), in collaboration with the Thermal Engines Group of the ETSII-UPM in Madrid and the Research Center for Environmental Energy and Technology (CIEMAT) in Madrid developed an electronic measurement equipment to relate the drive activity, vehicle state, and road conditions with pollutant emissions (gases and particles) in real traffic conditions [19]. This was the starting point to extend the study to radiated electromagnetic emissions. In order to start the modeling process with a basic configuration, this work concentrates on the effect of the input voltage of an electrical drive based on a DC motor on the drive radiated emissions.

This paper is arranged as follows. Section 2 highlights the context of this work by discussing the relationship between the driving profile and the vehicular radiated emissions. Section 3 describes the key aspects of the measurement methodology followed in the experimental tests. Details of ELM model development are discussed in the Section 4. Section 5 is devoted to commenting on the obtained experimental results. Finally, conclusions and future work are included in Section 6.

2. Driving Profile versus Vehicular Radiated Emissions

Current EMC standards, like CISPR 16-2-3 [20], do not discuss the effect of the driving characteristics on real traffic vehicular radiated emissions. There are a lot of driving signals that can describe the driving style such as vehicle speed, linear acceleration, frontal inclination, regime engine, following distance, relative lane position, yaw angle, position of throttle, clutch, and brake pedals [21, 22].

A straightforward solution to study the effect of the driving profile on the radiated emissions could be to measure a set of driving profile signals as well as EMI signals simultaneously in real traffic conditions. Measuring some driving profile signals in real time is an easy task because in most cases the onboard electronic system can easily provide information about most of these variables. On the other hand, measuring radiated electromagnetic emissions due to a specific electric vehicle in real traffic seems to be a very complicated task. This is because the onboard antenna that would receive the radiated emissions due to the vehicle under test (VUT) only would receive EMI signals from many other sources like radiated emissions coming from other vehicles, WiFi, AM radio, FM radio, television broadcast, mobile networks, satellite networks, Bluetooth devices, GPS, high voltage towers, and so forth. Mainly for this reason, but also for the size and weight of the required antennas for this purpose, it is impossible to measure the real traffic vehicular radiated emissions due to a specific VUT. Thus, the development of a model that would be able to estimate real traffic vehicular radiated emissions in terms of the corresponding driving style signals would be a novel work. In this way, the relative change of the real traffic radiated

emissions in terms of the driving profile parameters can be quantified. Therefore, guidelines can be determined in order to ensure green driving profiles in terms of minimization of vehicular radiated emissions.

In order to estimate the vehicular real traffic radiated emissions in terms of the driving behaviour signals, authors propose a process of three main stages as shown in Figure 1. Firstly, tests with the vehicle in a semianechoic chamber have to be done measuring some driving profile signals as well as the corresponding radiated emissions. Secondly, ELMs should be exploited to develop the desired model using data registered from the previous stage. Thirdly, real traffic experiments have to be done registering only the driving profile signals that will be simultaneously applied to the obtained model estimating the real traffic radiated emissions in terms of the registered real traffic driving profile signals.

As the electrical drive of electric vehicles is one of the most powerful sources of radiated emissions [23], the authors in this work have applied the first two steps of the above algorithm on an electrical drive as a first approach, that is, a DC-motor-based electrical drive has been tested in a semianechoic chamber measuring its input voltage that emulates the driving profile signals as well as its corresponding radiated emissions. Then, the authors have developed a neural network model that is capable of estimating the radiated emissions of the electrical drive in terms of the corresponding input voltage.

As a first trial, the authors presented a measurement methodology of the radiated emissions of electric vehicles as well as the driving profile [24]. This methodology is based on frequency domain EMI measurement procedure where a spectrum analyzer has been used to make sweeps of the radiated emissions signal. In the present work, the authors propose a time domain EMI measurement system based on a digital oscilloscope in order to save the overall measurement time.

3. Measurement Methodology

The basic idea of the proposed methodology is shown in Figure 2. Firstly, the EMI signal of the drive under test (DUT) is captured by a biconical antenna, filtered by an antialiasing low-pass filter, sampled as well as quantized by the digital oscilloscope's ADC, and sent to the converter via the GPIB bus which delivers it to the PC's USB port to be saved in a database for offline processing. The previous scenario is repeated until all the time domain sweeps are saved in the PC's database.

At the end of the experiment, spectrum and spectrogram of the electrical drive's EMI signal are computed via the discrete Fourier transform (DFT). Moreover, the time domain evolution of the electrical drive's radiated emissions is displayed with the input voltage profile applied to the electrical drive.

According to the measurement scenario described in the previous section, the digital oscilloscope provides a matrix of radiated emission sweeps. Then, the sweep mean power (SMP) is calculated as follows:

$$\text{SMP} = \frac{\left((1/N_s) \sum_{i=1}^{N_s} v_i^2 \right)}{50}, \quad (3.1)$$

where N_s is the sweep length, i is the sweep sample index, and v is the sweep sample value in Volts.

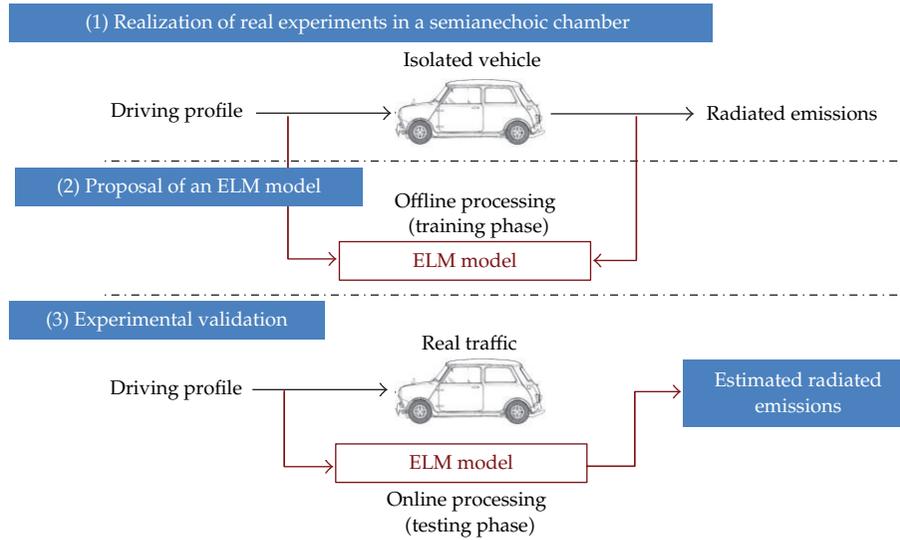


Figure 1: Complete process for estimating real traffic vehicular radiated emissions in terms of the driving profile.

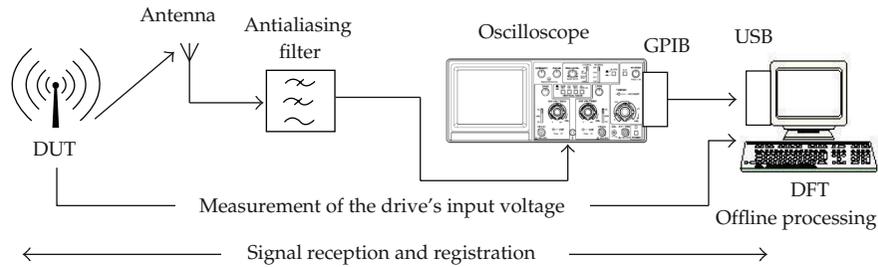


Figure 2: Block diagram of the proposed TDEMI measurement system.

The frequency range of interest in the study must be covered by the antenna bandwidth. In order to meet the sampling theorem and to avoid the aliasing effect, an antialiasing low-pass filters (LPF) must be connected to the antenna.

The following parameters of the oscilloscope should be properly adjusted: sampling time T_s (or the sampling frequency F_s), capture time T_c (duration of the time domain sweep of the oscilloscope), and sweeping time T_{sw} (or the sweeping frequency). The relation between the sampling, sweeping, and capturing times is illustrated in Figure 3.

According to the sampling theorem, the minimum sampling frequency equals twice the wanted maximum frequency (Nyquist frequency) of the calculated spectrum. To avoid aliasing errors, it is recommended to set the sampling frequency 2 to 4 times higher than Nyquist frequency depending on the steepness of the antialiasing filter. The sampling frequency used in this work was 12.5 MHz with an antialiasing filter of 1.9 MHz cutoff frequency.

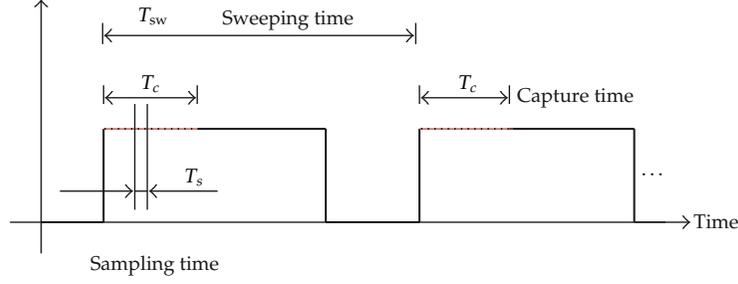


Figure 3: The relation between the sampling, sweeping, and capture times.

Besides, the capture time T_c depends on the distance between two neighboring frequency bins (frequency resolution or frequency step) Δf which depends on the resolution bandwidth B_r as follows [25–28]:

$$T_c = \frac{1}{\Delta f} = \frac{1}{1.06B_r}. \quad (3.2)$$

As a result of choosing these two parameters, that is, F_s and T_c , the time domain sweep length N_s , or number of samples per sweep is determined as follows:

$$N_s = F_s \times T_c. \quad (3.3)$$

It is noteworthy that the sweep length is limited by the oscilloscope's available memory as well as the maximum data transmission speed between the oscilloscope and the personal computer. As can be seen from (3.3), the sweep length depends on both the sampling frequency and the capture time. The capture time is already set by the resolution bandwidth determined by the currently available standards as can be seen in (3.2). On the other hand, the sampling frequency can be more than twice the maximum measured signal bandwidth. Thus, the lower bound on the sampling frequency is already imposed by the sampling theorem. Consequently, the upper bound on the sampling frequency is the value at which the oscilloscope's memory saturates or the maximum transmission speed between the oscilloscope and the PC is reached.

Finally, the sweeping time T_{sw} (time between the start of two consecutive records), or the recording frequency F_{sw} , has to be adjusted based on the profile of the drive input voltage and the maximum transmission speed between the PC and the oscilloscope. In other words, the details and changes included in the input voltage profile of the DUT impose an upper bound on the recording time, that is, the maximum separation between the starting of consecutive records has to be less than the minimum separation between two successive important changes in the input voltage profile of the DUT.

Moreover, the maximum data transmission speed between the oscilloscope and the PC passing by the GPIB/USB converter controls the lower limit on the recording time T_r . In other words, the minimum recording time corresponds to the maximum data transmissions speed between the oscilloscope and the PC.

The proposed TDEMI system has been validated against an ESIB26 EMI receiver from Rhode and Schwarz, comparing the spectra obtained by both systems due to emissions of the

Table 1: Comparison between generalization RMSE corresponding to different hidden neurons.

Hidden neurons	2	4	10	15	20	25	30
Testing RMSE * $1e - 8$	1.3127	1.2822	1.3363	1.3751	1.4408	1.3689	1.6281

same square wave signal source. The validation results showed the agreement between the spectra obtained by the TDEMI system and the EMI receiver. It is noteworthy that the authors in this work are not interested in obtaining accurate absolute levels of radiated emissions. On the other hand, they are interested in measuring the relative change in the level of radiated emissions due to the change in the applied voltage.

4. Extreme Learning Machines

Artificial intelligence techniques have been used in various applications. Among different computational intelligence techniques, ELM is considered a recently emerging technique that overcomes some challenges faced by artificial neural networks (ANNs) and support vector machines (SVMs) such as slow learning speed, trivial human intervene, and poor computational scalability [29]. Moreover, ELMs satisfy the universal approximation property [30].

When a particular ELM model fails, it could be due to one of two reasons. The model parameters fail to converge to the proper values, perhaps due to unsuitable model initialization, or the inability of the given model to implement the desired function, perhaps due to an insufficient number of hidden neurons. In this work, to avoid the first possibility, each ELM model was trained and tested 50 times. And the network architecture with the lowest root mean square error (RMSE) on the testing data set has been chosen.

Concerning the second reason, there is no theory yet to explain how many hidden neurons are needed to approximate any given function. If there are too few hidden neurons, a high training error and high generalization error would result from underfitting. On the other hand, if there are too many hidden neurons, there would be a low training error, but there would still be a high generalization error due to overfitting. In most situations, there is no way to determine the best number of hidden neurons without training several networks and estimating the generalization error of each [30–33]. In this paper, the network growing technique [31] is applied by adding hidden neurons sequentially from 1 to 30 comparing the testing RMSE error.

Table 1 shows that 4 hidden neurons have achieved the best generalization performance in terms of the testing RMSE calculated as follows:

$$\text{RMSE} = \left[\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right]^{1/2}, \quad (4.1)$$

where O is the vector of observed (measured) values, P is the vector of model estimated values, and N is the number of samples in the testing subset. The input to the ELM model is the drive input voltage, while the output is its radiated emissions. Therefore, this paper presents a single-input single-output ELM model whose structure is shown in Figure 4.

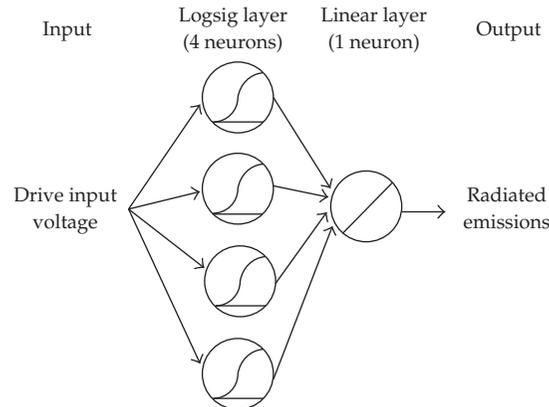


Figure 4: Structure of the ELM network.

5. Experimental Results

Experimental tests have been realized in the semianechoic chamber of the High Technology and Homologation Centre at the University of Alcalá. Figure 5 shows a part of the semianechoic chamber with experimental setup used in this work. It shows the biconical antenna applied to receive the electrical drive's radiated emissions as well as the DUT composed of a DC motor and a power electronics card.

The electrical drive has been subjected to two voltage profiles as can be seen in Figures 6 and 7. Radiated emissions registered by the proposed TDEMI system have been processed to calculate both the spectrograms shown in Figures 8 and 9, as well as the SMPs shown in Figures 6 and 7, using (3.1).

The data of the peaks profile, described in Figure 6, has been used in the training phase of the ELM network. This Figure shows that the ELM model has successfully learned to estimate the SMP of the EMI signal in terms of the corresponding input voltage value.

However, the steps profile data, depicted in Figure 7, has been exploited in the testing phase. It also shows that the developed ELM model has been successfully and accurately estimated the SMP of the radiated emissions signal in terms of unseen values of the drive input voltage. The proposed model has been able to precisely distinguish between 8, 10, and 15 volts of the electrical drive excitation.

Figures 8 and 9 illustrate the spectrograms of the radiated emissions EMI signals registered for the peaks and steps driving profiles, respectively. These figures also show a correspondence between the drive input voltage profile and the intensity of the radiated emissions.

6. Conclusions

This paper presents a proposal of estimating real traffic vehicular radiated emissions in terms of driving profile signals. The first two steps of this methodology have been applied to an electrical drive based on a DC motor. It has been shown that there is a direct relationship between the electrical drive's input voltage and its corresponding radiated emissions. An

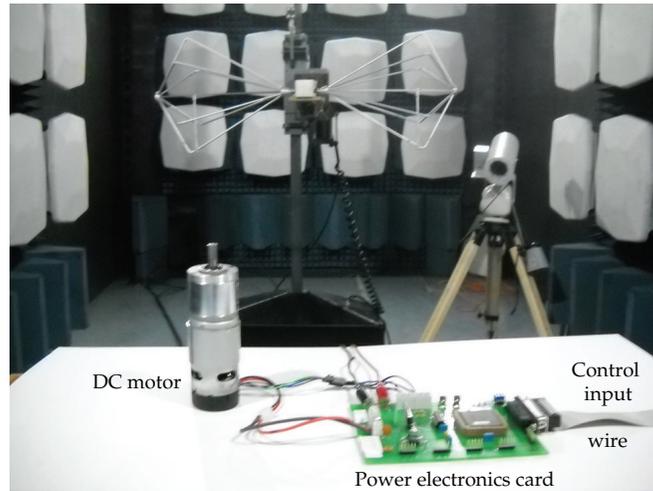


Figure 5: Experimental setup of an electrical drive including a DC motor (left side on the table) and the power electronics card (right side on the table), inside a semianechoic chamber.

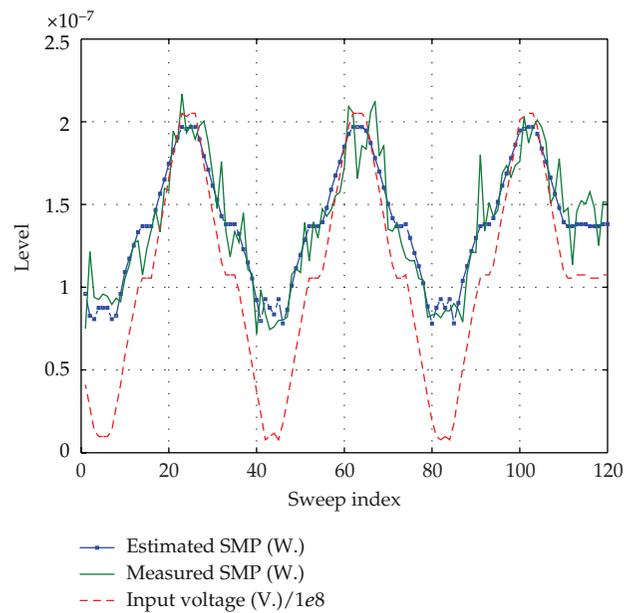


Figure 6: Training data (peaks profile).

ELM model with 4 hidden neurons has been developed to estimate the radiated emissions in terms of the drive input. The model has been validated with experimental data of a real electrical drive tested in a semianechoic chamber.

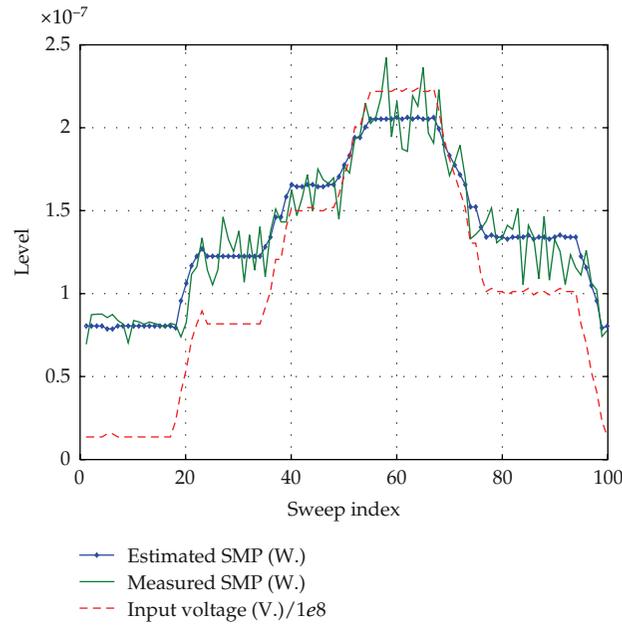


Figure 7: Testing data (steps profile).

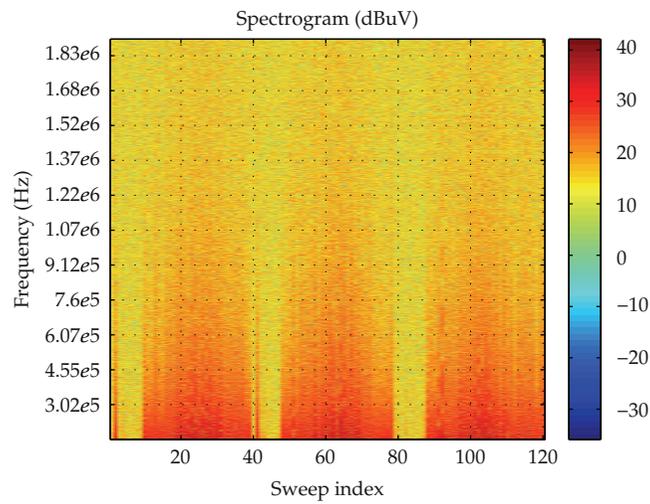


Figure 8: Spectrogram of the radiated emissions EMI signal (peaks profile).

Considering the successfully obtained results, as future work, the authors propose the online application of a model based on ELM to the radiated electromagnetic estimation of electric vehicles, according to the challenge described in Figure 1.

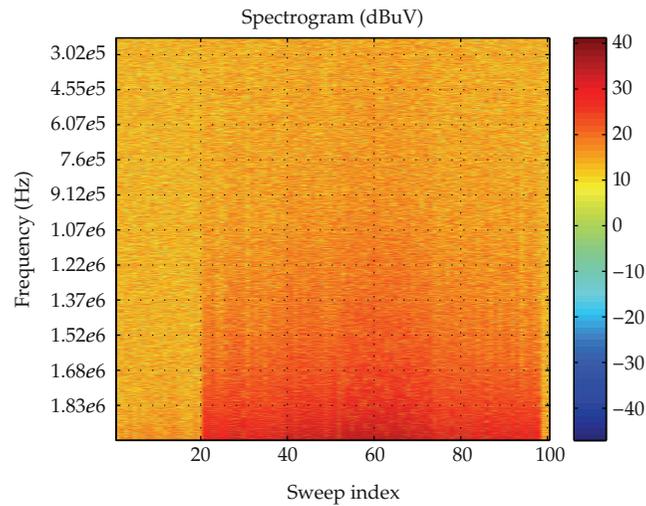


Figure 9: Spectrogram of the radiated emissions EMI signal (steps profile).

Acknowledgment

The authors would like to acknowledge the cooperation of the High Technology and Homologation Center at the University of Alcalá, Spain.

References

- [1] J. Ben Hadj Slama, S. Hrigua, F. Costa, B. Revol, and C. Gautier, "Relevant parameters of SPICE3 MOSFET model for EMC analysis," in *Proceedings of the IEEE International Symposium on Electromagnetic Compatibility (EMC '09)*, pp. 319–323, August 2009.
- [2] S. Chen, T. W. Nehl, J. S. Lai et al., "Towards EMI prediction of a PM motor drive for automotive applications," in *Proceedings of the 18th Annual IEEE Applied Power Electronics Conference and Exposition (APEC '03)*, pp. 14–22, February 2003.
- [3] L. Ferrer, J. Balcells, D. González, J. Gago, and M. Lamich, "Modelling of differential mode conducted EMI generated by switched power inverters," in *Proceedings of the 29th Annual Conference of the IEEE Industrial Electronics Society (IECON '03)*, pp. 2312–2315, November 2003.
- [4] R. Kahoul, Y. Azzouz, P. Marchal, and B. Mazari, "New behavioral modeling for DC motor armatures applied to automotive EMC characterization," *IEEE Transactions on Electromagnetic Compatibility*, vol. 52, no. 4, pp. 888–901, 2010.
- [5] W. Valente, M. H. Amaral, and A. Raizer, "EMC management: how to compare electromagnetic environmental measurements and equipment immunity levels," *Progress In Electromagnetics Research Letters*, vol. 18, pp. 165–177, 2010.
- [6] J. L. Guardado, J. A. Flores, V. Venegas, J. L. Naredo, and F. A. Uribe, "A machine winding model for switching transient studies using network synthesis," *IEEE Transactions on Energy Conversion*, vol. 20, no. 2, pp. 322–328, 2005.
- [7] M. K. Kazimierczuk, G. Sancineto, G. Grandi, U. Reggiani, and A. Massarini, "High-frequency small-signal model of ferrite core inductors," *IEEE Transactions on Magnetics*, vol. 35, no. 5, pp. 4185–4191, 1999.
- [8] S. Moreau, R. Kahoul, and J. P. Louis, "Parameters estimation of permanent magnetsynchronous machine without adding extra-signal input excitation," in *Proceedings of the IEEE International Symposium on Industrial Electronics (IEEE-ISIE '04)*, pp. 371–376, May 2004.
- [9] R. Kahoul, Y. Azzouz, and B. Ravelo, "Modelling of DC motors conducted low frequency EMI/EMC disturbance for automotive applications," *European Journal of Scientific Research*, vol. 63, no. 3, pp. 368–386, 2011.

- [10] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [11] Y. Xu, Z. Y. Dong, J. H. Zhao, P. Zhang, and K. P. Wong, "A reliable intelligent system for real-time dynamic security assessment of power systems," *IEEE Transactions on Power Systems*, vol. 27, no. 3, Article ID Article number6158623, pp. 1253–1263, 2012.
- [12] S. Samet and A. Miri, "Privacy-preserving back-propagation and extreme learning machine algorithms," *Data and Knowledge Engineering*, vol. 79–80, pp. 40–61, 2012.
- [13] Y. Song, J. Crowcroft, and J. Zhang, "Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine," *Journal of Neuroscience Methods*, vol. 210, no. 2, pp. 132–146, 2012.
- [14] S. Decherchi, P. Gastaldo, R. Zunino, E. Cambria, and J. Redi, "Circular-ELM for the reduced-reference assessment of perceived image quality," *Neurocomputing*, vol. 102, pp. 78–89, 2013.
- [15] S. Decherchi, P. Gastaldo, A. Leoncini, and R. Zunino, "Efficient digital implementation of extreme learning machines for classification," *IEEE Transactions on Circuits and Systems II*, vol. 59, no. 8, Article ID Article number6236105, pp. 496–500, 2012.
- [16] K. Choi, K.-A. Toh, and H. Byun, "Incremental face recognition for large-scale social network services," *Pattern Recognition*, vol. 45, no. 8, pp. 2868–2883, 2012.
- [17] R. Minhas, A. A. Mohammed, and Q. M. Wu, "Incremental learning in human action recognition based on snippets," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, 1 pages, 2011.
- [18] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 42, no. 2, Article ID 6035797, pp. 513–529, 2012.
- [19] F. Espinosa, J. A. Jiménez, E. Santiso et al., "Design and implementation of a portable electronic system for vehicle-driver-route activity measurement," *Measurement*, vol. 44, no. 2, pp. 326–337, 2011.
- [20] Commission, I. E., *CISPR 16-2-3 "Specification for radio disturbance and immunity measuring apparatus and methods Part 2-3: methods of measurement of disturbances and immunity—radiated disturbance measurements"*, 2008.
- [21] C. Miyajima, Y. Nishiwaki, K. Ozawa, T. Wakita, K. Itou, and K. Takeda, "Cepstral analysis of driving behavioral signals for driver identification," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, pp. V921–V924, May 2006.
- [22] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research Part C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [23] S. Alexandersson and M. Alaküla, "Automotive power electronic future—from an EMC perspective," in *Proceedings of the International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM '06)*, pp. 609–613, May 2006.
- [24] L. De Santiago, F. Espinosa, M. A. Ruiz et al., "Effect of electrical vehicle-driver interaction on the radiated electromagnetic emissions: measurement methodology," in *Proceedings of the IEEE-ICIT International Conference on Industrial Technology (ICIT '10)*, pp. 1113–1118, March 2010.
- [25] C. Keller and K. Feser, "Fast emission measurement in time domain," *IEEE Transactions on Electromagnetic Compatibility*, vol. 49, no. 4, pp. 816–824, 2007.
- [26] C. Keller and K. Feser, "A new method of emission measurement," in *Proceedings of the IEEE International Symposium on Electromagnetic Compatibility (EMC '02)*, pp. 599–604, August 2002.
- [27] C. Keller and K. Feser, "Fast emission measurement in time domain," in *Proceedings of the 14th International Zurich Symposium On Electromagnetic Compatibility (EMC '01)*, Zurich, Switzerland, 2001.
- [28] F. Krug, D. Mueller, and P. Russer, "Signal processing strategies with the TDEMI measurement system," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 5, pp. 1402–1408, 2004.
- [29] G. B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107–122, 2011.
- [30] G. B. Huang, L. Chen, and C. K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.
- [31] S. S. Haykin, *Neural Networks: A Comprehensive foundation*, Prentice Hall, 1999.
- [32] M. T. Hagan, *Neural Network Design*, Brooks/Cole, 1996.
- [33] C. M. Bishop, *Neural Networks for Pattern Recognition*, The Clarendon Press Oxford University Press, New York, NY, USA, 1995.

Research Article

Model for Microcirculation Transportation Network Design

Qun Chen and Feng Shi

School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

Correspondence should be addressed to Qun Chen, chenqun631@csu.edu.cn

Received 2 August 2012; Revised 8 November 2012; Accepted 21 November 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Q. Chen and F. Shi. 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 idea of microcirculation transportation was proposed to shunt heavy traffic on arterial roads through branch roads. The optimization model for designing micro-circulation transportation network was developed to pick out branch roads as traffic-shunting channels and determine their required capacity, trying to minimize the total reconstruction expense and land occupancy subject to saturation and reconstruction space constraints, while accounting for the route choice behaviour of network users. Since micro-circulation transportation network design problem includes both discrete and continuous variables, a discretization method was developed to convert two groups of variables (discrete variables and continuous variables) into one group of new discrete variables, transforming the mixed network design problem into a new kind of discrete network design problem with multiple values. The genetic algorithm was proposed to solve the new discrete network design problem. Finally a numerical example demonstrated the efficiency of the model and algorithm.

1. Introduction

Urban microcirculation transportation is an impersonate noun borrowed from human blood circulation system [1]. In the blood microcirculation system, blood flows from arterioles to microcirculation vessels and then flows from the microcirculation vessels back to venules. Similar to the blood microcirculation, traffic microcirculation can be defined as traffic flows from arterial roads to branch roads (microcirculation roads) and then flows from branch roads back to arterial roads. In general, most of vehicles run on the arterial roads, so arterial roads usually become very congested at peak hours. If microcirculation transportation network is designed around “jam points” of arterial roads, traffic on arterial roads can be shunted, and part of vehicles can go through the microcirculation roads (branch roads).

In reality, owing to narrow road surface and complicated functions, some branch roads are primarily for in-area traffic (e.g., pedestrian, bicycles, and few vehicles) [2, 3] and

do not have the ability of shunting traffic of arterials. For the purpose of traffic shunting, the microcirculation road systems need to be designed. Some branch roads with good conditions need to be picked out for reconstruction. So, there are two problems to solve: one is to determine which branch roads are picked out as traffic-shunting channels; the other is to determine the required capacity of these selected roads after reconstruction. Microcirculation transportation presented an effective and economical way for reducing traffic congestions because it does not need to add new roads but utilizes the existing branch roads to shunt traffic. Recently, in China, some big cities like Beijing and Kunming have established the microcirculation transportation systems in some congested segments.

Microcirculation transportation network design problem belongs to the family of network design problems (NDPs). The NDP is normally formulated as a mathematical program with equilibrium constraints (MPEC) in which the planner aims to define modifications to a network so as to optimize an objective function, whilst considering the response of travellers to the changes following an equilibrium condition. Often, the travellers' responses are assumed to follow Wardrop's user equilibrium condition (deterministic UE). Typical models for the NDP under DUE have been developed by Tobin and Friesz [4], Yang et al. [5], and Chiou [6]. Users' route choice behaviour is also usually characterized by the stochastic user equilibrium (SUE) [7]. Davis [8] and Uchida et al. [9] extended the NDP under the DUE to the SUE case.

The NDP is usually classified into three categories: the discrete network design problem (DNDP), the continuous network design problem (CNDP), and the mixed network design problem (MNDP) that combines both CNDP and DNDP in a network. The DNDP deals with the selection of the optimal locations of new links to be added and is normally applied in the design of new road systems. Leblance [10], Chen and Alfa [11], Gao et al. [12], and Jeon et al. [13] researched DNDP and developed mathematical models and solution algorithms. The CNDP determines the optimal capacity enhancement for a subset of existing links and is especially suitable for the design of widening the existing roads. Abdulaal and LeBlanc [14], Friesz et al. [15], Meng et al. [16], Chiou [17], and Wang and Lo [18] researched CNDP and developed mathematical models and solution algorithms. Yang and Bell [19] provided a comprehensive review of the models and algorithms for the NDP, in which MNDP was mentioned. The MNDP is normally formulated as a nonlinear mixed integral bilevel programming problem that is very hard to solve. Luathep et al. [20] developed a mixed-integer linear programming approach for solving the MNDP.

Shi et al. [21] modelled one-way traffic organization in microcirculation transportation network. In addition, Shi et al. [22] presented a model for reconstruction of urban branch road, but it only considered the cost target and optimized improvements of all branch roads. In fact, the microcirculation network design is a two-stage problem: the first is to determine which branch roads are picked out as traffic-shunting channels (0-1 variables); the second is to determine the required capacity of these selected roads (continuous variables). Microcirculation transportation network design problem includes both 0-1 discrete variables and continuous variables, so it also can be considered as one of the MNDPs. But it is different from the previous MNDPs. The conventional MNDPs combine both DNDP and CNDP in a network; discrete variables (for new road links) and continuous variables (for modified road links) are independent and for respective problems. However, in the microcirculation transportation network design problem, it needs to firstly select road links to be reconstructed, and then the required capacity of these selected road links can be determined. So it is a two-stage planning problem, in which determination of discrete variables' values is prior to determination of continuous variables' values. It is more difficult

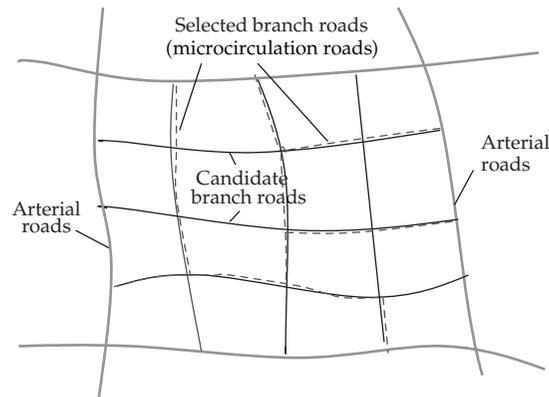


Figure 1: Arterial roads, candidate branch roads, and selected branch roads.

to solve than the conventional MNDPs. This paper presented a discretization way to convert two groups of variables (discrete variables and continuous variables) into one group of new discrete variables, and then the MNDP is transformed into a new kind of DNDP. The new DNDP is different from the conventional 0-1 DNDP because the variable of the new DNDP can take multiple values. The genetic algorithm was proposed to solve the new DNDP.

Moreover, compared with the conventional NDPs, the microcirculation transportation network design problem has different objectives. Microcirculation transportation network is a little local network whose objective is to shunt traffic from arterial roads. Because the size of network is small, passing time of vehicles is very short if the network is not congested, so the factor of travel time may be ignored in the model which is usually taken into account by the conventional NDPs. The main objective of the microcirculation network design problem is to minimize the total reconstruction expense under a saturation constraint. Also, the objective of minimizing land occupancy is taken into account to minimize interference with in-area residents. In addition, microcirculation transportation network design problem considers some other constraints, such as reconstruction space constraint and restriction for the number of cross-points of microcirculation roads and arterial roads.

The remainder of the paper is organized as follows. Section 2 presents the optimization model for designing the microcirculation transportation network. Section 3 introduces a discretization way to solve the model. In Section 4, a numerical example is given to demonstrate the application of the model and algorithm. The final section concludes the paper.

2. Optimization Model for Designing Microcirculation Transportation Network

In Figure 1, road network $N = (V, A \cup B)$, V is the set of all nodes, $n = |V|$. A is the set of arterial roads, and B is the set of candidate branch roads. $(q_{rs})_{n \times n}$ is traffic distribution between origins and destinations. For branch road $a (a \in B)$, $e(a)$ equals 1 if it is selected and 0 if not selected. All the selected branch roads construct the microcirculation transportation network. The existing capacity of each road is $C(a)$, $a \in A \cup B$. For the selected branch road $a (a \in B)$, its required capacity after reconstruction is $X(a)$, $a \in B$. Apparently, $X(a) \geq C(a)$, $a \in B$. $e(a)$, $X(a)$, $a \in B$ are the optimization variables.

In general, there are two road links with opposite directions between two adjacent nodes, and their capacities are usually the same.

Before reconstruction the branch roads do not have the ability of shunting traffic from arterial roads. They are for in-area traffic and often crowded with pedestrians and bicycles and even occupied by some other temporary facilities, and so would not have their original designed capacity unless they are cleaned up or reconstructed.

The main optimization objective is to minimize the total reconstruction cost which lies on the length and capacity of the reconstructed roads. If the capacity is improved more, the reconstruction expense will become more.

In addition, the objective of minimizing land occupancy (expressed as land use cost) should be taken into account to reduce interference with the area. Although microcirculation transportation can shunt arterials' traffic, the shunted traffic will interfere with residents' life inside the area and may cause environmental pollution. Reducing land occupancy (including road length and width) of microcirculation transportation can reduce interference scope. So, for those unselected branch roads, some management measures need to be taken to bar the traversing traffic, making traffic shunting restricted within the selected roads.

From the previous analysis, the cost function of candidate branch road a can be expressed as

$$P(a) = e(a)l(a)p(a) + e(a)l(a)h(a), \quad a \in B. \quad (2.1)$$

In (2.1), item 1 of the right side is reconstruction expense and item 2 is land use cost of microcirculation transportation. The optimization goal is to minimize $\sum_{a \in B} P(a)$, namely:

$$\min Y = \sum_{a \in B} P(a) = \sum_{a \in B} e(a)l(a)[p(a) + h(a)]. \quad (2.2)$$

$l(a)$ is the length of candidate branch road a , $p(a)$ is unit reconstruction expense, and $h(a)$ is unit land use cost:

$$p(a) = Z_a(X(a) - C(a)), \quad a \in B. \quad (2.3)$$

In (2.3), for branch road a , $p(a)$ is an increasing function of $X(a)$; namely, the required capacity is greater, the reconstruction expense is higher:

$$h(a) = T_a(X(a)), \quad a \in B. \quad (2.4)$$

Equation (2.4) implies that $h(a)$ is an increasing function of $X(a)$, because land use of microcirculation roads depends on their length and width, while land use cost $h(a)$ of unit length is decided by the road width which corresponds with the capacity after reconstruction ($X(a)$). In general, for branch road a , if the capacity $X(a)$ after reconstruction is greater, the road width should be greater, and so $h(a)$ becomes greater.

The constraints are as follows.

- (1) *Saturation constraint of arterial roads*: since the function of the microcirculation transportation network is to shunt traffic from arterial roads, the first target is to make the saturation of arterial roads less than an allowed value. But the saturation

of arterial roads should not be too small; otherwise, their capacity cannot be brought into full play. The key is to attain the goal of no more very congested:

$$u(a) = \frac{x(a)}{C(a)} < U(a), \quad a \in A, \quad (2.5)$$

where $u(a)$, $x(a)$, $a \in A$ are, respectively, the saturation and flow of arterial road a and $U(a)$ is the allowed saturation of arterial road a .

- (2) *Saturation constraint of branch roads*: the saturation of microcirculation branch roads should also be under an allowed value to avoid traffic congestions on branch roads and ensure the safety of pedestrians and bicycles on the branch roads:

$$v(a) = \frac{x(a)}{X(a)} < V(a), \quad a \in B, \quad (2.6)$$

where $v(a)$, $x(a)$, $a \in B$ are, respectively, the saturation and flow of branch road a and $V(a)$ is the allowed saturation of branch road a .

- (3) *Capacity constraint of branch roads*. Capacity enhancement of branch roads is affected by some actual conditions, such as land use restriction, building restriction and geological condition:

$$C(a) \leq X(a) \leq X_0(a), \quad a \in B, \quad (2.7)$$

where $X_0(a)$ is the available maximal capacity of branch road a after reconstruction.

- (4) *Restriction for the number of cross-points of microcirculation and arterial roads*: reducing the number of cross-points of microcirculation and arterial roads can reduce interference with arterial traffic. Microcirculation roads are for shunting arterial traffic and so generally have a relatively big traffic flow; signal controls normally need to be taken when they cross arterial roads. More signal-control intersections imply more of traffic delay on arterial roads (waiting time and the time needed for starting and braking of vehicles).

d_{i-j} denotes the number of cross-points of arterial road $(i - j)$ and microcirculation roads. Suppose d_{i-j} should not exceed D_{i-j} (maximal allowed value):

$$d_{i-j} \leq D_{i-j}. \quad (2.8)$$

$x(a)$, $a \in A \cup B$ is calculated via the user equilibrium (UE) traffic assignment model:

$$\min \sum_{a \in A \cup B} \int_0^{x(a)} t_a(w) dw \quad (2.9)$$

s.t.

$$\begin{aligned}
\sum_{k=1}^{L(r,s)} f_k^{rs} &= q_{rs}, \quad r, s = 1, 2, \dots, n, \\
x(a) &= \sum_{r=1}^n \sum_{s=1}^n \sum_{k=1}^{L(r,s)} f_k^{rs} \delta_{ak}^{rs}, \quad a \in A \cup B, \\
f_k^{rs} &\geq 0, \quad r, s = 1, 2, \dots, n, \quad k = 1, 2, \dots, L(r, s).
\end{aligned} \tag{2.10}$$

f_k^{rs} is the flow of path k between origin-destination (OD) pair (r, s) , $L(r, s)$ is the number of paths between OD pair (r, s) , and q_{rs} is traffic demand between OD pair (r, s) . $x(a)$ is the flow of link a . $\delta_{a,k}^{rs}$ equals 1 if link a is on path k between OD pair (r, s) , otherwise 0. t_a is travel time on link a . Here BPR (bureau of public road) link impedance function is applied:

$$t_a = t_{a0} \left[1 + \alpha \left(\frac{x(a)}{M(a)} \right)^\beta \right], \quad a \in A \cup B. \tag{2.11}$$

In (2.11), $M(a)$ is link capacity; for arterial roads, it is $C(a)$; for branch roads, it is $X(a)$. α, β are parameters, and BPR suggested that $\alpha = 0.15$, $\beta = 4$. t_{a0} is free-flow travel time of link a .

3. Solution Algorithms

There are two groups of variables $(e(a), X(a))$ in the above model, so the solution is very hard. But if the two groups of variables can be converted into one, then the solution will become much easier.

For branch road $a, a \in B$ (the existing capacity $C(a)$), its capacity enhancement via reconstruction can be discretized if it is selected. Let capacity enhancements be $0, \sigma, 2\sigma, 3\sigma, 4\sigma, \dots$, where σ denotes one added unit and 0 denotes that capacity enhancement is 0. This discretization way can accord with the real case. On the one hand, in reality, capacity enhancements via reconstruction are always discrete values instead of continuous; on the other hand, use of many discrete values is also able to reach the precision.

One group of new discrete variables $(\lambda(a))$ can be defined to convert two groups of variables $(e(a), X(a))$ into one group of variables $(\lambda(a))$:

$$\lambda(a) = \begin{cases} -1, & \text{do not select } a, \text{ here } e(a) = 0; \\ 0, & \text{select } a, \text{ here } e(a) = 1, \text{ the added value is } 0, X(a) = C(a); \\ 1, & \text{select } a, \text{ here } e(a) = 1, \text{ the added value is } \sigma, X(a) = C(a) + \sigma; \\ 2, & \text{select } a, \text{ here } e(a) = 1, \text{ the added value is } 2\sigma, X(a) = C(a) + 2\sigma; \\ 3, & \text{select } a, \text{ here } e(a) = 1, \text{ the added value is } 3\sigma, X(a) = C(a) + 3\sigma; \\ \vdots & \vdots \end{cases} \tag{3.1}$$

$\lambda(a), a \in B$ is the optimization variable. If $\lambda(a)$ is calculated, $e(a)$ and $X(a)$ can be obtained.

The real coded genetic algorithm is applied to solve the optimization model. The chromosome is made up of $\lambda(1), \lambda(2), \lambda(3), \dots$

Table 1: Traffic distribution between OD Pairs (unit: veh/h).

Node	1	2	3	4
1	0	800	2600	800
2	800	0	1000	2200
3	2600	1000	0	1000
4	800	2200	1000	0

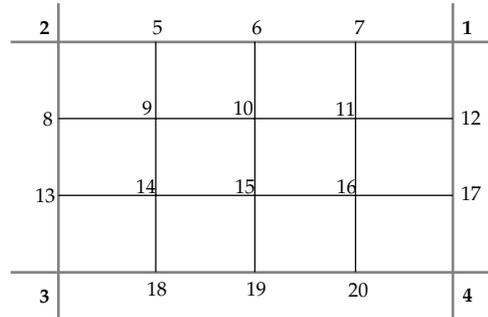


Figure 2: Original road network.

Steps of solving the model using genetic algorithm are as follows.

Step 1. Initialization: set population size (E), chromosome length (J), iteration number (g_{\max}), probability of crossover (P_c), and probability of mutation (P_m).

Step 2. Construct a fitness function: $F(m) = C_{\max} - O(m)$, where $F(m)$ is the fitness of individual m , $O(m)$ is the function value of individual m and C_{\max} is the estimated maximal function value. Randomly produce the initial population and set $g = 1$.

Step 3. Calculate link flows via UE traffic assignment model, and then calculate the fitness and excess over constraints of each individual. If $g = g_{\max}$, output the best individual; otherwise, turn to Step 4.

Step 4. Use roulette wheel selection operator based on ranking [23] to select the population of next generation. Feasible solutions rank from high to low by fitness, and then infeasible solutions rank from small to much by excess over constraints.

Step 5. According to probability of crossover (P_c), make multi-point crossover. Crossover points can be randomly selected without repeat. Variables between crossover points interchange alternately to produce two new individuals.

Step 6. According to probability of mutation (P_m), make single point mutation. Randomly produce an integer between $[-1 J]$ (J is the maximal value of $\lambda(a)$) to replace the current value of the variable. Set $g = g + 1$, and return to Step 3.

4. A Numerical Example

In Figure 2, the thick lines around the area denote arterial roads and the thin lines inside the area denote candidate branch roads. Each line includes two links with opposite directions and equal capacity. Traffic distribution during peak hours is in Table 1.

Table 2: Flows and saturations of arterial roads.

Link	Flow (veh/h)	Saturation (%)
1 → 7	2978	99.3
1 → 12	2959	98.6
2 → 5	2930	97.7
2 → 8	2993	99.8
3 → 13	2978	99.3
3 → 18	2998	99.9
4 → 17	2940	98.0
4 → 20	2936	97.9
5 → 2	2930	97.7
5 → 6	2930	97.7
6 → 5	2930	97.7
6 → 7	2978	99.3
7 → 1	2978	99.3
7 → 6	2978	99.3
8 → 2	2993	99.8
8 → 13	2978	99.3
12 → 1	2959	98.6
12 → 17	2940	98.0
13 → 3	2978	99.3
13 → 8	2978	99.3
17 → 4	2940	98.0
17 → 12	2940	98.0
18 → 3	2998	99.9
18 → 19	2936	97.9
19 → 18	2936	97.9
19 → 20	2936	97.9
20 → 4	2936	97.9
20 → 19	2936	97.9

The capacity of arterial road is 3000 (veh/h); the existing capacity of candidate branch road is 500 (veh/h). The length of each link is 1 km. Unit reconstruction cost function of branch roads is $p(a) = (X(a) - 500) \times 10^4$ (\$/km); unit land use cost function is $h(a) = (1/4)X(a) \times 10^4$ (\$/km). Road saturation should not exceed 1. The available maximal capacity of each branch road after reconstruction is 1000 (veh/h). t_{a0} of arterial roads is 1 min and that of branch roads is 1.1 min. $d_{1-2} \leq 1$, $d_{1-4} \leq 1$, $d_{3-4} \leq 1$, $d_{2-3} \leq 1$.

Let $\sigma = 100$; the solution set of $\lambda(a)$ is $\{-1, 0, 1, 2, 3, 4, 5\}$ since the available maximal capacity is 1000 and the existing capacity is 500. The selected branch roads are shown in Figure 3; the total cost is 8000×10^4 \$. Saturations of arterial and branch roads are all less than 1. Flows and saturations of arterial roads are in Table 2; capacities, flows, and saturations of the selected branch roads for constructing the microcirculation network are in Table 3.

Comparatively, if only arterial roads exist (without microcirculation road network), the saturation of arterial roads goes beyond 1 (Table 4).

5. Conclusions

This paper defined the concept of urban microcirculation transportation. Microcirculation transportation network is a little local network and can shunt traffic from arterial roads.

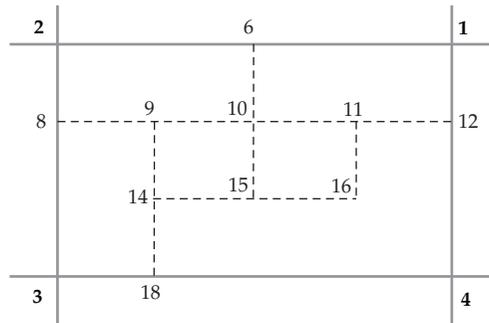


Figure 3: Microcirculation road network.

Table 3: Capacities, flows, and saturations of the selected branch roads.

Link	Capacity (veh/h)	Flow (veh/h)	Saturation (%)
6 → 10	1000	714	71.4
8 → 9	800	620	77.5
9 → 8	800	620	77.5
9 → 10	700	374	53.4
9 → 14	700	323	46.1
10 → 6	1000	714	71.4
10 → 9	700	374	53.4
10 → 11	700	424	60.6
10 → 15	500	101	20.2
11 → 10	700	424	60.6
11 → 12	800	686	85.8
11 → 16	500	262	52.4
12 → 11	800	686	85.8
14 → 9	700	323	46.1
14 → 15	600	360	60.0
14 → 18	800	683	85.4
15 → 10	500	101	20.2
15 → 14	600	360	60.0
15 → 16	500	262	52.4
16 → 11	500	262	52.4
16 → 15	500	262	52.4
18 → 14	800	683	85.4

Table 4: Saturations and flows.

Link	Flow (veh/h)	Saturation (%)
1 → 2	3278	109.3
1 → 4	3297	109.9
2 → 1	3278	109.3
2 → 3	3303	110.1
3 → 2	3303	110.1
3 → 4	3321	110.7
4 → 1	3297	109.9
4 → 3	3321	110.7

Through the microcirculation transportation network design model in this paper, the branch roads as traffic-shunting channels and their required capacity after reconstruction can be decided.

Since microcirculation transportation network design problem includes both discrete variables and continuous variables, this paper developed a discretization method to convert two groups of variables (discrete variables and continuous variables) into one group of new discrete variables, transforming the solution of MNDP into the solution of a new kind of DNDP with multiple values, and the genetic algorithm was proposed to solve the new DNDP.

A numerical example demonstrated the application of the model and algorithm and compared the results with or no microcirculation transportation network. The method and model proposed in this paper provided a new effective way for solving urban traffic congestions.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant no. 50908235) and the Hunan Province Social Science Fund (Grant no. 12YBB274).

References

- [1] F. Shi, E. H. Huang, and Y. Z. Wang, "Study on the functional characteristics of Urban transportation micro-circulation system," *Urban Studies*, vol. 15, no. 3, pp. 34–36, 2008.
- [2] W. H. Wang, Q. Cao, K. Ikeuchi, and H. Bubb, "Reliability and safety analysis methodology for identification of drivers' erroneous actions," *International Journal of Automotive Technology*, vol. 11, no. 6, pp. 873–881, 2010.
- [3] W. H. Wang, H. W. Guo, K. Ikeuchi, and H. Bubb, "Numerical simulation and analysis procedure for model-based digital driving dependability in intelligent transport system," *KSCE Journal of Civil Engineering*, vol. 15, no. 5, pp. 891–898, 2011.
- [4] R. L. Tobin and T. L. Friesz, "Sensitivity analysis for equilibrium network flow," *Transportation Science*, vol. 22, no. 4, pp. 242–250, 1988.
- [5] H. Yang, Q. Meng, and M. G. H. Bell, "Simultaneous estimation of the origin-destination matrices and travel-cost coefficient for congested networks in a stochastic user equilibrium," *Transportation Science*, vol. 35, no. 2, pp. 107–123, 2001.
- [6] S. W. Chiou, "TRANSYT derivatives for area traffic control optimisation with network equilibrium flows," *Transportation Research B*, vol. 37, no. 3, pp. 263–290, 2003.
- [7] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.
- [8] G. A. Davis, "Exact local solution of the continuous network design problem via stochastic user equilibrium assignment," *Transportation Research B*, vol. 28, no. 1, pp. 61–75, 1994.
- [9] K. Uchida, A. Sumalee, D. Watling, and R. Connors, "A study on network design problems for multi-modal networks by probit-based stochastic user equilibrium," *Networks and Spatial Economics*, vol. 7, no. 3, pp. 213–240, 2007.
- [10] L. J. Leblanc, "An algorithm for the discrete network design problem," *Transportation Science*, vol. 9, no. 3, pp. 183–199, 1975.
- [11] M. Chen and A. S. Alfa, "Network design algorithm using a stochastic incremental traffic assignment approach," *Transportation Science*, vol. 25, no. 3, pp. 215–224, 1991.
- [12] Z. Y. Gao, J. J. Wu, and H. J. Sun, "Solution algorithm for the bi-level discrete network design problem," *Transportation Research B*, vol. 39, no. 6, pp. 479–495, 2005.
- [13] K. Jeon, J. S. Lee, S. Ukkusuri, and S. T. Waller, "Selectorecombinative genetic algorithm to relax computational complexity of discrete network design problem," *Transportation Research Record*, no. 1964, pp. 91–103, 2006.
- [14] M. Abdulaal and L. J. LeBlanc, "Continuous equilibrium network design models," *Transportation Research B*, vol. 13, no. 1, pp. 19–32, 1979.

- [15] T. L. Friesz, H. J. Cho, N. J. Mehta, R. L. Tobin, and G. Anandalingam, "Simulated annealing approach to the network design problem with variational inequality constraints," *Transportation Science*, vol. 26, no. 1, pp. 18–26, 1992.
- [16] Q. Meng, H. Yang, and M. G. H. Bell, "An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem," *Transportation Research B*, vol. 35, no. 1, pp. 83–105, 2001.
- [17] S. W. Chiou, "Bilevel programming for the continuous transport network design problem," *Transportation Research B*, vol. 39, no. 4, pp. 361–383, 2005.
- [18] D. Z. W. Wang and H. K. Lo, "Global optimum of the linearized network design problem with equilibrium flows," *Transportation Research B*, vol. 44, no. 4, pp. 482–492, 2010.
- [19] H. Yang and M. G. H. Bell, "Models and algorithms for road network design: a review and some new developments," *Transport Reviews*, vol. 18, no. 3, pp. 257–278, 1998.
- [20] P. Luathep, A. Sumalee, W. H. K. Lam, Z. C. Li, and H. K. Lo, "Global optimization method for mixed transportation network design problem: a mixed-integer linear programming approach," *Transportation Research B*, vol. 45, no. 5, pp. 808–827, 2011.
- [21] F. Shi, E. H. Huang, Q. Chen, and Y. Z. Wang, "Optimization of one-way traffic organization for urban micro-circulation transportation network," *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 4, pp. 30–35, 2009.
- [22] F. Shi, E. H. Huang, Q. Chen, and Y. Z. Wang, "Bi-level programming model for reconstruction of urban branch road network," *Journal of Central South University of Technology*, vol. 16, no. 1, pp. 172–176, 2009.
- [23] X. P. Wang and L. M. Cao, *Theory, Application and Software Implementation*, Xi'an Jiaotong University Press, Xi'an, China, 2002.

Research Article

Adaptability Analysis of Service Facilities in Transfer Subway Stations

Liya Yao,¹ Lishan Sun,² Wuhong Wang,¹ and Hui Xiong¹

¹ Department of Transportation Engineering, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China

² Key Laboratory of Traffic Engineering, Beijing University of Technology, 100 Pingleyuan, Chaoyang District, Beijing 100124, China

Correspondence should be addressed to Liya Yao, yaoliya@bit.edu.cn

Received 19 July 2012; Accepted 4 November 2012

Academic Editor: Huimin Niu

Copyright © 2012 Liya Yao 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.

Service capability and matching degree of transfer facilities are directly related to the operational efficiency and safety of a subway station. Owing to differences in planning and construction, the transfer subway stations in developing countries have some defects in facility size and serviceability, which cause a decline in service performance, operation efficiency, and security level. In order to solve the problems, traffic investigations were conducted on the form, size, and operation status of several typical transfer subway facilities. The service facilities were classified within a subway station in this research by considering service objects, service forms, service functions, and several other features. In addition, pedestrian behavior and pedestrian flow characteristics in different service facilities were analyzed in detail. The research results are deemed meaningful for the optimization of service facilities in subway stations and for the development of urban pedestrian transportation systems.

1. Background

As important modes of public transportation, the subway has become part of daily commuter life in China, especially in Beijing. During the past half century, a great amount of changes have taken place both in the construction conditions and design standards of subways. Moreover, the design scale of transfer facilities and the connecting modes in each subway line have become dramatically different. Along with the rapid construction of the subway network in Beijing, subway lines that were constructed under different standards during different periods are now linked at transfer hubs. As a result, unreasonable infrastructures and mismatched capacities for subway facilities led to pedestrian interweaving and jamming. Several research interests on subway facilities have emerged, including service level of

various transfer facilities, subway facilities match, characteristics of pedestrian flow at different subway facilities, and efficiency of a subway station.

The method of adaptability has been broadly used in the biological and engineering domains in the past years [1, 2]. The service level of facilities was first recommended in the *Highway Capacity Manual 2000* (HCM2000) [3]. Pedestrian facility service level is divided based on the quantitative observation of pedestrian parameters, such as velocity, density, and flow. Fruin [4] proposed an algorithm for calculating the service level in pedestrian facilities, including footways, stairs, and queuing areas. The division of the level is based on pedestrian velocity, space, and conflict probability. Domestic scholars have obtained preliminary results in the characterization of pedestrians and facilities [5–7].

Indexes such as walking time, speed, waiting time, walking distance, choice of walking route, selection of stairways, or elevators were studied for the traffic characteristics of a single pedestrian. Lam et al. [8, 9], Young [10], and Paul [11, 12] obtained pedestrian walking speed at different facilities for different pedestrian characteristics. Hoskin [13] stated the definition of facility volume. Ayano et al. [14] studied the influence of different pedestrian volumes on actual capacity by cellular automata, which yielded the maximum pedestrian volume in a single direction, in two directions, and in four directions. Thompson [15] described pedestrian behavior and divided it into five types: route-selecting behavior, crash-avoiding behavior, pursuing behavior, arriving and leaving behavior, and lingering behavior. Hine [16] showed that traffic circumstance has a great influence on pedestrian behavior. Lam et al. [17–19] and Delft [20] studied the rule of pedestrian flow characteristics and route selection at subway stations and surveyed and simulated facility service level. Wang et al. [21, 22] conducted numerous studies on the theory of behavior simulation. Christian et al. [23] studied the distribution of walking speed and route selection behavior. Helbing et al. [24] performed numerous experiments on emergency evacuation.

The characteristics of pedestrian flow, pedestrian volume, density, and average speed were studied weightily. Hughes [25] revealed the maximum speed and density of pedestrian flow. The relationship between pedestrian speed and density was studied by Ando et al. [26], Thompson and Marchant [27], Hughes [28], Hankin and Wright [29], and so on. Cheung and Lam [18] and Tanaboriboon et al. [30] studied the relationship between pedestrian volume and density. However, existing methods fail to consider the matching of various transfer facilities and to evaluate the adaptability of facilities with pedestrian flow characteristics. This paper conducts an in-depth study on the adaptability of subway transfer facilities with a consideration to pedestrian flow characteristics.

The remainder of the paper is structured as follows. Section 2 briefly discusses the current status of subway lines and the classification of transfer facilities in Beijing. Section 3 introduces the basic methods of the survey and the behavior indicators of pedestrian flow on transfer facilities. Section 4 presents the relationship among density, volume, and speed on stairways, corridors, platforms, and so on. Section 5 presents the adaptability analysis. Finally, Section 6 presents the conclusions.

2. Transfer Facility Classification

No references for a uniform design standard concerning the early years of subway construction in China are available. Subway designers had to refer to specifications abroad, which resulted in the difference in transfer facility capabilities among different subway lines. Arbitrary design became a serious problem in the first decades of subway construction

Table 1: Design parameter comparison between the new and old standards of China.

Minimum width	GB50157-92	GB50157-2003
Corridor	2.5 m	2.4 m
Stair	2 m	2.4 m
Island platform	8 m	8 m
Interisland platform	2 m	2.5 m
Column-free side platform	3.5 m	2.5 m
Column side platform	3 m	3.5 m

in China. In 1992, the first national standard for subway design (underground railway design standard (GB50157-92)) was issued by the Ministry of Housing and Urban-Rural Development of China. This design standard was replaced by a completely revised version in 2003: subway design standard (GB50157-2003). Numerous differences exist between these two standards, especially in the design parameter requirements of transfer facilities. Design parameter comparisons are given in Table 1 [31, 32].

According to the national standard, plenty of existing studies classified transfer facilities as traffic service facilities, information service facilities, assistant service facilities, and so on. In this paper, traffic service facilities, such as stairs, corridors, and platforms are used as examples to analyze the adaptability of a facility.

3. Research on Characteristics of Transfer Facilities

3.1. Investigation Scheme Design

Pedestrian behavior depends largely on the adaptability of transfer facilities. We investigated pedestrian behavior in typical service facilities of transfer subway stations in Beijing in December, 2011, to obtain typical pedestrian characteristics at various facilities, with reference to the classification of the aforementioned transfer facilities.

3.1.1. Investigation Time

The great pressure of pedestrian flow in transfer facilities and the significance of pedestrian flow characteristics presented in peak periods were taken into account when choosing investigation times; thus, the morning peak hour (8:00–9:00) and evening peak hour (17:00–18:00) were selected.

3.1.2. Investigation Area

The selected stations for investigation were Jianguomen Subway Station, where early-built lines are connected; Dongzhimen Subway Station, where early-built, midterm-built, and newly-built lines are connected, Haidianhuangzhuang Subway Station, where recently-built lines are connected (Figure 1). Our investigations focused on the different forms of staircases, corridors, and platforms to analyze the adaptability of transfer facilities in different eras under different criteria.

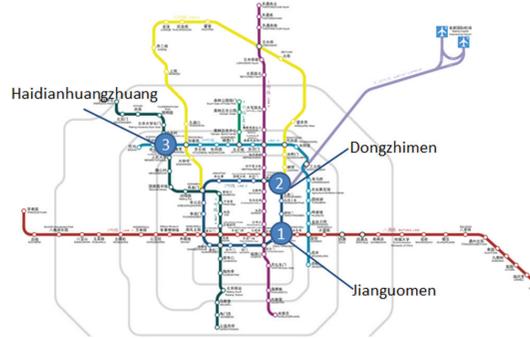


Figure 1: Distribution of subway lines and subway stations investigated in Beijing.

Table 2: Design parameters of main transfer facilities of subway stations in different years.

Construction year	Staircase width/m	Corridors width/m	Platform width/m	Reference standard
1960~1990	3.5	3.5	7	Soviet standard
1991~2003	2.5~5	3.5	7	GB50157-92
2003~2010	4	4	10	GB50157-2003

3.1.3. Data Collection Method

Artificial methods and video data collection of pedestrian behavior characteristics were adopted in the investigation with regards to heavy pedestrian flow transfer and high pedestrian walking speed.

3.2. Characteristics of Transfer Facilities

3.2.1. Width Parameters of Main Transfer Facilities

Service facilities of various subway stations differ in size, connecting characteristics, location, service level, and so on. The parameters of recently constructed service facilities are better compared with earlier ones (Table 2).

3.2.2. Characteristics of Pedestrian Behavior at Main Transfer Facilities

Staircases, corridors, and platforms are not only the main function facilities for pedestrian transfer within the spatial dimensions but also important routes for pedestrian emergency evacuation. The following indicators were selected after taking into account the characteristics of the distribution and interweaving of high-density pedestrian flow: average walking speed, average space, average flow rate, and peak 15-min volume of pedestrian traffic flow. These indicators were used to describe the characteristics of various types of transfer facilities.

Average walking speed (V): average pedestrian walking distance in a specified direction per unit time is generally expressed in m/s.

Table 3: Indicators describing characteristics of pedestrian flow at stairways and corridors.

Subway station	Facility	Average width (m)	Average space (m ² /per)	Average speed (m/s)	Average flow rate (per/min · m)	Pek-15min volume (per/15 min)
Jianguomen	Staircase	2.5	0.59	0.42	37.65	1412
	Corridor	3.5	1.76	1.22	37.45	1966
Dongzhimen	Staircase	3.5	1.00	0.46	29.04	1525
	Corridor	4.0	1.21	1.01	53.25	3195
Haidianhuan-gzhuang	Staircase	4.0	1.23	0.51	25.05	1503
	Corridor	3.0	1.98	1.17	42.00	1890

Table 4: Indicators describing characteristics of pedestrian flow at platforms.

Platform	Jianguomen		Dongzhimen		Haidianhuangzhuang	
	Line 1	Line 2	Line 2	Line 13	Line 4	Line 10
Average pedestrian space m ² /person	0.31	0.42	0.38	0.59	0.79	0.72

Average pedestrian space (S): average area provided for each pedestrian by the transfer facilities is calculated by Formula (3.1). Average pedestrian space is the reciprocal of pedestrian density, generally expressed as m²/person:

$$S = \frac{V \times W_E}{Q}, \quad (3.1)$$

where W_E is effective width of facilities, m; Q is pedestrian flow volume, person/h.

Peak 15-min pedestrian volume (Q_{15}): maximum pedestrian flow volume in 15 minutes, generally expressed as person/15 min.

Pedestrian flow rate (P): the number of pedestrians passing through per unit width of a certain section of the facilities per unit time is generally expressed as person/min·m:

$$P = \frac{Q_{15}}{15 \times W_E}. \quad (3.2)$$

Pedestrians naturally slow down while arriving at stairs or corridors because of changes in facility types, leading to serious congestions or queues in front of stairs or corridors. The characteristics of moving pedestrian flow at stairways and corridors, such as average speed, average flow rate, and peak 15-min pedestrian volume, are described by the survey data (Table 3).

Subway station platforms provide a place for pedestrians to realize traffic functions, such as waiting, getting on and disembarking from the train, evacuating, and transferring. Intertwined pedestrian flow frequently occurs on the platform in all directions, causing significant changes in pedestrian density and walking space. The characteristics of pedestrian flow at platforms are described by the average pedestrian space in this paper (Table 4).

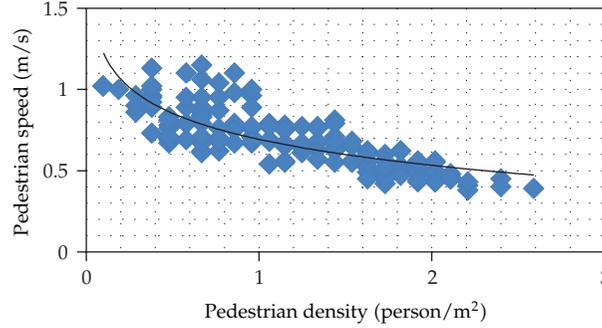


Figure 2: Relationship between pedestrian density and speed at stairways.

4. Relationship among Density, Volume, and Speed at Typical Service Facilities

4.1. Stairways

Pedestrian characteristics at stairways are not only influenced by the stairway itself but also by pedestrian gravity and interaction. Based on video observations, we found that walking speed is decreased when going downstairs. The down stairway located in the west of the Haidianhuangzhuang Subway Station was taken as an example. The investigation point was very crowded because all pedestrians converge at this location from the south and north of the transfer corridors of subway Line 10.

The data obtained at the down stairway area yielded the scatter diagram between pedestrian density and speed. Furthermore, the curve between density and speed should be fitted. Comparisons of all regression equations reveal that the logarithm model is best used in the regression analysis of pedestrian density and speed at stairways. The scatter diagram and regression curve are shown in Figure 2.

The logarithm equation obtained by fitting is expressed as

$$y = 0.23 \ln(x) + 0.6928. \quad (4.1)$$

Pedestrian flow volume is the number of pedestrians passing through per unit width, which is calculated by the product of pedestrian speed and density. Based on the calibrated parameters, the maximum pedestrian flow volume and the corresponding pedestrian density of the unit width were obtained and called pedestrian flow rate.

The analyzed data obtained at the down stairway yielded the scatter diagram between pedestrian density and volume. Furthermore, the curve between density and speed should be fitted. Comparisons of all regression equations reveal that the quadratic model is best used for the regression analysis of pedestrian density and volume at down stairways (Figure 3).

The quadratic equation obtained by fitting is expressed as

$$y = 0.2473x^2 + 0.9796x + 0.0065. \quad (4.2)$$

The maximum pedestrian flow rate and the corresponding pedestrian density were calculated according to the parameters obtained by fitting. Based on Figure 3, the maximum

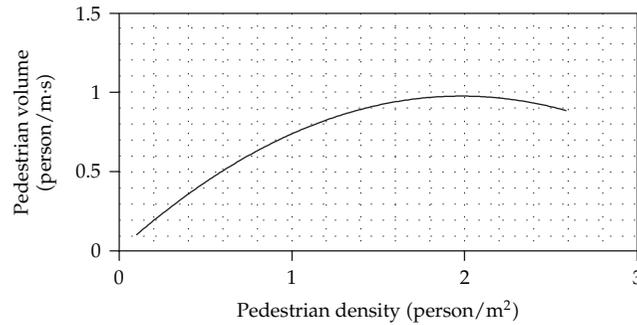


Figure 3: Relationship between pedestrian density and volume at stairs.

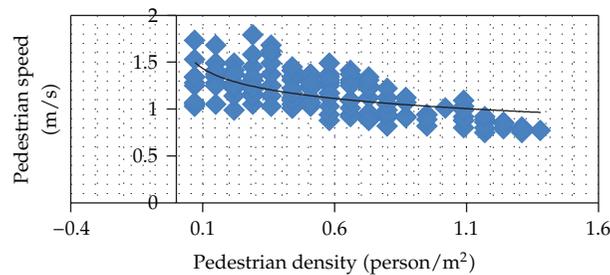


Figure 4: Relationship between pedestrian density and speed at corridors.

pedestrian flow rate is 0.97 person/m·s, and the corresponding pedestrian density is 1.98 person/m².

The curve begins to embody the falling trend at this point, which is the passing capacity of the stairways. As this trend develops, the stairways will become unstable for pedestrian flow, which means that pedestrian volume exceeds the capacity of the stairways. Queuing phenomenon will occur, and pedestrian volume will decrease instead of increase until the pedestrian volume becomes zero, and the corresponding density is called jamming density.

4.2. Corridor

Similar to stairways, corridors also belong to passing facilities where pedestrian characteristics are influenced by connecting service facilities, such as platforms. The corridor in the south channel from subway Line 10 to Line 4 of Haidianhuangzhuang Subway Station was taken as an example. From the investigation and video observations of the selected corridor, we found that pedestrian flow is more fluent at the corridor compared with the selected stairway. The fluent pedestrian flow is mainly because of the square-built connecting area, where the phenomenon of crowding is infrequent.

The analyzed data obtained at the corridor yielded the scatter diagram between pedestrian density and speed. Furthermore, the curve between density and speed should be fitted. Comparisons of all regression equations revealed that the logarithm model is best used in the regression analysis of pedestrian density and speed at corridors. The scatter diagram and regression curve are shown in Figure 4.

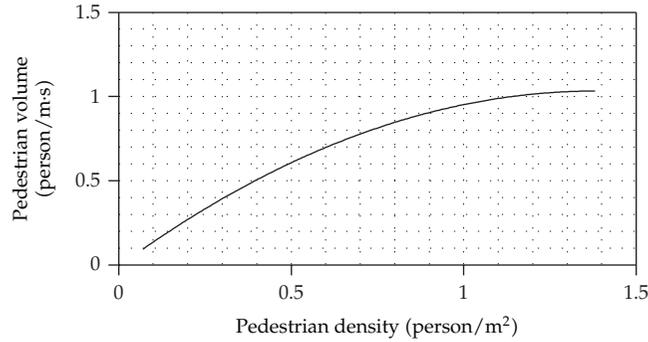


Figure 5: Relationship between pedestrian density and volume at corridors.

The logarithm functional relation obtained by fitting is expressed as

$$y = 0.178 \ln(x) + 1.0224. \quad (4.3)$$

The analyzed data obtained at the corridors yielded the scatter diagram between pedestrian density and volume. Furthermore, the curve between density and speed should be fitted. Comparisons of all regression equations reveal that the quadratic model is best used in the regression analysis of pedestrian density and volume at corridors (Figure 5).

The quadratic functional relation obtained by fitting is expressed as

$$y = 0.5417x^2 + 1.5025x + 0.0087. \quad (4.4)$$

The maximum pedestrian flow rate and the corresponding pedestrian density were calculated according to the parameters obtained by fitting. Based on Figure 5, the maximum pedestrian flow rate is 1.04 person/m·s, and the corresponding pedestrian density is 1.42 person/m².

The curve is smooth at this point, and the theoretic maximum value is still not reached. However, calculating the service level of such points is also meaningful because of the potential service capacity at such points.

4.3. Platform

Pedestrian characteristics, such as flow density and speed, are influenced by connecting service facilities. The platform in the east of the Haidianhuangzhuang Subway Station at Line 4 was taken as an example. According to the investigation, pedestrian flow density is largest at the point where the platform and stairs connect.

The analyzed data obtained at the platform area yielded the scatter diagram between pedestrian density and speed. Furthermore, the curve between density and speed should be fitted. Comparisons of all regression equations reveal that the logarithm model is best used in the regression analysis of pedestrian density and speed at platforms. The scatter diagram and regression curve are shown in Figure 6.

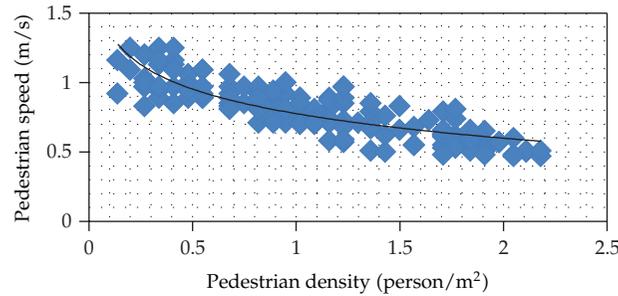


Figure 6: Relationship between pedestrian density and speed at platform.

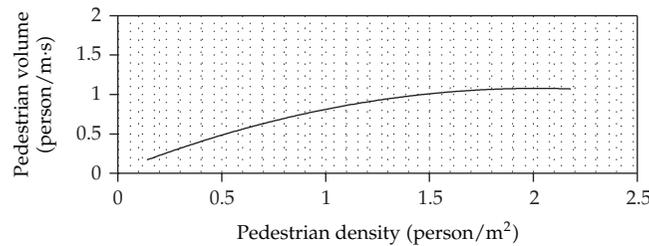


Figure 7: Relationship between pedestrian density and volume at platform.

The logarithm functional relation obtained by fitting is expressed as

$$y = 0.255 \ln(x) + 0.7758. \quad (4.5)$$

The analyzed data obtained at the platforms yielded the scatter diagram between pedestrian density and volume. Furthermore, the curve between density and volume should be fitted. Comparisons of all regression equations reveal that the quadratic model is best used in the regression analysis of pedestrian density and volume at the platforms (Figure 7).

The quadratic functional relation obtained by fitting is expressed as

$$y = 0.2576x^2 + 1.0387x + 0.0305. \quad (4.6)$$

The maximum pedestrian flow rate and the corresponding pedestrian density were calculated according to the parameters obtained by fitting. Based on Figure 7, the maximum pedestrian flow rate is 1.08 person/m·s, and the corresponding pedestrian density is 2.02 person/m².

The curve begins to embody the falling trend at this point. The curve is stable although the decreasing pedestrian flow at the platform is not as obvious as that at the stairs.

5. Adaptability Analysis of Transfer Facilities

The adaptability of transfer facilities can be defined as the adaptation degree of the facilities faced with pedestrian volume based on physical properties and abilities. Based on the design parameters of the studies, such as the connecting layout and usage status, this paper analyzed

Table 5: Service level of main transfer facilities in peak hours.

Subway station	Jianguomen			Dongzhimen			Haidianhuangzhuang		
	Stairs	Corridor	Platform	Stairs	Corridor	Platform	Stairs	Corridor	Platform
v/c	0.77	0.50	—	0.59	0.71	—	0.51	0.56	—
Service level*	E	D	D	D	E	D	C	D	C

Service level*: according to *Highway Capacity Manual 2000* (HCM2000).

the service level of subway station facilities. Furthermore, the capability of the facilities to meet the transfer needs of pedestrians was also determined. The service level of the selected transfer facilities in the three subway stations at peak hour was obtained according to investigation data of pedestrian flow characteristics at certain facilities in the subway stations (Table 5). The service level was obtained using the aforementioned method and the existing standard on facility service level as a reference.

Based on Table 5, the service level of various transfer facilities is low during peak hours and the service ability of most facilities is able to meet the basic transfer demand of pedestrians. The stairs at the Jianguomen Station and the corridors at the Dongzhimen Station are the key facilities in improving the service level of the subway station. The main reasons that lead to e-class are as follows.

(1) Imbalance between the service level of old facilities at subway stations and the increasing transfer demand of pedestrians: the subway stations constructed in the 1970s, such as the Jianguomen Station, mainly referred to the subway design standard of the former Soviet Union, which mostly adhered to combat readiness and evacuation protocols, and only took into account a few concepts of traffic function. Along with the rapidly increasing pedestrian volume, existing transfer facilities are not prepared to meet the traffic demand. In addition, the adaptability of the facilities is poor. For example, the Jianguomen Subway Station has four-segmented stairways that connect subway Line 1 to subway Line 2. Excessive stairways promote pedestrian queuing before the stairways, which even spreads to the station platform during peak hours. Pedestrian queuing not only occupies platform space but also decreases transfer efficiency.

(2) Unbalanced service level of transfer facilities built in different periods: because of different standards referred to during different periods, subway facilities face balance problems, which occur when two or more subway lines connect to the same subway station. Three subway lines are connected at Dongzhimen Station: subway Line 2, built in 1970; subway Line 13, built in 1999; the airport express line, built in 2008. Differences in scale, size, and layout exist among these transfer facilities. The sudden change in the size or layout of the transfer facility makes crowding and interweaving of pedestrians at the corridor entrances common.

6. Conclusions

The transfer facilities of numerous subway stations in Beijing obviously vary in terms of form, scale, and layout because of difference in design ideas and standards. A systematic analysis of the adaptability of transfer facilities in transfer subway stations is important in enhancing the overall performance of Beijing's public transport system. This paper conducted an adaptability analysis framework using commonly available variables in response to the absence of a pedestrian behavior model for transfer facilities.

This paper investigated and analyzed the design scale, layout form, and operating status of typical transfer subway stations in Beijing. The service adaptability of transfer facilities was studied based on the investigation data. This paper will enable subway designers to match different transfer facilities effectively and improve the operation performance of transfer subway stations according to comparisons between pedestrian behavior and facility adaptability. The main conclusions are as follows.

- (1) Obvious differences exist in the type and size of transfer facilities in subway stations constructed in different periods and according to different criteria. Differences in design standards are the most important factors that lead to the low operational efficiency of subway networks.
- (2) Old facilities in subway stations are unable to cater to heavy pedestrian volume in Beijing. Heavy pedestrian volume is reflected by a sudden increase in pedestrian density, reduced walking speed, and the frequent occurrence of interweaving pedestrians at the connecting points of old facilities.

Acknowledgments

This research was supported by National Nature Science Foundation of China (no. 51108028), Beijing Municipal Natural Science Foundation (no. 8122009), Programme of Introducing Talents of Discipline to Universities (no. B12022), and 973 Program (no. 2012CB725403).

References

- [1] R. Cattalani, D. Corsini, L. Posteraro, M. Agosti, and M. Saccavini, "The Italian version of the Mayo-Portland adaptability inventory-4. A new measure of brain injury outcome," *European Journal of Physical and Rehabilitation Medicine*, vol. 45, no. 4, pp. 513–519, 2009.
- [2] L. S. Kang and B. C. Paulson, "Adaptability of information classification systems for civil works," *Journal of Construction Engineering and Management*, vol. 123, no. 4, pp. 419–426, 1997.
- [3] National Research Council (NRC), *Highway Capacity Manual*, National Research Council (NRC), Washington, DC, USA, 2000.
- [4] J. J. Fruin, *Pedestrian Planning and Design*, Metropolitan Association of Urban Designers and Environmental Planners, New York, NY, USA, 1971.
- [5] S. H. Cao, Z. Z. Yuan, and D. Zhao, "Queuing mechanism of passengers at exit stairs of urban rail transit," *Journal of Jilin University*, vol. 39, no. 6, pp. 1463–1468, 2009.
- [6] C. Ran and D. Liyun, "Observations and preliminary analysis of characteristics of pedestrian traffic in Chinese metropolis," *Journal of Shanghai University*, vol. 11, no. 1, pp. 93–97, 2005.
- [7] H. F. Jia, L. L. Yang, M. Tang, and D. Meng, "Micro-characteristics and modeling requirement of pedestrians in multi-modal transport hub," *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 2, pp. 17–22, 2009.
- [8] W. H. K. Lam and C. Y. Cheung, "Pedestrian speed/flow relationships for walking facilities in Hong Kong," *Journal of Transportation Engineering*, vol. 126, no. 4, pp. 343–349, 2000.
- [9] W. H. K. Lam, J. Y. S. Lee, and C. Y. Cheung, "A study of the bi-directional pedestrian flow characteristics at Hong Kong signalized crosswalk facilities," *Transportation*, vol. 29, no. 2, pp. 169–192, 2002.
- [10] S. B. Young, "Evaluation of pedestrian walking speeds in airport terminals," *Transportation Research Record*, no. 1674, pp. 20–26, 1999.
- [11] J. L. Paul, "A personal perspective on research, consulting and codes standards development in fire-related human behavior, 1969–1999, with an emphasis on space and time factors," *Fire and Materials*, vol. 23, no. 1, pp. 265–272.
- [12] J. L. Paul, "Suggestions on evacuation models and research questions," in *Proceedings of the 3rd International Symposium on Human Behavior in Fires*, pp. 23–33, Inter-science communications, London, UK, 2004.

- [13] K. Hoskin, *Fire protection and evacuation procedures of Stadia Venues in New Zealand [Ph.D. dissertation]*, Department of Civil Engineering University of Canterbury, Christchurch, New Zealand, 2004.
- [14] N. Ayano, I. Mai, and H. Hisamoto, "Uni-directional pedestrian movement model based on Voronoi diagrams," in *Proceedings of the 8th International Symposium on Voronoi Diagrams in Science and Engineering*, pp. 123–126, 2011.
- [15] P. A. Thompson, *Developing new techniques for modeling crowd movement [Ph.D. dissertation]*, Edinburgh University of Edinburgh, 1994.
- [16] J. Hine, "Assessing the impact of traffic on behaviour and perceptions of safety using an in-depth interview technique," *Journal of Transport Geography*, vol. 4, no. 3, pp. 179–199, 1996.
- [17] C. Y. Cheung and W. H. K. Lam, "A study of the Bi-directional pedestrian flow characteristics in the Hong Kong mass transit railway stations," *Journal of Transportation Engineering*, vol. 2, no. 5, pp. 1607–1619, 1997.
- [18] C. Y. Cheung and W. H. K. Lam, "Pedestrian route choices between escalator and stairway in MTR stations," *Journal of Transportation Engineering*, vol. 124, no. 3, pp. 277–285, 1998.
- [19] J. Y. S. Lee and W. H. K. Lam, "Levels of service for stairway in Hong Kong underground stations," *Journal of Transportation Engineering*, vol. 129, no. 2, pp. 196–202, 2003.
- [20] W. Delft, *Modeling Passenger Flows in Public Transport Facilities*, Delft University, Delft, The Netherlands, 2004.
- [21] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research Part C-Emerging Technologies*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [22] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: a fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.
- [23] R. Christian, M. Thomas, S. Stefan, and B. Dietmar, "Can walking behavior be predicted? Analysis of calibration and fit of pedestrian models," *Transportation Research Record*, vol. 2264, no. 1, pp. 101–109, 2011.
- [24] D. Helbing, L. Buzna, A. Johansson, and T. Werner, "Self-organized pedestrian crowd dynamics: experiments, simulations, and design solutions," *Transportation Science*, vol. 39, no. 1, pp. 1–24, 2005.
- [25] R. L. Hughes, "A continuum theory for the flow of pedestrians," *Transportation Research Part B*, vol. 36, no. 6, pp. 507–535, 2002.
- [26] K. Ando, H. Ota, and T. Oki, "Forecasting the flow of people," *Railway Research Review*, vol. 45, no. 8, pp. 8–14, 1988.
- [27] P. A. Thompson and E. W. Marchant, "A computer model for the evacuation of large building populations," *Fire Safety Journal*, vol. 24, no. 2, pp. 131–148, 1995.
- [28] R. L. Hughes, "The flow of large crowds of pedestrians," *Mathematics and Computers in Simulation*, vol. 53, no. 4–6, pp. 367–370, 2000.
- [29] B. D. Hankin and R. A. Wright, "Passenger flow in subways," *Operational Research Quarterly*, vol. 9, no. 2, pp. 81–88, 1958.
- [30] Y. Tanaboriboon, S. S. Hwa, and C. H. Chor, "Pedestrian characteristics study in Singapore," *Journal of Transportation Engineering*, vol. 112, no. 3, pp. 229–235, 1986.
- [31] Beijing Urban Engineering Design Research Institute, *Code for Metro Design (GB50157-92)*, China planning press, Beijing, China, 1992.
- [32] Beijing Urban Engineering Design Research Institute, *Code for Design of Metro (GB50157-2003)*, China planning press, Beijing, China, 2003.

Research Article

Optimal Sizing and Control Strategy Design for Heavy Hybrid Electric Truck

Yuan Zou, Dong-ge Li, and Xiao-song Hu

*National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering,
Beijing Institute of Technology, Beijing 100081, China*

Correspondence should be addressed to Yuan Zou, zouyuan@bit.edu.cn

Received 6 September 2012; Accepted 28 October 2012

Academic Editor: Huimin Niu

Copyright © 2012 Yuan Zou 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.

Due to the complexity of the hybrid powertrain, the control is highly involved to improve the collaborations of the different components. For the specific powertrain, the components' sizing just gives the possibility to propel the vehicle and the control will realize the function of the propulsion. Definitely the components' sizing also gives the constraints to the control design, which cause a close coupling between the sizing and control strategy design. This paper presents a parametric study focused on sizing of the powertrain components and optimization of the power split between the engine and electric motor for minimizing the fuel consumption. A framework is put forward to accomplish the optimal sizing and control design for a heavy parallel pre-AMT hybrid truck under the natural driving schedule. The iterative plant-controller combined optimization methodology is adopted to optimize the key parameters of the plant and control strategy simultaneously. A scalable powertrain model based on a bilevel optimization framework is built. Dynamic programming is applied to find the optimal control in the inner loop with a prescribed cycle. The parameters are optimized in the outer loop. The results are analysed and the optimal sizing and control strategy are achieved simultaneously.

1. Introduction

The parallel power-train is one of the most effective configurations for hybrid electric vehicles (HEVs). The benefits of the parallel power-train result from its ability to drive with the engine or the electric motor only, or with both. How to minimize the fuel consumption of this type of HEV is presently quite hot in the academic community. Several energy management strategies have been studied or implemented in the literatures [1–5]. Sciarretta and Guzzella [6] suggested that HEV energy control strategy can be mainly categorized into four groups—the numerical optimization method, the analytic optimization method, the equivalent consumption minimization strategy, and the heuristic strategy. Dynamic

programming (DP) is a numerical method for solving multi-stage decision-making problems and has been widely applied to explore the possible maximum fuel saving for the parallel HEVs [7, 8]. However, an optimal control strategy with the inappropriate component sizing could not guarantee the best fuel economy. It means that component sizing should be studied along with power management strategy to acquire the optimal performance. Hence, the combined optimization problem of the power management and component sizing for HEV is important. The combined plant/controller optimization problem has been researched a lot. For example, [9] discussed several implementations for the combined optimization strategy: the sequential, iterative, bilevel, and simultaneous manners, in which bilevel form was most commonly used [10]. Wu et al. [11, 12] optimized the components' sizes and rule-based control strategy parameters for a hybrid electric vehicle. The highly accurate models were considered in the bilevel framework in [13]. A parameterized powertrain model and the near-optimal controller constituted a combined optimization problem for a fuel cell hybrid vehicle [14]. However, due to the near-optimal controller, the vehicle fuel saving was a bit unsatisfactory. Delphine et al. built a scalable powertrain model to form an integrated optimization problem, in which the outer loop chose the battery capacity, maximum torque of engine and motor as the variables, while dynamic programming was applied to find the optimal control strategy in the inner loop. Each simulation adjusted merely one parameter while keeping the remainder fixed [15]. Therefore, the coupling effects among component parameters were neglected.

In this study, the combined power management and sizing optimization problem for a heavy parallel hybrid electric truck is formulated and solved in a bilevel manner. The paper starts from the power train modelling, including the engine, the motor, the battery, and the transmission. Through the bilevel framework, a scalable vehicle model is developed and integrated in the optimal design process. DP is applied for the power management in the inner loop and the main parameters of the components are optimized simultaneously in the outer loop. The coupling among the component parameters is studied and the considerable fuel economy improvements are achieved.

2. Vehicle Model

2.1. Vehicle Configuration

The baseline vehicle is shown in Figure 1. The hybrid electric truck is a pre-transmission parallel HEV. The engine is connected to an automatic clutch, and then to the transmission. The parameters of this vehicle are given in Table 1.

2.2. Model Simplification

It is highly desirable to perform the extensive simulations for HEVs with the different component configurations at the preliminary system design and optimization. It also means that the scalable model is in great demand at that stage. To avoid the dependence on the specific efficiency maps, a universal representation of the internal combustion engine based on the Willans line concept has been adopted [16]. Considering the complexity of the combined optimization, a simplified scalable motor model is also built later. Those models only consider the dynamic effects related to the low frequency power flows. The transient phenomena, such as chemical reactions in the battery and electric dynamics in the motor, are

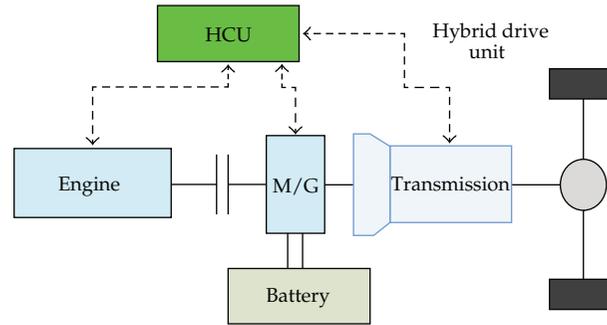


Figure 1: Schematic diagram of the hybrid electric truck.

Table 1: Parameters of the hybrid electric truck.

DI diesel engine	7.0 L, 155 kw@2000 rpm, 900 Nm@1300–1600 rpm
AC motor	Maximum power: 90 kw
	Maximum torque: 600 Nm
	Maximum speed: 2400 rpm
Lithium-ion battery	Capacity: 60 Ah
	Number of modules: 25
	Nominal voltage: 12.5 (volts/module)
AMT	9 speed, gear ratio: 12.11/8.08/5.93/4.42/3.36/2.41/1.76/1.32/1
Vehicle	Curb weight: 16000 kg

ignored. Due to the fact that the computation cost increases exponentially as the number of state increases, only the gear number and SOC are chosen to be the system states.

(1) Engine Modeling

The mean effective pressure p_{me} and the mean piston speed c_m are used to describe the engine power and the operating condition. The following three normalizations are used to define the engine efficiency by avoiding the quantities which depends on the engine size [17]:

$$\begin{aligned}
 p_{me} &= \frac{4 \cdot \pi}{V_d} \cdot T_e, \\
 p_{ma} &= \frac{4 \cdot \pi \cdot H_{LHV}}{V_d} \cdot \frac{\dot{m}}{\omega}, \\
 c_m &= \frac{S}{\pi} \cdot \omega,
 \end{aligned} \tag{2.1}$$

where V_d is the engine's displaced volume, S is the stroke, \dot{m} is the fuel mass flow rate, and H_{LHV} is the fuel low heating value. T_e is the engine effective torque, ω is the engine angular

speed, and p_{ma} can be interpreted as an available mean pressure. When the energy converting efficiency is considered, the following exist:

$$\begin{aligned} T_e \cdot \omega_e &= \eta \cdot \dot{m} \cdot H_{\text{LHV}}, \\ T_e &= e \cdot T_a - T_{\text{loss}} = e \cdot \frac{\dot{m} \cdot H_{\text{LHV}}}{\omega} - T_{\text{loss}}, \end{aligned} \quad (2.2)$$

where η is the engine efficiency, e is the thermodynamic efficiency, and T_a is the available torque that would be generated by engine if all the chemical energy were converted into mechanical form. T_{loss} is the inner loss. Associating (2.1) and (2.2), a dimensionless definition of the engine efficiency can be acquired:

$$\begin{aligned} p_{\text{me}} &= e \cdot p_{\text{ma}} - p_{\text{mloss}}, & \eta &= \frac{p_{\text{me}}}{p_{\text{ma}}}, \\ p_{\text{mloss}} &= \frac{4 \cdot \pi}{V_d} \cdot T_{\text{loss}}. \end{aligned} \quad (2.3)$$

The two parameters e and p_{mloss} are the functions of the engine speed and load. The following parameterizations have been experimentally validated on the different engines [18]:

$$\begin{aligned} e &= e_0(c_m) - e_1(c_m) \cdot p_{\text{ma}}, \\ e_0(c_m) &= e_{00} + e_{01} \cdot c_m + e_{02} \cdot c_m^2, \\ e_1(c_m) &= e_{10} + e_{11} \cdot c_m, \\ p_{\text{mloss}} &= p_{\text{mloss0}} + p_{\text{mloss2}} \cdot c_m^2. \end{aligned} \quad (2.4)$$

The coefficients, $e_{00}, e_{01}, e_{02}, e_{10}, e_{11}, p_{\text{mloss0}}$, and p_{mloss2} , remain unchanged for the different engines in the same family and are obtained through the bench test and parameter identification. Hence the actual engine behavior from the same family is defined by the two size parameters, the swept volume V_d and the piston stroke S . Figure 2 compares the engine model with the actual data collected from the bench experiments for a prototype 7.0 L compression-ignition engine.

(2) Motor Modeling

The motor is modeled based on the experimental data. The motor efficiency is considered as a constant because of its high average efficiency in its feasible working area. Due to the battery power and the motor torque limits, the final motor torque becomes

$$T_m = \begin{cases} \min(T_{m,\text{req}}, T_{m,\text{dis}}(\omega_m), T_{\text{bat},\text{dis}}(\text{SOC}, \omega_m)), & \text{if } T_{m,\text{req}} > 0, \\ \max(T_{m,\text{req}}, T_{m,\text{chg}}(\omega_m), T_{\text{bat},\text{chg}}(\text{SOC}, \omega_m)), & \text{if } T_{m,\text{req}} < 0, \end{cases} \quad (2.5)$$

where $T_{m,\text{req}}$ is the requested motor torque. $T_{m,\text{dis}}(\omega_m)$ and $T_{m,\text{chg}}(\omega_m)$ are the maximum motor torques in the motoring and charging modes, respectively. $T_{\text{bat},\text{dis}}(\text{SOC}, \omega_m)$

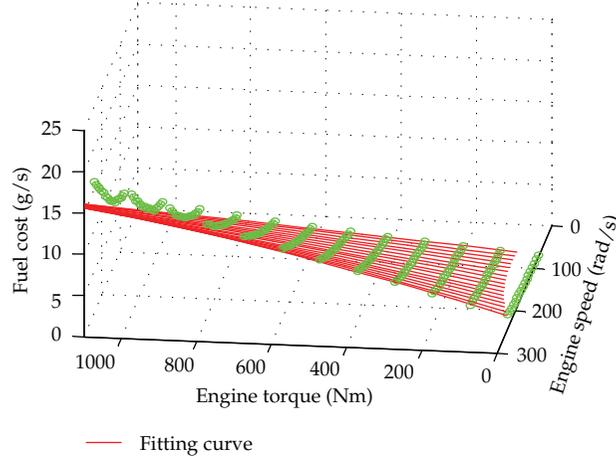


Figure 2: The comparison of Willans line model with the test data of the engine.

and $T_{\text{bat,chg}}(\text{SOC}, \omega_m)$ are the torque bounds due to the battery current limits in the discharging and charging modes.

(3) Battery Modeling

The thermal-temperature effects and transients are ignored. SOC is calculated by

$$\text{SOC}(k+1) = \text{SOC}(k) - \frac{V_{\text{oc}} - \sqrt{V_{\text{oc}}^2 - 4(R_{\text{int}} + R_t) \cdot T_m \cdot \omega_m \cdot \eta_m^{-\text{sgn}(T_m)}}}{2(R_{\text{int}} + R_t) \cdot C_b}, \quad (2.6)$$

where the internal resistance R_{int} and the open circuit voltage V_{oc} are functions of the battery SOC, obtained through the bench test. C_b is the maximum battery charge, R_t is the terminal resistance, and η_m is the efficiency of the motor.

(4) Driveline Modeling

The driveline is defined as the system from the transmission input shaft to wheels. Assuming perfect clutches and gear shifting, the following equations describe the transmission and final drive gear models:

$$\begin{aligned} T_{\text{wheel}} &= \eta_{\text{gear}} \cdot \eta_{\text{FD}} \cdot i_g \cdot i_0 \cdot T_i - \eta_t \cdot \omega_i, \\ \omega_i &= i_g \cdot i_0 \cdot \omega_{\text{wheel}}, \end{aligned} \quad (2.7)$$

where i_g is the transmission gear ratio, i_0 is the final drive gear ratio, η_{gear} and η_{FD} are the transmission and final drive efficiency, respectively. T_i and T_{wheel} are the transmission input torque and output torque, respectively. η_t is the transmission viscous-loss coefficient, ω_i is the transmission input speed, and ω_{wheel} is the wheel speed.

The gear-shifting sequence of the AMT is modeled as a discrete dynamic system:

$$\text{gear}(k+1) = \begin{cases} 9, & \text{gear}(k) + \text{shift}(k) > 9 \\ 1, & \text{gear}(k) + \text{shift}(k) < 1 \\ \text{gear}(k) + \text{shift}(k), & \text{otherwise,} \end{cases} \quad (2.8)$$

where the state $\text{gear}(k)$ is the gear number, and the control $\text{shift}(k)$ to the transmission is constrained to take on the values of $-1, 0$, and 1 , representing down shifting, sustaining, and upshifting, respectively.

(5) Vehicle Dynamics

It is a common practice that only the vehicle longitudinal dynamics is considered. The longitudinal vehicle dynamics is modeled as a point-mass:

$$\omega_{\text{wheel}}(k+1) = \omega_{\text{wheel}}(k) + \frac{T_{\text{wheel}} - T_{\text{brake}} - r_{\omega} \cdot (F_r + F_a)}{M_r \cdot r_{\omega}^2}, \quad (2.9)$$

where T_{brake} is the friction brake torque, F_r and F_a are the rolling resistance force, and the aerodynamic drag force, and r_{ω} is the dynamic tire radius. $M_r = M_V + J_r/r_{\omega}^2$ is the effective mass of the vehicle, and J_r is the equivalent inertia of the rotating components in the vehicle.

3. Combined Optimization Problem Formulation

3.1. Combined Optimization Framework

Given the particular system parameters, DP can be used to find the optimal control theoretically subject to some constraints under a specific driving schedule. When the system parameters vary in the feasible scope, DP is iteratively applied. The optimal combination of the parameters and control will be identified simultaneously. The bilevel combined plant/controller optimization is adopted, consisting of two nested optimization loops. The outer loop evaluates the system parameters. The inner loop generates the optimal control strategy for the parameters selected by the outer loop. These two loops form the integrated plant/controller optimization, which guarantees the global optimal design for the system parameters and control strategy. The combined optimization problem is complicated, due to the interaction between system parameters and control optimization, and computationally expensive due to the bilevel iterative search process. In order to improve the computational efficiency, once the constraints in the inner loop are violated, the current search stops, and the current cost will be set to a huge infeasible value. The flow chart of the bilevel combined optimization process is shown in Figure 3.

3.2. The Scaled Model and Optimization Problem Formulation

The scaled models are needed to parameterize the system conveniently in the optimization. The scope of the motor torque, the motor speed, the motor power, the engine volume, the cylinder stroke, the battery numbers, and the capacity of battery are scaled by `mot_tor`,

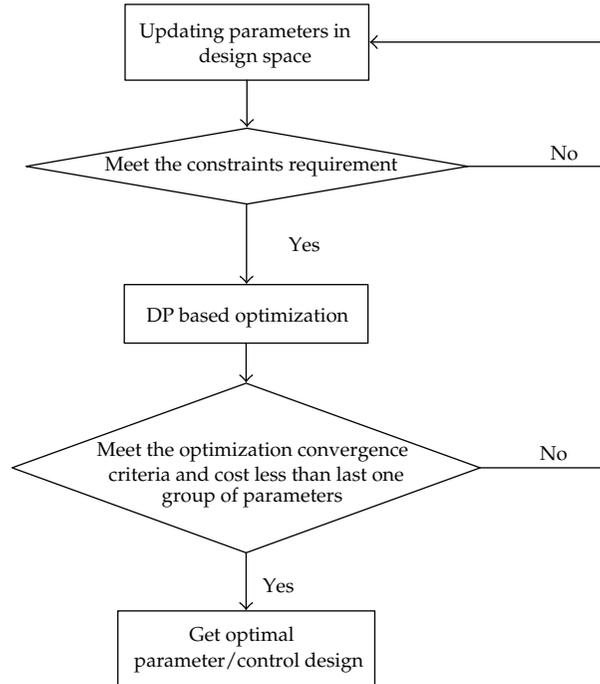


Figure 3: The bilevel combined optimization process.

mot_spd , $motorp$, $Vdscale$, $Sscale$, bat_num , and bat_ah , respectively. The final drive ratio i_0 , varying within a certain range without a scale enlarging, is one of the design parameters. The component parameters are described as follows:

$$\begin{aligned}
 C_b &= bat_ah \cdot C_{b_bas}, \\
 V_{oc} &= bat_num \cdot N_{bas} \cdot V_{oc_bas}, \\
 V_d &= V_{d_bas} \cdot Vdscale, \\
 S &= Sscale \cdot S_{bas}, \\
 Mt &= mot_tor \cdot Mt_{bas}, \\
 Ms &= mot_spd \cdot Ms_{bas}, \\
 Mp &= motorp \cdot Mp_{bas}, \\
 R_{int} &= \frac{bat_num}{bat_ah} \cdot R_{int_bas}, \\
 R_t &= \frac{bat_num}{bat_ah} \cdot R_{t_bas},
 \end{aligned} \tag{3.1}$$

where C_{b_bas} and N_{bas} are the baseline battery capacity and the baseline number of the battery pack, V_{oc_bas} is the baseline open circuit voltage of battery pack as a function of SOC. R_{int_bas}

and $R_{t,\text{bas}}$ are the baseline internal resistance and terminal resistance. Mt_{bas} , Ms_{bas} , and Mp_{bas} are the baseline parameters of the motor, while $V_{d\text{bas}}$ and S_{bas} are those of the engine. The baseline parameters are listed in Table 1. The variables in the left hand of the equations are the scaled parameters that need to be transferred to the inner loop.

The degree of hybridization (DOH) is often adopted to measure the relative contributions of the primary and second power sources. As to the parallel hybrid electric vehicles, the engine is often the primary power source and the battery the secondary power source. The DOH is constrained to be within $[0, 0.4]$ and calculated by

$$x_h = \frac{P_{m.\text{max}}}{P_{e.\text{max}} + P_{m.\text{max}}}, \quad (3.2)$$

where $P_{m.\text{max}}$ is the maximum power that the motor offers, and $P_{e.\text{max}}$ is the maximum power that the engine provides. The combined optimal problem is formulated with all the constraints by

$$\min_{\substack{\text{mot.tor, mot.spd, bat.num,} \\ \text{bat.ah, } i_0, \text{ motorp, Sscale, Vdscale}}} \left\{ \sum_{i=0}^{N-1} T_s \cdot F_d(n_{\text{eng}}(k), T_{\text{eng}}(K)) \right\} \quad (3.3)$$

subject to

$$\begin{aligned} x(k+1) &= f(x(k), u(k)), \\ 0.3 &\leq \text{mot.tor} \leq 2, \\ 0.9 &\leq \text{mot.spd} \leq 2, \\ 0.5 &\leq \text{motorp} \leq 1.5, \\ 0.7 &\leq V\text{dscale} \leq 1.5, \\ 0.9 &\leq S\text{scale} \leq 1.2, \\ 0.5 &\leq \text{bat.num} \leq 3, \\ 0.5 &\leq \text{bat.ah} \leq 3, \\ 2 &\leq i_0 \leq 8, \\ 0 &< x_h \leq 0.4, \\ \text{max speed} &\geq 50 \text{ mph}, \\ \text{acceleration time (0 - 50 mph)} &\leq 45 \text{ sec}, \\ \text{grade (at the speed of 6 mph)} &\geq 20\%, \end{aligned} \quad (3.4)$$

where f represents the dynamics (2.1)–(2.9). The dynamic performance should be limited in the constraints when both the engine and motor propel the car. The constraints on the scaled parameters constitute the design space of the component sizing optimization.

4. Algorithms and Methods

Design of experiments (DOE) technique is first applied to explore the response map in all the feasible design space based on Optimal Latin Hypercube sampling. Then the Nonlinear Programming by Quadratic Lagrangian (NLPQL) algorithm is applied to obtain the global optimal solution [19]. The group of parameters derived from DOE is optimal among the randomly selected points and will be the initial design point for NLPQL algorithm which can build a quadratic approximation. The quadratic programming problem is iteratively solved to find an improved solution until the final convergence to the optimum design.

Dynamic Programming (DP) is a powerful tool for solving optimization problems due to its guaranteed global optimality even for nonlinear dynamic systems with constraints. For a given driving cycle, DP can obtain the optimal operating strategy minimizing fuel consumption.

For maximizing the fuel saving of HEV, the cost function to be minimized has the following form:

$$J = \sum_{k=0}^{N-1} [L_{\text{fuel}}(k) + \beta \cdot |\text{shift}(k)|] + G_N(x_{\text{SOC}}(N)), \quad (4.1)$$

where $L_{\text{fuel}}(k)$ is the instantaneous cost of the fuel use. The vehicle drivability is constrained by $\beta \cdot |\text{shift}(k)|$ to avoid excessive shifting, in which β is a positive weighting factor. A terminal constraint on SOC, represented by $G_N(x_{\text{SOC}}(N))$, is imposed on the cost function due to the charge-sustaining strategy. During the optimization, it is necessary to enforce the following inequality constraints to ensure safe and smooth operation for the engine, the battery, and the motor:

$$\begin{aligned} \omega_{e.\text{min}} &\leq \omega_e(k) \leq \omega_{e.\text{max}}, \\ \text{SOC}_{\text{min}} &\leq \text{SOC}(k) \leq \text{SOC}_{\text{max}}, \\ T_{e.\text{min}}(\omega_e(k)) &\leq T_e(k) \leq T_{e.\text{max}}(\omega_e(k)), \\ T_{m.\text{min}}(\omega_e(k), \text{SOC}(k)) &\leq T_m(k) \leq T_{m.\text{max}}(\omega_e(k), \text{SOC}(k)), \end{aligned} \quad (4.2)$$

where ω_e is the engine speed, SOC is the battery state of charge. SOC_{min} and SOC_{max} are selected to be 0.4 and 0.8, respectively. T_e is the engine torque, and T_m is the motor torque. A generic DP algorithm is implemented in MATLAB and applied to solve the above optimal control problem [20, 21].

5. Simulations and Results

The heavy-duty vehicle driving schedule used to evaluate the fuel economy of the hybrid electric truck is shown in Figure 4.

The Pareto figure indicating the influence of the various factors on the fuel consumption is shown in Figure 5. It is determined by ordering the scaled and normalized coefficients of a standard least-squares second-order polynomial fit to the contribution to the fuel consumption from the different parameters. It is evident that motor_p , $vdscale$, and i_0 individually have a significant effect on the fuel consumption. These three parameters

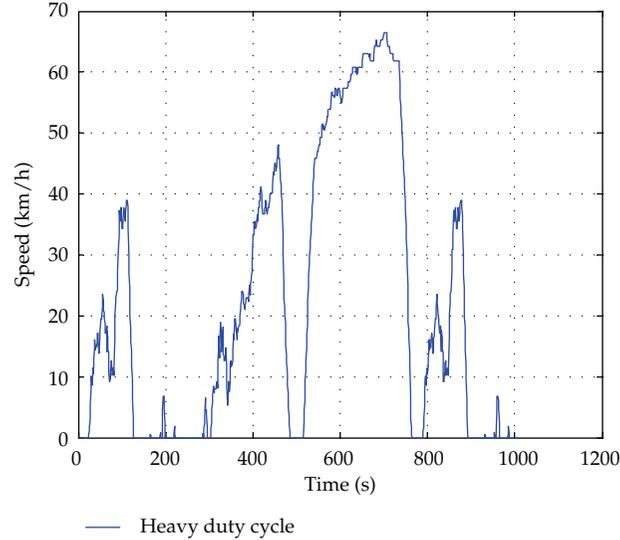


Figure 4: The heavy-duty vehicle driving schedule.

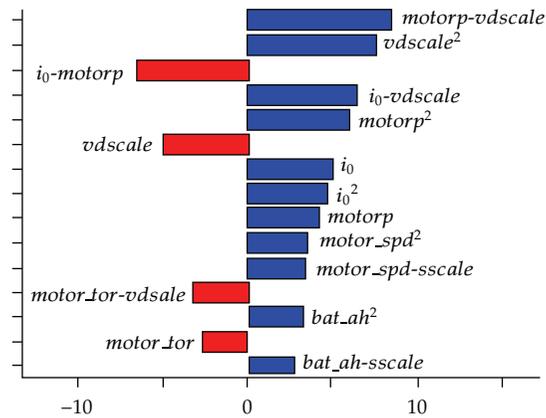


Figure 5: The Pareto plot for the various factors' influence on the fuel consumption.

represent the motor's maximum power, the engine's maximum power, and the final drive ratio, respectively. However, the interaction between the maximum motor power and the engine volume has the largest impact on the fuel consumption. The effects of the battery capacity on the fuel consumption are not as significant as other parameters; the percentage is less than 3%. Therefore, the battery supplying enough power for the motor can be chosen based on the cost effectiveness.

The specific influences on the fuel economy from the power sizing of the engine and the motor are shown in Figure 6. Note that alteration of the engine volume brings the change of the engine maximum power. It can also be concluded that the fuel consumption does not decrease as the engine size reduces or the motor size increases. Both of them should be chosen within a specific range in order to obtain the impressive fuel economy.

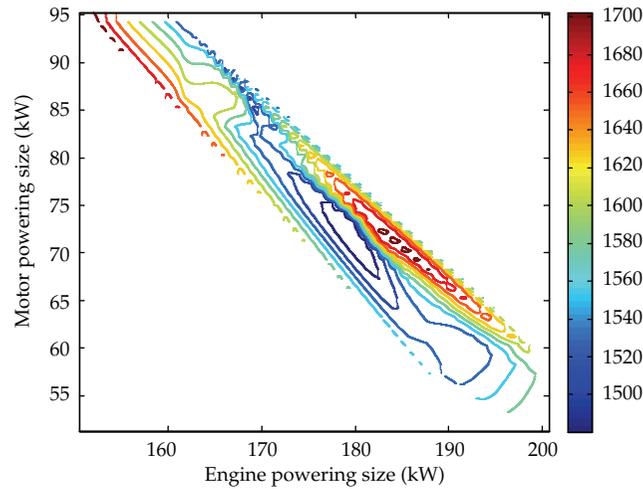


Figure 6: The fuel consumption versus the motor and engine powering size.

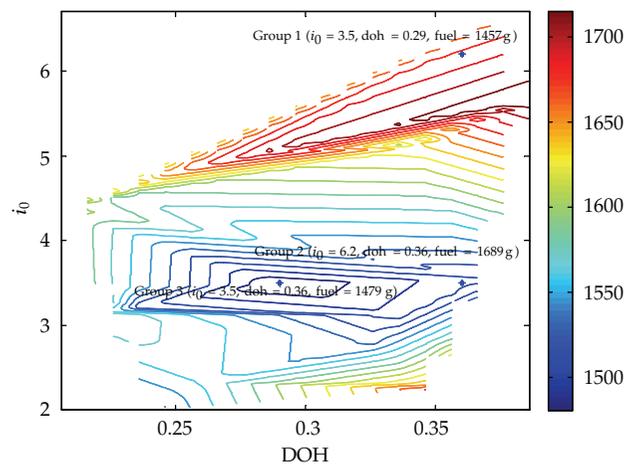


Figure 7: The fuel consumption versus i_0 and DOH.

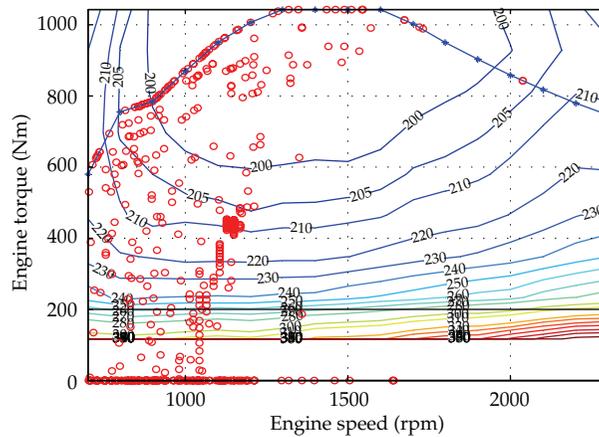
The parameters of the three typical groups with the different components' sizes are listed in Table 2. The second and third group only differs in the final drive ratio and the third one has the same final ratio with the first one. The three groups of parameters are marked in Figure 7. It may allow the conclusion that the final drive ratio i_0 should be selected within a limited range, roughly between 3 and 4, slightly smaller than the initial value, to keep the good fuel economy, regardless of the DOH. The improper selection of the final drive ratio can lead to the increasing fuel consumption despite the optimal control strategy.

The engine working area and the gear shifting of the first and second group is shown in Figures 8, 9, 10, and 11. The second group with a smaller engine has a fundamentally different gear shifting from the first one. It is easy to extract the shifting rule from Figure 9, whereas difficult to obtain a shifting line for the second group because there is no apparent boundary between neighboring gears in Figure 11. The improper selection of the final drive

Table 2: The comparison within the different groups.

Group number	bat_ah	bat_num	i_0	motor_spd	motor_tor	motorp
1	0.98	0.7	3.50	1.15	1.93	0.77
2	2.59	0.82	6.20	1.55	1.44	1.00
3	2.59	0.82	3.50	1.55	1.44	1.00

Group number	sscale	vdscale	DOH	Engine's max power (kW)	Total power (kW)	Fuel consumption (g)
1	0.94	1.16	0.29	179	252	1457
2	0.9	1.00	0.36	154	248	1689
3	0.9	1.00	0.36	154	248	1479

**Figure 8:** The working area of engine in the first group.

ratio will result in low efficiency working area for the engine more possibly and could not be compensated by optimizing gear shifting and power distribution.

It is clear that the component parameters can affect HEV fuel economy directly. Sometimes a slight parameter discrepancy may lead to the considerable change of the fuel consumption. It emphasizes that the component sizing of HEV should be designed with a great cautiousness.

The optimal and initial parameters are listed in Table 3. The battery capacity decreases to 30 Ah from the original value, 60 Ah, although its voltage increases a bit. The final ratio decreases to 3.3 from the original value 4.769. Although the motor power is decreased, the motor max torque is found to increase by 63% to meet the performance constraints. Around 9% improvement is observed in the fuel economy through the combined optimization. The feasibility of the components in the actual engineering applications, however, needs more investigation in the view of the reliability and cost effectiveness.

6. Conclusion

A bilevel optimization problem for the combined component sizing and power distribution of a heavy hybrid electric truck is formulated and solved. DOE and NPQRL algorithms are

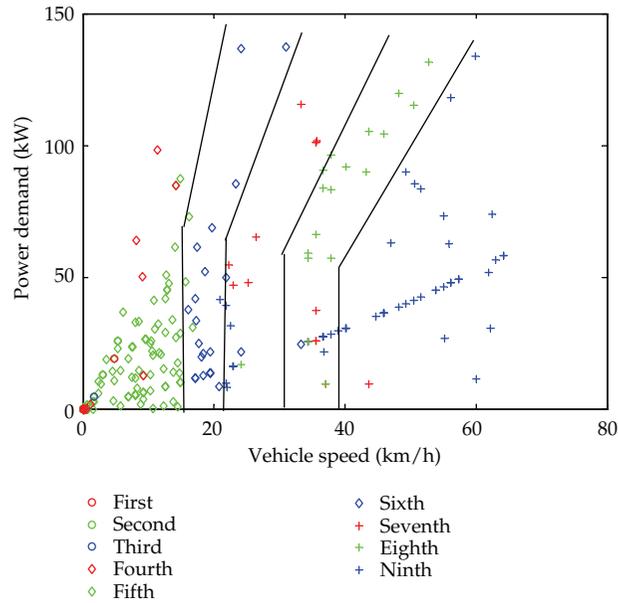


Figure 9: The gear shifting in the first group.

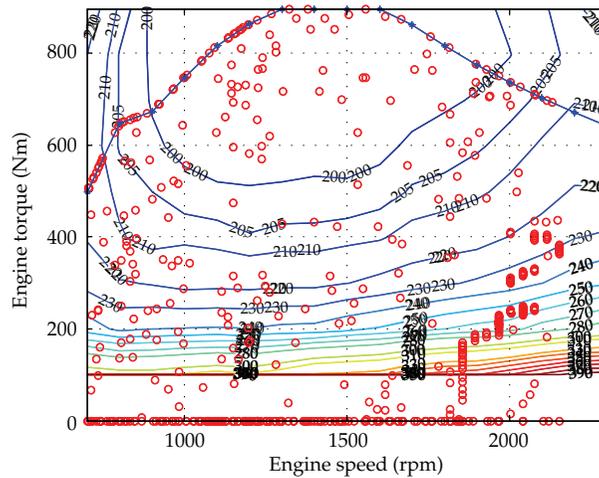


Figure 10: The working area of the engine in the second group.

Table 3: The values of the baseline and optimization parameters.

C (Ah)	V (V)	i_0	Max motor speed (rpm)	Max motor torque (nm)	Max motor power (kW)	Max engine power (kW)	Fuel economy (mile/gallon)
The initial values							
60	312	4.769	2400	600	94	155	26.4
The optimal values							
30	393	3.3	2400	980	83	163	28.7

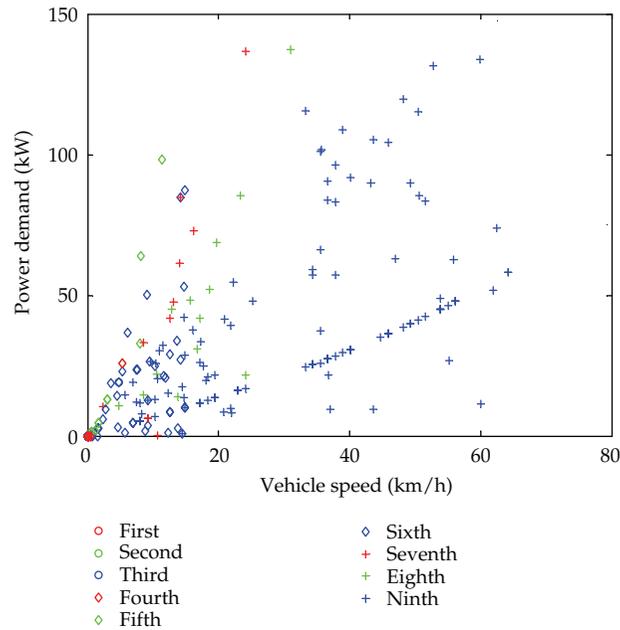


Figure 11: The gear shifting in the second group.

applied to find the optimal component parameters in the outer loop, while DP is used to find the optimal energy strategy in the inner loop. Simulation results show that the complex relationships between the component sizes and fuel consumption can be efficiently analyzed by solving the combined optimization problem. The law extracted from the optimization results can provide the suggestions for the actual hybrid vehicle system optimization and control. The results also indicate that the comprehensive bilevel optimization framework can facilitate the enhancement of HEV fuel economy, and the components sizing is as important as the control strategy.

Acknowledgments

This research is financially supported by China Natural Science Funding Project (50905015), China 863 High Technology Project (2011AA11A223), and China University Discipline Talent Introduction Program (B12022).

References

- [1] X. Wei, L. Guzzella, V. I. Utkin, and G. Rizzoni, "Model-based fuel optimal control of hybrid electric vehicle using variable structure control systems," *Journal of Dynamic Systems, Measurement and Control*, vol. 129, no. 1, pp. 13–19, 2007.
- [2] A. Sciarretta, M. Back, and L. Guzzella, "Optimal control of parallel hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 12, no. 3, pp. 352–363, 2004.
- [3] C. C. Lin, M. J. Kim, H. Peng, and J. W. Grizzle, "System-level model and stochastic optimal control for a PEM fuel cell hybrid vehicle," *Journal of Dynamic Systems, Measurement and Control*, vol. 128, no. 4, pp. 878–890, 2006.

- [4] S. E. Lyshevski, "Energy conversion and optimal energy management in diesel-electric drivetrains of hybrid-electric vehicles," *Energy Conversion and Management*, vol. 41, no. 1, pp. 13–24, 2000.
- [5] C.-C. Lin, H. Peng, J. W. Grizzle, and J. M. Kang, "Power management strategy for a parallel hybrid electric truck," *IEEE Transactions on Control Systems Technology*, vol. 11, no. 6, pp. 839–849, 2003.
- [6] A. Sciarretta and L. Guzzella, "Control of hybrid electric vehicles," *IEEE Control Systems Magazine*, vol. 27, no. 2, pp. 60–70, 2007.
- [7] B. Gu, *Supervisory control strategy development for a hybrid electric vehicle [M.S. thesis]*, The Ohio State University, Columbus, Ohio, USA, 2006.
- [8] L. V. Pérez, G. R. Bossio, D. Moitre, and G. O. García, "Optimization of power management in a hybrid electric vehicle using dynamic programming," *Mathematics and Computers in Simulation*, vol. 73, no. 1–4, pp. 244–254, 2006.
- [9] H. K. Fathy, J. A. Reyer, P. Y. Papalambros, and A. G. Ulsoy, "On the coupling between the plant and controller optimization problems," in *Proceedings of the American Control Conference*, pp. 1864–1869, Arlington, Va, USA, June 2001.
- [10] J. F. Bonnans, Th. Guilbaud, A. Ketfi-Cherif, C. Sagastizabal, D. Wissel, and H. Zidani, "Parametric optimization of hybrid car engines," *Optimization and Engineering*, vol. 5, no. 4, pp. 395–415, 2004.
- [11] J. Wu, C.-H. Zhang, and N.-X. Cui, "Psoalgrithm-based parameter optimization for HEV power-train and its control strategy," *International Journal of Automotive Technology*, vol. 9, no. 1, pp. 53–69, 2008.
- [12] C. Desai and S. S. Williamson, "Optimal design of a parallel hybrid electric vehicle using multi-objective genetic algorithms," in *Proceedings of the 5th IEEE Vehicle Power and Propulsion Conference (VPPC '09)*, pp. 871–876, Dearborn, Mich, USA, September 2009.
- [13] D. Assanis, G. Delagrammatikas, R. Fellini et al., "Optimization approach to hybrid electric propulsion system design," *Mechanics of Structures and Machines*, vol. 27, no. 4, pp. 393–421, 1999.
- [14] M.-J. Kim and H. Peng, "Power management and design optimization of fuel cell/battery hybrid vehicles," *Journal of Power Sources*, vol. 165, no. 2, pp. 819–832, 2007.
- [15] D. Sinoquet, G. Rousseau, and Y. Milhau, "Design optimization and optimal control for hybrid vehicles," *Optimization and Engineering*, vol. 12, no. 1-2, pp. 199–213, 2011.
- [16] X. Wei and G. Rizzoni, "A scalable approach for energy converter modeling and supervisory control design," in *Proceedings of the ASME International Mechanical Engineering Congress and Exposition*, pp. 1281–1288, November 2001.
- [17] G. Rizzoni, L. Guzzella, and B. M. Baumann, "Unified modeling of hybrid electric vehicle drivetrains," *IEEE/ASME Transactions on Mechatronics*, vol. 4, no. 3, pp. 246–257, 1999.
- [18] A. Urlaub, *Verbrennungsmotoren*, Springer-Verlag, Berlin, Germany, 1994.
- [19] Dassault Company, *Isight User Guide (4.5 Release)*, Dassault Company, Velizy-Villacoublay, France, 2010.
- [20] Y. Zou, F. Sun, C. Zhang, and J. Li, "Optimal energy management strategy for hybrid electric tracked vehicles," *International Journal of Vehicle Design*, vol. 58, no. 2–4, pp. 307–324, 2012.
- [21] O. Sundström and L. Guzzella, "A generic dynamic programming Matlab function," in *Proceedings of the IEEE International Conference on Control Applications (CCA '09)*, pp. 1625–1630, July 2009.

Research Article

A Compromise Programming Model for Highway Maintenance Resources Allocation Problem

Hui Xiong,¹ Qixin Shi,² Xianding Tao,³ and Wuhong Wang¹

¹ *Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China*

² *Department of Civil Engineering, Tsinghua University, Beijing 100084, China*

³ *ITS Division, District Department of Transportation, Washington, DC 20003, USA*

Correspondence should be addressed to Hui Xiong, xionghui@bit.edu.cn

Received 5 August 2012; Accepted 22 October 2012

Academic Editor: Geert Wets

Copyright © 2012 Hui Xiong 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 formulates a bilevel compromise programming model for allocating resources between pavement and bridge deck maintenances. The first level of the model aims to solve the resource allocation problems for pavement management and bridge deck maintenance, without considering resource sharing between them. At the second level, the model uses the results from the first step as an input and generates the final solution to the resource-sharing problem. To solve the model, the paper applies genetic algorithms to search for the optimal solution. We use a combination of two digits to represent different maintenance types. Results of numerical examples show that the conditions of both pavements and bridge decks are improved significantly by applying compromise programming, rather than conventional methods. Resources are also utilized more efficiently when the proposed method is applied.

1. Introduction

By the end of the 1960s, major construction of highway networks ended in developed countries such as the US and Canada. The main tasks of highway agencies shifted from planning, design, and construction to management and maintenance of the highway infrastructure, which usually consumes large resources. In the US, for instance, the total investment in highway infrastructure has reached \$1 trillion [1]. In recent years, highway agencies have to face tough budgeting problems in maintaining existing highway facilities. Due to ever-increasing transportation demand, more funds are needed to rehabilitate the deteriorating highway facilities, and the gap between required funds and budgets is increasing. It becomes critical that the limited resources should be more effectively allocated and used.

Consequently, highway asset management, as an effective means for infrastructure management, has received increasing attention. The principle of highway asset management can be defined as a strategic approach to the optimal allocation of resources for management, operation, and preservation of highway infrastructure. Highway asset management emphasizes resource sharing and optimal allocation among all sections of the whole highway network, beyond the scope of single facility management (e.g., pavement management or bridge management). Through highway asset management, the limited resources can be used more effectively. Hence, the infrastructure will better serve the needs of road users during its life cycle. The conceptual core of highway asset management is resource allocation [2].

Highway asset maintenance management (HAMM) is a critical element of highway asset management. Through its integrated management for all components of the entire highway system during its life cycle, the objective of HAMM is to optimize resource allocation in a wider range and find the highway maintenance solution that minimizes the costs or maximizes the benefits, considering the resources constraints in funds, labor, and equipment. Specifically, HAMM will solve such questions as when and how the maintenance should be implemented. Since different maintenance types and implementation times would yield different performances, the HAMM process finds the optimal combinations in order to achieve stated objectives. The resource allocation problem in HAMM is usually a multiobjective problem. The main task in solving multiobjective problem is to obtain Pareto-optimal solutions. As an effective method for solving a multiobjective problem, compromise programming (CP) was developed by Yu [3] and Zeleny [4]. In recent several decades, CP was broadly applied in many fields.

Lounis and Cohn [5] applied compromise programming approach to select satisfying solution for multicriteria optimization of engineering structures and structural systems. To solve multicriteria decision making in irrigation planning, Zarghaami [6] formulated compromise programming models with multiobjectives such as regulation of reliable water at the demand time, improving rice and tea production, domestic water supply, environmental needs, as well as reducing social conflicts. Diaz-Balteiro et al. [7] applied compromise programming to rank sustainability of European paper industry countries in terms of 14 indicators including economic, environmental, and social indices. It was found that the results were robust when different preferential weights were attached. And the methodology can be applied at a more disaggregated level and other indicators can be introduced. Amiri et al. [8] proposed a Nadir compromising programming (NCP) model by expanding a CP-based method for optimization of multiobjective problems in financial portfolios. The NCP model was formulated on the basis of the nadir values of each objective. Numerical example of a multiobjective problem to select optimal portfolio in Iran stock market proved that the NCP model can satisfy decision maker's purposes better. Andre et al. [9] assessed Spanish economy by taking compromise programming as an analytical tool and studied several Pareto-efficient policies that represent compromises between economic growth and inflation rate. Hashimoto and Wu [10] proposed a data envelopment analysis—compromise programming model for comprehensive ranking including preference voting to rank candidates. Shiao and Wu [11] applied compromise programming to optimize water allocation scheme under multiobjective criteria to minimize the hydrologic alteration and water supply shortages. By combination of fuzzy compromise programming and group decision making under fuzziness, Prodanovic and Simonovic [12] formulated a new multiple criteria multiple expert decision support methodology. However, few have applied CP in pavement management, bridge deck maintenance, or highway asset maintenance management except [13, 14].

This paper applies the concepts of compromise programming and formulates a bilevel model for the resource allocation problem. The first level of the model solves the resource allocation problems for pavement management and bridge deck maintenance, without considering resource sharing between them. The second level of the model solves the resources allocation problems considering resource sharing. The following content is organized as below. Section 2 presents the bilevel optimization model based on compromising programming. Section 3 is a numerical example applying genetic algorithms to solve the model. The conclusions and recommendations for further study are presented in Section 4.

2. Model

The basic concepts of CP are presented below. Consider the following general multiobjective problem [13, 14]:

$$\begin{aligned} \min \quad & \{f_1(x), f_2(x), \dots, f_k(x)\}, \\ \text{s.t.} \quad & x \in X, \end{aligned} \quad (2.1)$$

where there are k (≥ 2) objective functions $\{f_i(x), i = 1, 2, \dots, k\}$.

The constraints can be expressed in a general form as $x \in X = \{x \in R^n \mid g_j(x) \leq 0, j = 1, 2, \dots, q; h_l(x) = 0, l = 1, 2, \dots, r\}$, where q and r are numbers of inequality and equality constraints, respectively.

If the problem does not have any conflicting objectives, one can apply conventional optimization approaches to obtain a final solution that optimizes all objective functions. However, the objectives are mutually conflicting in many real-world engineering problems. In addition, those objectives can have different measurement scales. Zeleny [4] developed a method that transforms these objectives into a set of comparable scales. He then formulated a new single-objective optimization problem as in (2.2) and showed that the optimal solution from the new problem is a Pareto-optimal solution to the original problem (2.1). Consider

$$\begin{aligned} \min \quad & L_p = \left[\sum_{i=1}^k \left(\lambda_i \cdot \frac{f_i(x) - y_i^0}{y_i^0} \right)^p \right]^{1/p}, \\ \text{s.t.} \quad & x \in X, \quad \lambda \in \Lambda. \end{aligned} \quad (2.2)$$

In the Previous expression, y_i^0 is the optimal solution to the original problem containing only the i th objective function $f_i(x)$. p is a parameter satisfying $1 \leq p < \infty$. $\Lambda = \{\lambda \in R^k \mid \lambda \geq 0, \sum_{i=1}^k \lambda_i = 1\}$. Let $Z_i(x) = (f_i(x) - y_i^0)/y_i^0$. The meanings of the objective function in (2.2) for different parameter p are discussed in the following.

(a) If $p = 1$, (2.2) can be simplified to (2.3),

$$\begin{aligned} \min \quad & L_p = \sum_{i=1}^k |\lambda_i \cdot Z_i(x)|, \\ \text{s.t.} \quad & x \in X, \quad \lambda \in \Lambda. \end{aligned} \quad (2.3)$$

The new objective function, L_p , is a weighted average of all shortest distances of objective values. L_p is called the Manhattan distance.

(b) If $1 < p < \infty$, then (2.2) can be transformed into (2.4)

$$\begin{aligned} \min \quad L_p &= \left[\sum_{i=1}^k (\lambda_i \cdot Z_i(x))^2 \right]^{1/2}, \\ \text{s.t.} \quad x &\in X, \quad \lambda \in \Lambda. \end{aligned} \quad (2.4)$$

In this case, the new objective function, L_p , is a weighted summation of geometric distances. When $p = 2$, L_p is a Euclidean weighted distance.

(c) If $p = \infty$, then (2.2) can be transformed into (2.5). In this case, the new objective function, L_p , is a Chebyshev distance as follows

$$\begin{aligned} \min \quad L_p &= \max |\lambda_i \cdot Z_i(x)|, \\ \text{s.t.} \quad x &\in X, \quad \lambda \in \Lambda. \end{aligned} \quad (2.5)$$

Under such transformations, the multiobjective problem in (2.1) has an equivalent single-objective problem, which can be expressed as

$$\begin{aligned} \min \quad L_p &= \left[\sum_{i=1}^k (\lambda_i \cdot Z_i(x))^p \right]^{1/p}, \\ \text{s.t.} \quad x &\in X, \quad \lambda \in \Lambda. \end{aligned} \quad (2.6)$$

Earlier approaches to highway infrastructure management were usually developed for single facility management. In this section, we present an optimization model for resource allocation in pavement maintenance. A model for bridge deck maintenance is very similar. The major difference between them is that in the pavement maintenance model, road surface is divided into segments of different lengths, while in the bridge deck maintenance model a bridge deck is divided into different areas. In addition, performance indices are different between the two models.

The objective for the pavement maintenance is to maximize the weighted average surface performance after maintenance. It can be formulated as

$$\max \left\{ \frac{1}{\sum_{i=1}^I L_i \times \omega_i} \times \sum_{i=1}^I \sum_{a=0}^A X_{ia}^0 \times L_i \times PQI_{ia} \times \omega_i \right\}, \quad (2.7)$$

where PQI_{ia} is the pavement quality index of segment i after a type a maintenance, $a = 0, 1, \dots, A$, L_i is the length of segment i , ω_i is the associated weight for segment i , I is the total number of road segments in the whole network, A is the total number of maintenance types, X_{ia}^0 is a binary decision variable. $X_{ia}^0 = 1$ when a type a maintenance is selected for segment i , and 0 otherwise.

Four sets of constraints are considered in the problem: funding constraints, human resource constraints, equipment constraints, and constraints on road maintenance types.

Funding constraints are expressed in (2.8). It indicates that the total maintenance expenses for all segments must not exceed the total available funds in a specific planning year as follows:

$$\sum_{i=1}^I \sum_{a=0}^A X_{ia}^0 \times L_i \times C_{ia}^p \leq B_p, \quad (2.8)$$

where C_{ia}^p in the above expression is the maintenance expense when a type a maintenance is selected for segment i and B_p is the total available funds for all pavement of the network.

Similar to the funding constraints, human resource constraints can be expressed as

$$\sum_{i=1}^I \sum_{a=0}^A X_{ia}^0 \times L_i \times m_{ia}^p \leq M_u^p, \quad (2.9)$$

where m_{ia}^p is the required amount of type u labor when a type a maintenance is selected for segment i and M_u^p is the total available amount of type u labor.

The equipment constraints can be expressed as

$$\sum_{i=1}^I \sum_{a=0}^A X_{ia}^0 \times L_i \times e_{iak}^p \leq E_k^p, \quad (2.10)$$

where e_{iak}^p is the required amount of type k equipment when a type a maintenance is selected for segment i and E_k^p is the total available amount of type k equipment.

It is also assumed that in a maintenance cycle, only one type of maintenance is selected for one segment. Thus, one has the following constraints on road maintenance types:

$$\sum_{a=0}^A X_{ia}^0 = 1, \quad \forall i. \quad (2.11)$$

Typically, highway assets include pavement and bridges. The objective of highway asset maintenance management is to maximum the overall quality of the entire highway network. In this paper, we only consider two types of maintenance: pavement and bridge deck maintenance. Based on the concepts of compromise programming, we formulate a single-objective function as a weighted summation of two distances as below:

$$\max \left\{ \left[\left(\lambda_1 \cdot \frac{\sum_i (\sum_a X_{ia} L_i \omega_i PQI_{ia} - \sum_a X_{ia}^0 L_i \omega_i PQI_{ia})}{\sum_i \sum_a X_{ia}^0 L_i \omega_i PQI_{ia}} \right)^2 + \left(\lambda_2 \cdot \frac{\sum_j (\sum_r Y_{jr} S_j \omega_j CR_{jr} - \sum_r Y_{jr}^0 S_j \omega_j CR_{jr})}{\sum_j \sum_r Y_{jr}^0 S_j \omega_j CR_{jr}} \right)^2 \right]^{1/2} \right\}, \quad (2.12)$$

where parameters λ_1, λ_2 represent the weights of pavement and bridge decks, respectively, and ω_i, ω_j are the weights of pavement segment i and bridge deck unit j , respectively; X_{ia} and Y_{jr} are binary decision variables. $X_{ia} = 1$ if a type a maintenance is selected for pavement segment i , and 0 otherwise; $Y_{jr} = 1$ if a type r maintenance is selected for bridge deck unit j , and 0 otherwise. In (2.12), we take p as 2 to avoid cancellation of positive and negative values in the objective function.

The problem has five sets of constraints, which are discussed in the following content.

2.1. Funding Constraints

As the funds for pavement maintenance and bridge deck maintenance can be shared, the total expenses for both pavement and bridge deck maintenance must not exceed the total available funds. Thus, one has the following:

$$\sum_{i=1}^I \sum_{a=0}^A X_{ia} L_i C_{ia}^p + \sum_{j=1}^J \sum_{r=0}^R Y_{jr} S_j C_{jr}^b \leq B, \quad (2.13)$$

where B is the total available funds for entire network maintenance.

2.2. Human Resource Constraints

In highway maintenance, labor can be classified into three groups: personnel only capable of managing pavement, personnel only capable of managing bridge decks, and personnel having both capabilities. Among these three groups, personnel in the third group can be shared between pavement maintenance and bridge deck maintenance. Equations (2.14), (2.15) and (2.16) describe constraints on labor in these three groups, respectively:

$$\sum_{i=1}^I \sum_{a=0}^A X_{ia} L_i m_{ia}^p \leq M_u^p \quad u = f + 1, \dots, U \quad f + 1 \leq U, \quad (2.14)$$

$$\sum_{j=1}^J \sum_{r=0}^R Y_{jr} S_j m_{jr}^b \leq M_v^b \quad v = f + 1, \dots, V \quad f + 1 \leq V, \quad (2.15)$$

$$\sum_{i=1}^I \sum_{a=0}^A X_{ia} L_i m_{ia}^p + \sum_{j=1}^J \sum_{r=0}^R Y_{jr} S_j m_{jr}^b \leq M_u^p + M_v^b \quad u = v = 1, \dots, f, \quad (2.16)$$

where f is the total number of labor types that can be shared between pavement maintenance and bridge deck maintenance; U is the total number of labor types that can only be used for pavement maintenance; V is the total number of labor types that can only be used for bridge deck maintenance.

2.3. Equipment Constraints

Equipment resources can also be classified into three groups: equipments only usable in pavement maintenance, equipments only usable in bridge deck maintenance, and equipments usable in both pavement and bridge deck maintenance. Equations (2.17), (2.18), and (2.19) describe constraints on equipment in these three groups, respectively:

$$\sum_{i=1}^I \sum_{a=0}^A X_{ia} L_i e_{iak}^p \leq E_k^p \quad k = g+1, \dots, K \quad g+1 \leq K, \quad (2.17)$$

$$\sum_{j=1}^J \sum_{r=0}^R Y_{jr} S_j e_{jrl}^b \leq E_l^b \quad l = g+1, \dots, L \quad g+1 \leq L, \quad (2.18)$$

$$\sum_{i=1}^I \sum_{a=0}^A X_{ia} L_i e_{iak}^p + \sum_{j=1}^J \sum_{r=0}^R Y_{jr} S_j e_{jrl}^b \leq E_k^p + E_l^b \quad k = l = 1, \dots, g, \quad (2.19)$$

where g is the total number of equipment types that can be shared, K is the total number of equipment types that can be used in pavement maintenance, and L is the total number of equipment types that can be used in bridge deck maintenance.

2.4. Constraints on Maintenance Types

In each maintenance cycle, it is assumed that only one type of maintenance can be selected for a pavement segment or bridge deck unit. Hence, we have additional constraints as follows:

$$\begin{aligned} \sum_{a=0}^A X_{ia} &= 1 \quad \forall i, \\ \sum_{r=0}^R Y_{jr} &= 1 \quad \forall j. \end{aligned} \quad (2.20)$$

2.5. Least Performance Constraints

From the transformations of objective functions in compromise programming, it is noted that when parameter p is an even number, the value within the parentheses in (2.12) can be either positive or negative. When this value becomes negative, highway asset maintenance management will not achieve any benefits. Hence, we have additional constraints that the performance of highway infrastructure under resource sharing must exceed or at least equal that without resource sharing. Such constraints can be presented in the following:

$$\begin{aligned} \sum_{i=1}^I \left(\sum_{a=0}^A X_{ia} L_i \omega_i P Q I_{ia} - \sum_{a=0}^A X_{ia}^0 L_i \omega_i P Q I_{ia} \right) &\geq 0, \\ \sum_{j=1}^J \left(\sum_{r=0}^R Y_{jr} S_j \omega_j C R_{jr} - \sum_{r=0}^R Y_{jr}^0 S_j \omega_j C R_{jr} \right) &\geq 0. \end{aligned} \quad (2.21)$$

Table 1: Human resource demands for different types of pavement maintenance.

Maintenance types	Labor Demands (Man-days)			
	Technical chiefs	Drivers	Ordinary workers	Equipment operators
No maintenance	0	0	0	0
Minor maintenance	1	1	1	2
Moderate maintenance	1	3	3	4
Major maintenance	2	5	6	7

Table 2: Productivities for types of maintenances and types of road facilities.

Highway types	Productivities (Lane-km/Day)			
	No maintenance	Minor maintenance	Moderate maintenance	Major maintenance
State highways	0.0	5.1	2.3	0.8
Provincial highways	0.0	6.7	3.5	1.4
County roads	0.0	8.5	4.2	2.1

3. Numerical Analysis

In this numerical example, the highway network consists of four types of facilities: 16.7 lane-km of state highways, 30.7 lane-km of provincial highways, 59.6 lane-km of county roads, and 20.3 lane-km of bridges. For each pavement unit or bridge deck unit, one can choose from the following four options: major maintenance, moderate maintenance, minor maintenance, and no maintenance. The resources (e.g., funds, human resources, equipments) amounts required for different maintenance options and for different facility types are assumed to be known. It is also assumed that the information on infrastructure quality after different types of maintenance is available for analysis.

Table 1 lists the amounts of human resources required in different types of maintenance. Table 2 shows the productivities for different maintenance types against facility types. In addition, the model requires other information such as unit maintenance expense, equipment allocation, before-and-after facility quality, interest rates, total available funds, and other resources, least performance requirements for each facility unit, and variations of traffic volumes. The input information for each bridge deck unit is similar to that for a pavement segment.

Although the highway network is not large, the space of all feasible solutions to the problem is very large, and complete enumeration is computationally infeasible as a method of finding the optimal solution. In this paper, we apply genetic algorithms (GA) to search for the optimal solution.

A chromosome in GA is a string of 0-1 numbers. Here we use a combination of two digits to represent various maintenance types. For instance, "00" stands for "no maintenance", while "01," "10," and "11" stand for minor, moderate, and major maintenances, respectively.

In this example, a blue cell represents a pavement segment, and a green cell represents a bridge deck segment. A chromosome, that is, a feasible solution to the model, consists of 10 fractions. Each fraction represents a one-year pavement and bridge deck maintenance plan, see Figure 1. As for any specific fraction, there are 2 parts, blue one and green one. The blue part indicates pavement maintenance plan and the green part corresponds to bridge deck maintenance plan for a given year.

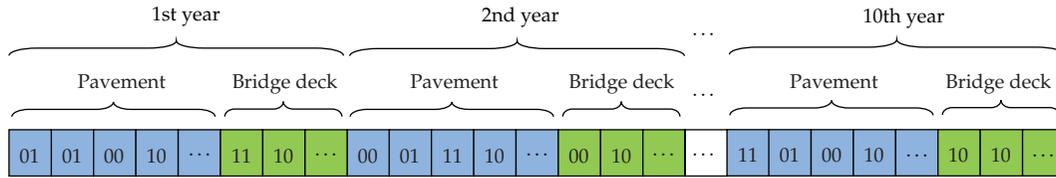


Figure 1: GA code for the solution.

Table 3: Human resources utilization in highway maintenance with and without resource sharing (working days).

Labor types	Maintenance without resource sharing			Maintenance under resource sharing	Total available human resources
	Pavement	Bridge deck	Total		
Technical chiefs	4.7	8.6	13.3	21.7	26.0
Drivers	8.3	15.9	24.2	35.0	35.0
Ordinary workers (1)	8.3	0.0	8.3	16.1	30.0
Ordinary workers (2)	0.0	19.5	19.5	18.9	30.0
Equipment operators (1)	13.0	0.0	13.0	28.3	40.0
Equipment operators (2)	0.0	18.3	18.3	18.9	30.0

There are two steps in solving the problem. At the first step, we obtain the optimized solutions for pavement maintenance and bridge deck maintenance, without considering resource sharing. The second step uses the results from the first step as input and generates the final solution to the resource sharing problem.

The GA parameters are determined with some preliminary analyses. In solving the pavement maintenance problem at the first step, the population size and iteration number, mutation rate are set at 500, 600, and 0.02, respectively. In solving the bridge deck maintenance problem, the population size, iteration, mutation rate to 300, 200, and 0.02, respectively. At the second step computation (with resource sharing), the values of these 3 parameters are set at 500, 2000, and 0.03, respectively.

Table 3 shows the human resources utilization in the base year for maintenances with and without resource sharing. It is apparent that the utilization efficiency for the six types of labors is increased, respectively. Table 4 compares equipments utilization in maintenances with and without resource sharing.

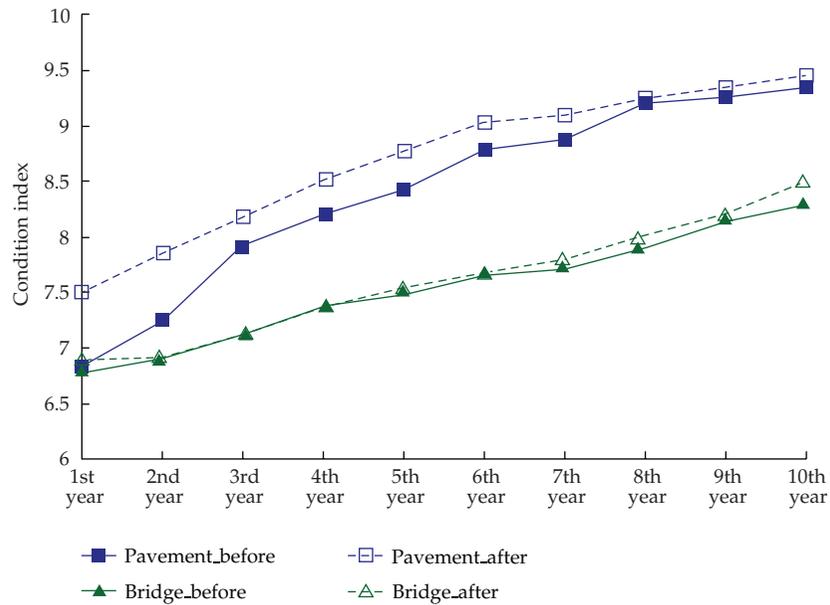
Figure 2 indicates the performance changes of pavement and bridge decks after maintenance. It can be easily seen that for all planning years, both pavement performance and bridge deck performance are improved with resources sharing.

4. Conclusions

The objective of highway asset maintenance management is to maximize the overall conditions of the highway system after maintenance with limited resources. This is achieved through integrated management for different types of highway facilities. The paper presents a new approach for optimal allocation of highway maintenance resources based on compromise programming. With the concepts of compromise programming, one can consider the relative importance of different highway facilities by introducing associated weights into the objective function. A bilevel model has been developed to analyze the

Table 4: Equipments utilization in highway maintenance with and without resource sharing (Working Days).

Equipment types	Maintenance without resource sharing			Maintenance under resource sharing	Total available equipments
	Pavement	Bridge deck	Total		
Dual-use mini vans	4.7	7.4	12.1	21.7	25.0
Backhoe/loaders	3.6	2.4	6.0	3.8	11.0
Pavement saws	2.9	8.6	11.5	19.8	23.0
Grinders	2.9	9.7	12.6	19.8	26.0
Compressors	6.5	1.2	7.7	14.2	18.5
Scrapers	1.8	0.0	1.8	1.9	5.0
Welding torch	0.0	1.2	1.2	0.0	3.0
Vibrating compactors	0.0	8.6	8.6	9.5	10.0
Cement mixers	0.0	1.2	1.2	0.0	3.2

**Figure 2:** Comparison of condition changes for pavement and bridge-decks with and without resource sharing.

resource allocation problem in pavement and bridge deck maintenance. Two scenarios in maintenance activities are analyzed: first without and then with sharing of resources between pavement maintenance and bridge deck maintenance. With a robust GA searching for an optimal solution to the problem, it is found that the performance of pavements and bridge decks improves significantly under resource sharing. The maintenance resources (e.g., funds, labor, and equipment) are utilized more efficiently in the resource-sharing scenario. The results of experimental analyses clearly show the promising features of the model in solving complex resource allocation problems in highway maintenance management. The method developed in this paper is useful to highway agencies in their decision-making activities such as developing maintenance programs, budgeting, and resource allocation.

Acknowledgment

The authors acknowledge the support from Beijing Municipal Natural Science Foundation with Grant no. 8092026.

References

- [1] FHWA, Asset Management Primer, Office of Asset Management, FHWA, US DOT, 1999.
- [2] FHWA, Annual Report 2001, Office of Asset Management, US DOT, 2001.
- [3] P. L. Yu, "A class of solutions for group decision problems," *Management Science*, vol. 19, no. 8, pp. 936–946, 1973.
- [4] M. Zeleny, "Compromise Programming," in *Multiple Criteria Decision Making*, J. L. Cochrane and M. Zeleny, Eds., pp. 262–301, University of South Carolina, Columbia, 1973.
- [5] Z. Lounis and M. Z. Cohn, "An engineering approach to multicriteria optimization of bridge structures," *Computer-Aided Civil and Infrastructure Engineering*, vol. 10, no. 4, pp. 233–238, 1995.
- [6] M. Zarghaami, "Integrated water resources management in Polrud irrigation system," *Water Resources Management*, vol. 20, no. 2, pp. 215–225, 2006.
- [7] L. Diaz-Balteiro, R. Voces, and C. Romero, "Making sustainability rankings using compromise programming. An application to European paper industry," *Silva Fennica*, vol. 45, no. 4, pp. 761–773, 2011.
- [8] M. Amiri, M. Ekhtiari, and M. Yazdani, "Nadir compromise programming: a model for optimization of multi-objective portfolio problem," *Expert Systems with Applications*, vol. 38, pp. 7222–7226, 2011.
- [9] F. J. Andre, M. A. Cardenete, and C. Romero, "Using compromise programming for macroeconomic policymaking in a general equilibrium framework: theory and application to the Spanish economy," *Journal of the Operational Research Society*, vol. 59, no. 7, pp. 875–883, 2008.
- [10] A. Hashimoto and D.-A. Wu, "A DEA-compromise programming model for comprehensive ranking," *Journal of the Operations Research Society of Japan*, vol. 47, no. 2, pp. 73–81, 2004.
- [11] J. T. Shiau and F. C. Wu, "Compromise programming methodology for determining instream flow under multiobjective water allocation criteria," *Journal of the American Water Resources Association*, vol. 42, no. 5, pp. 1179–1191, 2006.
- [12] P. Prodanovic and S. P. Simonovic, "Fuzzy compromise programming for group decision making," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 33, no. 3, pp. 358–365, 2003.
- [13] H. Xiong, "Optimal allocation of highway maintenance resources based on compromise programming," *Journal of Tsinghua University*, vol. 46, no. 12, pp. 1961–1964, 2006 (Chinese).
- [14] H. Xiong, "Multiobjective optimization of pavement management using compromise programming," *Journal of Beijing Institute of Technology*, vol. 26, no. 9, pp. 793–797, 2006 (Chinese).

Research Article

Equivalent Mechanical Model for Lateral Liquid Sloshing in Partially Filled Tank Vehicles

Zheng Xue-lian, Li Xian-sheng, and Ren Yuan-yuan

College of Traffic, Jilin University, No. 5988 Renmin Street, Changchun 130022, China

Correspondence should be addressed to Ren Yuan-yuan, maggie170101@gmail.com

Received 25 July 2012; Revised 17 September 2012; Accepted 10 October 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Zheng Xue-lian 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 reports a new approach to investigating sloshing forces and moments caused by liquid sloshing within partially filled tank vehicles subjected to lateral excitations. An equivalent mechanical model is used in the paper to approximately simulate liquid sloshing. The mechanical model is derived by calculating the trajectory of the center of gravity of the liquid bulk in tanks as the vehicle's lateral acceleration changes from 0 to 1 g. Parametric expressions for the model are obtained by matching the dynamic effect of the mechanical model to that of liquid sloshing. And parameter values of a liquid sloshing dynamic effect, such as sloshing frequency and forces, are acquired using FLUENT to simulate liquid sloshing in tanks with different cross-sections and liquid fill percentages. The equivalent mechanical model for liquid sloshing in tank vehicles is of a great significance for simplifying the research on roll stability of tank vehicles and for developing active/passive roll control systems for these vehicles.

1. Introduction

Road tank vehicles are commonly used in carrying a wide range of liquid cargoes, mainly of a dangerous nature, such as chemical and petroleum products. At the same time, they are more frequently involved in rollover-related road accidents, which can seriously harm peoples and the environment. Statistical data collected by Statistique Canada have shown that 83% of lorry rollover accidents on highways are caused by tank vehicles [1]. And a US study has reported that the average annual number of cargo tank rollovers is about 1265, which takes up 36.2% in the total number of heavy vehicle highway accidents [2].

Although there are many reasons that lead to tank vehicle rollover accidents, such as driver's fatigue, overtaking, bad road and weather conditions, and so forth, liquid sloshing in tanks is the main factor [3, 4]. Due to different liquid densities and axle load limits on roads, tanks are in a partially filled state for the majority of the time. This phenomenon

causes liquid sloshing in tanks when vehicle driving conditions change, meaning that strong sloshing forces are generated and vehicle roll stability is weakened. Therefore, research on liquid sloshing in partially filled tanks is one of the most important aspects when studying the roll stability of tank vehicles.

To date, many studies have been carried out on liquid flow and sloshing characteristics that happened in tanks, and the main methods can be summarized as follows.

- (1) The quasi-static (QS) method. The cargo's static moment at a specified point on a tank vehicle can be approximated by calculating the transient center of gravity (CG) of the liquid bulk in the tank. Then, the liquid sloshing effect on tank walls can be analyzed. It is convenient and simple to obtain liquid sloshing force using the QS method. However, the analysis results have poor accuracy [5–7].
- (2) The hydrodynamics method. By theoretically analyzing liquid flow characteristics in partially filled tanks, sloshing parameters can be acquired using basic hydrodynamic equations. Although the results so obtained are accurate, the analysis and the solution procedure are complicated. Due to the limited studies on turbulence and the fact that in reality the majority of flow can be categorized as turbulence, a large number of liquid flow phenomena cannot be explained using this method [8–12].
- (3) The experimental method. By building a test platform or using test tank vehicles, liquid sloshing phenomenon can be observed and relevant parameters can be monitored by reproducing liquid sloshing [13, 14]. The experimental results will depend on the test devices used, the sensor accuracy, and the operation of the tests, and so forth. And the method requires significant human and material resources.
- (4) Computer simulation. Simulation software is used to simulate liquid sloshing and to obtain the values of a corresponding sloshing dynamic effect [15, 16].
- (5) The equivalent mechanical model. Here, mechanical models are used to simulate liquid sloshing, which was created by NASA [17] and widely used for its simplicity and accuracy. Until now, most of the researches using this method have focused on spacecraft tanks and other vertical tanks [17–21]. Researches on horizontal tanks, such as those in tank vehicles, are limited [22–26].

By analyzing the present domestic and overseas conditions, the paper uses the equivalent mechanical model to simulate liquid sloshing in tank vehicles. The research outcomes have great importance for studying the roll stability of tank vehicles and for developing active/passive roll control systems for them.

2. Derivation of the Equivalent Mechanical Model

2.1. Mathematical Form of the Mechanical Model

Tanks with circular or oval cross-sections have larger volumes but the same surface area. Therefore, they are more popular in market applications and are the focus of study in this paper.

Theoretical analysis and experimental studies have shown that the first-order sloshing mode, which can be described by the oscillation of liquid-free surface, is the most important mode of liquid sloshing in partially filled tanks [18, 19]. Therefore, we start the research by studying liquid sloshing in a partially filled tank with different liquid fill levels and solving for the trajectory of the CG of the liquid bulk.

As shown in Figure 1, the tank cross-section is circular when $a/b = 1$ and oval when $a/b > 1$. In Figure 1, a is a half of the tank width, b is a half of the tank height, h_0 is the intersection point between the liquid level and the y -axis, and φ is the tilt angle of the liquid-free surface.

Define the ratio of the height of the liquid level to the tank height as the liquid fill percentage or fill level in tanks, which can be expressed by

$$\text{liquid fill percentage} = \frac{(h_0 + b)}{2b} = \Delta. \quad (2.1)$$

Locus of the CG of the liquid bulk can be obtained from the following equations:

$$X = \frac{\int_{x_1}^{x_2} \int_{y_1}^{y_2} x \, dy \, dx}{\int_{x_1}^{x_2} \int_{y_1}^{y_2} dy \, dx}, \quad (2.2)$$

$$Y = \frac{\int_{x_1}^{x_2} \int_{y_1}^{y_2} y \, dy \, dx}{\int_{x_1}^{x_2} \int_{y_1}^{y_2} dy \, dx}.$$

And the cross-sectional area of the liquid in the tank can be expressed by

$$A = \frac{3ah_0}{b} \sqrt{b^2 - h_0^2} + ab \arcsin \frac{\sqrt{b^2 - h_0^2}}{b}, \quad (2.3)$$

when $h_0 < 0$, and

$$A = \frac{a}{b} \left[h_0 \sqrt{b^2 - h_0^2} + b^2 \left(\arcsin \frac{h_0}{b} + \frac{\pi}{2} \right) \right], \quad (2.4)$$

when $h_0 > 0$.

The intersection point between the liquid-free surface and the y -axis, which is defined as h , changes with each tilt angle of the liquid-free surface. Therefore, the liquid-free surface in the (x, y) coordinate system can be described by

$$y = x \tan \varphi + h. \quad (2.5)$$

The intersection points of the liquid-free surface with the tank periphery are given by

$$\left(\frac{-a^2nh + ab\sqrt{a^2n^2 + b^2 - h^2}}{a^2n^2 + b^2}, \frac{b^2h + abn\sqrt{a^2n^2 + b^2 - h^2}}{a^2n^2 + b^2} \right), \quad (2.6)$$

$$\left(\frac{-a^2nh - ab\sqrt{a^2n^2 + b^2 - h^2}}{a^2n^2 + b^2}, \frac{b^2h - abn\sqrt{a^2n^2 + b^2 - h^2}}{a^2n^2 + b^2} \right).$$

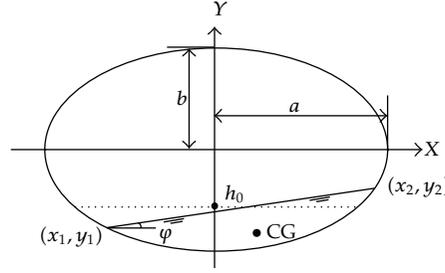


Figure 1: Schematic diagram for a partially filled tank with circular or oval cross-section.

The cross-sectional area of the liquid, which is defined as S , and its static moments on the x -axis and the y -axis, can be obtained from (2.3)–(2.6). And the acquired functions are all functions of h .

Regardless of the tilt angle of the liquid-free surface, the cross-sectional area of the liquid remains constant. Make h vary within a given range with a quite small step size and calculate S and its static moments at each value of h . Then, the CG of the liquid bulk can be obtained using (2.2), ensuring that the determinant condition of $|S - A| \leq \delta$ (δ is a very small positive value depending on the step size of h) is satisfied.

The trajectory of the CG of the liquid bulk while the tilt angle of the liquid-free surface varies over a suitable range is shown in Figure 2, which shows that the trajectory of the CG of the liquid bulk remains parallel to the tank periphery.

In a vehicle's roll stability analysis using QS method, the liquid sloshing effect can be approximated by the static moment of the liquid bulk at a specified point on the tank vehicle [5–7]. The results have great errors from the actual condition, which cannot be neglected. However, simple mechanical devices, such as springs or pendulums, not only accurately calculate the liquid sloshing force and its influence on tank vehicles but also reflect liquid sloshing characteristics. For the problem discussed in this paper, the trammel pendulum, whose oscillation trajectory is an ellipse, is more appropriate; see [25].

2.2. Equations of Motion for the Trammel Pendulum

The oscillation trajectory and basic parameters of the trammel pendulum are shown in Figure 3. Suppose that the pendulum's oscillation trajectory is different from that of the CG of the liquid bulk, then $a_p + b_p$ is the arm length of the pendulum, where a_p is a half of the major axis of the pendulum's oscillation trajectory and b_p is a half of its minor axis. a_{cg} is a half of the major axis of the elliptical trajectory of the CG of the liquid bulk and b_{cg} is a half of its minor axis. θ is the pendulum amplitude, which is the maximum angle the pendulum swings away from the vertical position. And α is the angle between the line that connects the origin to the pendulum mass which is short for the mass of the bob on a pendulum and the y -axis.

The tank periphery, the oscillation trajectory of the pendulum, and the CG of the liquid bulk are all parallel to each other, which can be expressed as follows:

$$\frac{a}{b} = \frac{a_{cg}}{b_{cg}} = \frac{a_p}{b_p} = \Lambda. \quad (2.7)$$

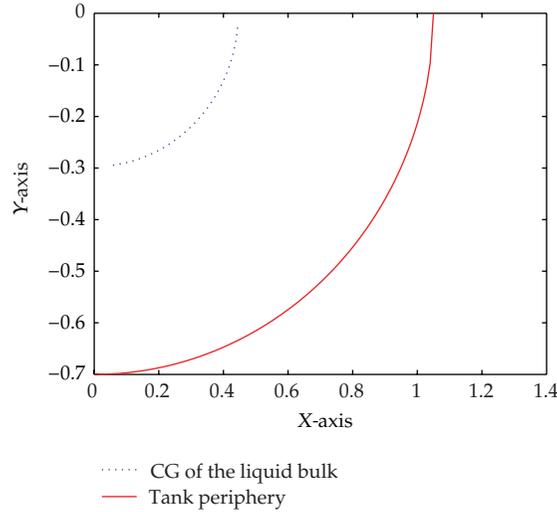


Figure 2: The trajectory of the CG of the liquid bulk.

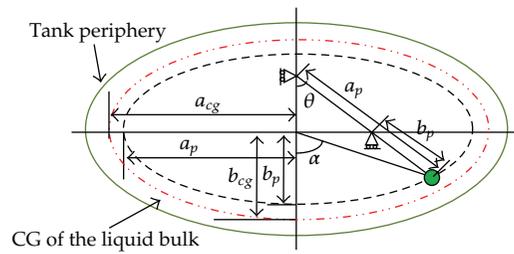


Figure 3: Schematic diagram for oscillation trajectory and basic parameters of the trammel pendulum.

The trammel pendulum’s oscillation is affected by its arm length, amplitude, and the vehicle’s lateral acceleration, not the pendulum mass.

The motion analysis for the trammel pendulum is shown in Figure 4, where xy is the tank-fixed coordinate and XY is the earth-fixed coordinate. l is the distance between the origin of XY and that of xy .

According to Figure 4, the absolute location of the pendulum mass can be expressed as

$$\vec{r} = (a_p \sin \theta + l)\vec{i} - b_p \cos \theta \vec{j}. \tag{2.8}$$

Therefore, the velocity and acceleration of the pendulum mass can be expressed as follows:

$$\dot{\vec{r}} = (a_p \dot{\theta} \cos \theta + \dot{l})\vec{i} + b_p \dot{\theta} \sin \theta \vec{j}, \tag{2.9}$$

$$\ddot{\vec{r}} = (\ddot{l} + a_p \ddot{\theta} \cos \theta - a_p \dot{\theta}^2 \sin \theta)\vec{i} + b_p (\ddot{\theta} \sin \theta + \dot{\theta}^2 \cos \theta)\vec{j}. \tag{2.10}$$

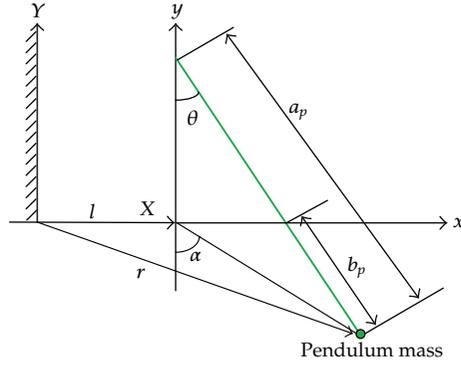


Figure 4: Diagram for motion analysis of the trammel pendulum.

The kinetic energy of the moving pendulum mass is defined by

$$T = \frac{1}{2}mv^2 = \frac{1}{2}m\left(a_p^2\dot{\theta}^2\cos^2\theta + \dot{l}^2 + 2a_p\dot{l}\dot{\theta}\cos\theta + b_p^2\dot{\theta}^2\sin^2\theta\right). \quad (2.11)$$

Assume that the zero of the potential energy is located at the surface of the equilibrium position of the trammel pendulum. Therefore, the gravitational potential energy of the trammel pendulum can be expressed as

$$Q = mgb_p(1 - \cos\theta). \quad (2.12)$$

According to (2.11)-(2.12), a Lagrangian function can be used to obtain the kinetic equation for the pendulum system, which can be written as follows:

$$L = T - Q = \frac{1}{2}m\left(a_p^2\dot{\theta}^2\cos^2\theta + \dot{l}^2 + 2a_p\dot{l}\dot{\theta}\cos\theta + b_p^2\dot{\theta}^2\sin^2\theta\right) + mgb_p(\cos\theta - 1). \quad (2.13)$$

The motion of the trammel pendulum system can be expressed by

$$\frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0, \quad (2.14)$$

where

$$\frac{\partial L}{\partial \dot{\theta}} = m\left(a_p^2\dot{\theta}\cos^2\theta + a_p\dot{l}\cos\theta + b_p^2\dot{\theta}\sin^2\theta\right), \quad (2.15)$$

$$\begin{aligned} \frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \dot{\theta}}\right) &= m\left(a_p^2\ddot{\theta}\cos^2\theta - a_p^2\dot{\theta}^2\sin 2\theta + a_p\ddot{l}\cos\theta - a_p\dot{l}\dot{\theta}\sin\theta + b_p^2\ddot{\theta}\sin^2\theta + b_p^2\dot{\theta}^2\sin 2\theta\right), \\ \frac{\partial L}{\partial \theta} &= m\left(-0.5a_p^2\dot{\theta}^2\sin 2\theta - a_p\dot{l}\dot{\theta}\sin\theta + 0.5b_p^2\dot{\theta}^2\sin 2\theta\right) - mgb_p\sin\theta. \end{aligned} \quad (2.16)$$

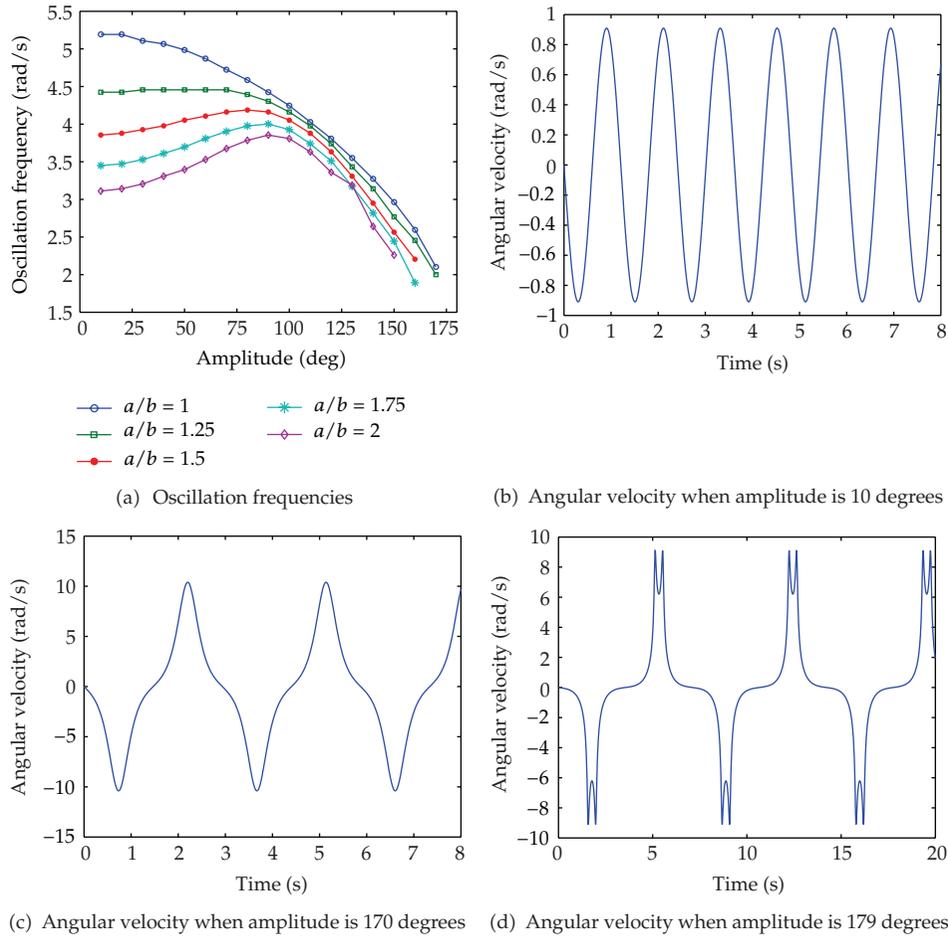


Figure 5: Motion characteristics of trammel pendulum.

Substituting (2.16) into (2.14) we get

$$\left(a_p^2 \cos^2 \theta + b_p^2 \sin^2 \theta\right) \ddot{\theta} + \frac{1}{2} \left(b_p^2 - a_p^2\right) \theta^2 \sin 2\theta + g b_p \sin \theta + \dot{l} a_p \cos \theta = 0. \quad (2.17)$$

MATLAB's ODE algorithm is used to solve (2.17). During the solution procedure, we make a/b varies between 1 and 2 with a 0.25 step size and the pendulum amplitude varies between 10 degrees and 180 degrees with a 10-degrees step size.

Oscillation frequencies and angular velocities for pendulums with small and large amplitudes are presented in Figure 5. Tanks with different cross-sections have the same cross-sectional area and $a = b = 0.3602$ m when the cross-section is circular ($a/b = 1$).

As Figure 5(a) shows the pendulum's oscillation frequency depends on its arm length and amplitude, the oscillation frequency decreases with an increase in amplitude when $a/b = 1$. However, for the other pendulums, the oscillation frequency rises with an increase in amplitude, reaching the maximum frequency when the amplitude reaches a certain value, and then decreasing after that. For instance, the maximum frequency appears at an amplitude

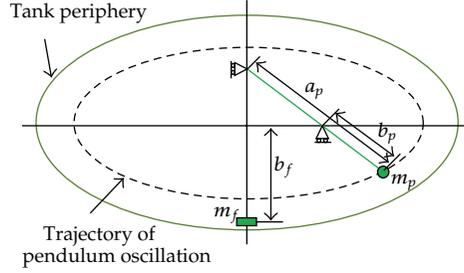


Figure 6: Parameters to be determined for the pendulum.

of 90 degrees when $a/b = 2$. For all pendulums, the oscillation frequency remains almost constant when the amplitude is below a certain value.

As seen in Figures 5(b)–5(d), with an increase in amplitude, the motion of the pendulum becomes irregular with more nonlinearity, especially for the amplitude larger than 170 degrees. Fortunately, the lateral acceleration of the tank vehicles is smaller than 0.45 g in reality to avoid vehicle rollovers, which means that the tilt angle of the liquid-free surface is always smaller than 90 degrees. Therefore, the nonlinear characteristics of the pendulum can be neglected and the pendulum can be assumed approximately linear.

3. Parametric Expressions for the Trammel Pendulum Model

The parameters that need to be determined for the pendulum are presented in Figure 6, where m_p is the pendulum mass, m_f is the fixed liquid mass, and b_f is the distance between the center of the ellipse and the location of the fixed liquid mass.

Due to the fact that not all of the liquid participate in the sloshing [11–14, 17, 18, 20], the pendulum parameters a_p , b_p , and m_p are not equal to a_{cg} , b_{cg} , and the liquid mass m , respectively.

Because the pendulum parameters cannot be obtained directly, and taking into consideration that the pendulum parameters must have relations with the liquid sloshing parameters, analogy method is used in obtaining parametric expressions for the mechanical model.

3.1. Derivation of the Pendulum Arm Length Parameters

According to Section 2.2, the pendulum's oscillation frequency partly depends on its arm length which can be expressed as $a_p + b_p$. For a partially filled tank with a specified cross-section, the liquid sloshing frequency is known and, using (2.7), a_p and b_p can be obtained. Therefore, (2.7) and (2.17) are sufficient to derive the pendulum arm length parameters.

Given the fact that the oscillation frequency remains almost constant when the pendulum amplitude is quite small, define the following quantities:

$$\begin{aligned} \sin \theta &\approx \theta, & \cos \theta &\approx 1, & \sin^2 \theta &\approx 0, \\ \cos^2 \theta &\approx 1, & \sin 2\theta &\approx 2\theta. \end{aligned} \quad (3.1)$$

Then, (2.17) can be rewritten as follows (see in [25]):

$$\ddot{\theta} + \left(\frac{b_p^2 - a_p^2}{a_p^2} \right) \dot{\theta}^2 \theta + \frac{g b_p}{a_p^2} \theta = 0. \quad (3.2)$$

For analytical simplicity, define the following quantities:

$$\theta = x; \quad \frac{dx}{dt} = y. \quad (3.3)$$

Then, (3.2) can be rewritten as follows:

$$\frac{dy}{dt} = C_1 y^2 x - C_2 x, \quad (3.4)$$

where $C_1 = -((b_p^2 - a_p^2)/a_p^2)$, $C_2 = g b_p / a_p^2$.

Now, the orientation field equation for the trammel pendulum can be expressed as

$$\frac{dy}{dx} = \frac{C_1 y^2 x - C_2 x}{y}. \quad (3.5)$$

Equation (3.5) can be transformed into the following form:

$$\frac{y dy}{C_1 y^2 - C_2} = x dx. \quad (3.6)$$

By solving (3.6), the phase plane trajectory equation for the pendulum can be written as follows:

$$y = \frac{\pm \sqrt{C_1 (C_2 + A e^{C_1 x^2})}}{C_1}, \quad (3.7)$$

where A is an integral constant that can be obtained by setting $x = x_{\max}$ and $y = 0$.

Substituting A into (3.7) gives

$$y = \frac{\pm \sqrt{C_1 C_2 (1 - e^{C_1 (x^2 - x_{\max}^2)})}}{C_1}. \quad (3.8)$$

Based on (3.8), the phase trajectories for trammel pendulums with different amplitudes are presented in Figure 7. It is concluded that the pendulum system moves back and forth and keeps a circular motion with the same amplitude.

Now, define the following quantities for (3.8):

$$C_3 = \sqrt{\frac{C_2}{C_1}}; \quad z = C_1(x^2 - x_{\max}^2). \quad (3.9)$$

Then, (3.8) can be rewritten as follows:

$$y = C_3(1 - e^z)^{1/2}. \quad (3.10)$$

Rewriting (3.10) using a Taylor series expansion and neglecting higher order terms gives

$$y = C_3 \sqrt{C_1(x_{\max}^2 - x^2)} = \sqrt{\frac{gb_p}{a_p^2}} \sqrt{x_{\max}^2 - x^2}. \quad (3.11)$$

At the instance of tanks with circular cross-section, the coefficient of (3.11) is equal to $\sqrt{g/a_p}$, which is the frequency expression for simple pendulums. Therefore, the phase plane trajectory equation of a simple pendulum is thus given by

$$y = \omega \sqrt{x_{\max}^2 - x^2}. \quad (3.12)$$

Comparing (3.11) to (3.12), the natural oscillation frequency of the trammel pendulum with small amplitude can be expressed as follows:

$$\omega = \sqrt{gb_p/a_p^2}. \quad (3.13)$$

The oscillation frequencies of pendulums with small amplitudes obtained from (2.17) are used to verify the accuracy of (3.13). The results show that (3.13) is very consistent with (2.17).

As the liquid sloshing frequency in a partially filled tank vehicle is already known, a_p and b_p can easily be obtained based on (2.7) and (3.13).

3.2. Derivation of the Pendulum Mass Parameters

The lateral liquid sloshing force in a tank vehicle is caused by the liquid mass that participates in the sloshing. When a pendulum is used to simulate the liquid sloshing, the liquid mass that participates in the sloshing is equal to the pendulum mass. According to the law of conservation of mass, the liquid mass that does not participate in the sloshing is equal to the fixed part.

A direct solution for the pendulum mass is difficult and requires hydrodynamic theory analysis. Thus, an alternative method is used.

First of all, suppose that all of the liquid mass participate in the sloshing. If the maximum lateral acceleration of the liquid bulk is known, then the sloshing force of the entire

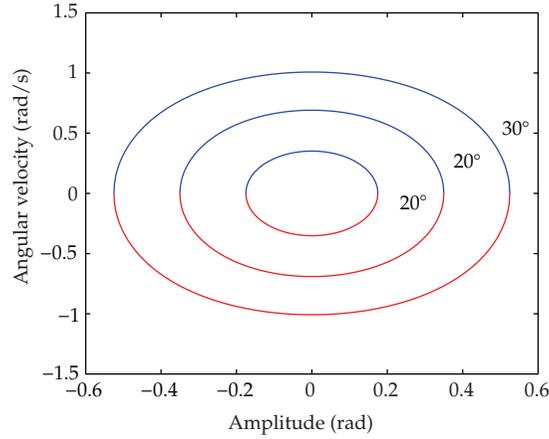


Figure 7: Phase trajectories for trammel pendulum with different amplitudes.

liquid mass can be obtained. By comparing it with the actual sloshing force, the ratio of the pendulum mass to the entire liquid mass can be acquired. Since the entire liquid mass is known, the pendulum mass can thus be calculated.

According to Newton's second law, the sloshing force can be expressed by

$$F_t = \max[ma_y], \quad (3.14)$$

where F_t is the sloshing force caused by the entire liquid mass, a_y is the maximum lateral acceleration of the liquid bulk, and m is the entire liquid mass.

Assume that the tank length is 1 m and the liquid density is known. Then, given the size of the tank cross-section and the value of the liquid fill percentage, m can be obtained.

Make the lateral acceleration of the tank vehicle equal to zero and liquid oscillate only under the action of gravity. According to (2.10), the maximum lateral acceleration of the liquid can be expressed as

$$a_y = \max[a_p(-\dot{\theta} \sin \theta + \ddot{\theta} \cos \theta)]. \quad (3.15)$$

From (3.14)-(3.15), the maximum sloshing force caused by the entire liquid mass can be obtained.

The actual sloshing force is given by

$$F_p = \max[m_p a_y]. \quad (3.16)$$

Equation (3.16) divided by (3.14) gives

$$\frac{m_p}{m} = \frac{F_p}{F_t}. \quad (3.17)$$

Then, the fixed liquid mass is

$$m_f = m - m_p. \quad (3.18)$$

No matter where the locations of m_f and m_s are, the action point of the two parts always coincides with that of the CG of the entire liquid mass. Therefore, the sum of the static moments of m_f and m_s at the lowest point of the tank is equal to that of m at the same point when the liquid-free surface is level, which can be expressed by

$$m_f(b - b_f) + m_p(b - b_p) = m(b - b_{cg}). \quad (3.19)$$

If all the other parameters are already known, then b_f can be obtained from (3.19).

4. Simulation and Discussion

4.1. Settings for FLUENT Simulation Conditions

Based on Section 3.1, parameters used to describe the liquid sloshing dynamic effect, such as the sloshing frequency and the maximum lateral sloshing force, will be obtained in this section to completely specify the equivalent mechanical model.

The FLUENT software is used to simulate liquid sloshing that occurs in tank vehicles and to obtain the values of relevant parameters. Before performing the simulation, the sizes of tank cross-sections should be decided and the corresponding simulation conditions should be set.

According to a market survey, the cross-sectional area of oval tanks is usually just under 2.4 m^2 . *XH9140G*, a typical tank semitrailer of *PieXin* brand, is chosen as the simulation object [27]. The long axis of the tank cross-section is 2.3 m and the short one is 1.3 m; the tank's wall thickness is neglected. According to the principle that tanks with different cross-sections have the same surface area, the sizes of the tank cross-sections are presented in Table 1.

The tilt angle of the liquid-free surface is set to be 5 degrees to maintain the linear characteristics of the pendulum and to ensure that the liquid will oscillate gently under the action of gravity. Water is chosen as the simulation liquid. The liquid fill percentage is set to vary from 10% to 90% with a 10% step size.

The maximum velocity of water can be obtained when it moves to the lowest position in the tank, which is presented as follows:

$$v = \sqrt{2g\Delta h}, \quad (4.1)$$

where v is the water's velocity, and Δh is the vertical distance that the CG of the water bulk moves.

Suppose that $b_{cg} = 0.2$, which is a quite small value compared with b , the flow Reynolds number can be expressed as follows:

$$\text{Re} = \frac{Dv}{\nu} = 2.44 \times 10^5 \gg 2000. \quad (4.2)$$

Table 1: Sizes of tank cross-sections (unit: m).

	$a/b = 1$	$a/b = 1.25$	$a/b = 1.5$	$a/b = 1.75$	$a/b = 2$
a	0.857	0.9585	1.05	1.134	1.2124
b	0.857	0.7668	0.7	0.6481	0.6062

According to (4.2), the liquid sloshing that occurs in the tank vehicles can be categorized as turbulence.

Although the liquid velocity in the region near to the wall is quite low and its order of magnitude is around 10^{-2} , the turbulence characteristic of the water flow is still quite apparent. In order to choose the standard wall function for the near-wall treatment of the viscous model, the meshing for the tank model must be qualified and the wall $Y\text{-}plus$, which is the dimensionless distance between the CG of the first layer of the grid and the wall, should be within the range from 10 to 100.

Based on real-life conditions, the reference pressure location for the operating conditions is in the pressure inlet and the gravity is 9.81 m/s^2 in the negative direction of the y -axis.

The intensity and hydraulic diameter are chosen as the turbulence specification method. For the liquid that oscillates freely under the action of gravity generated by the small tilt angle of the liquid-free surface, the turbulence intensity will be within the range 0.1%–0.5%.

The hydraulic diameter is calculated as follows:

$$D = \frac{4A}{\chi}, \quad (4.3)$$

where D is the hydraulic diameter, A is the cross-sectional area of the liquid, and χ is the wetted perimeter.

The PISO algorithm and the Body Force Weighted method are chosen for the pressure-velocity coupling and the pressure discretization, respectively; see [28].

According to the above settings, a schematic diagram for the liquid sloshing model can be obtained and is shown in Figure 8.

To obtain the cycle time of the liquid sloshing, a point located in the tank that will always be immersed in the liquid is specified. Then, the cycle time of the liquid sloshing can be obtained by monitoring the lateral velocity of this point.

4.2. Simulation Results

The cycle time of the lateral velocity of a point $(-0.84, 0)$ in a tank with circular cross-section and 20% liquid fill level is presented in Figure 9.

The relation between the cycle time and the angular frequency is given by

$$\omega = \frac{2\pi}{T}. \quad (4.4)$$

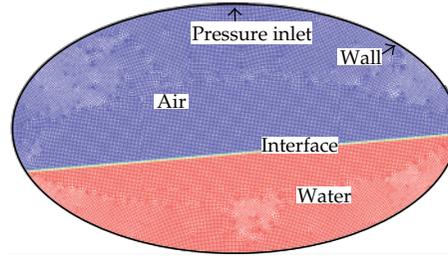


Figure 8: Schematic diagram for fluid sloshing model.

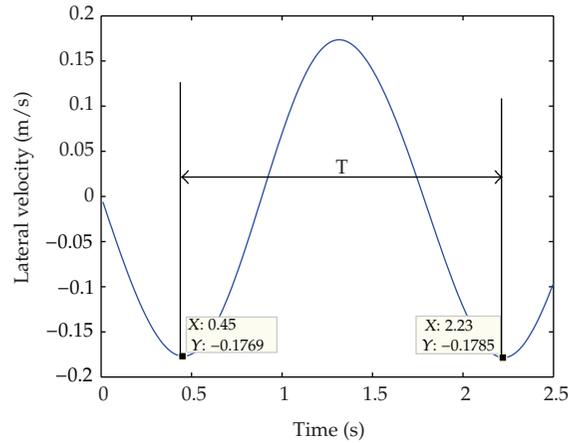


Figure 9: Lateral velocity of a point in a tank with circular cross-section and 20% liquid fill level.

The natural frequencies of liquid sloshing in tanks with different cross-sections and liquid fill levels are obtained using (4.4) and plotted in Figure 10.

Equation (3.13) can be rewritten as follows:

$$\frac{b_p}{b} = \frac{g}{\omega^2 r^2 b}. \quad (4.5)$$

As the sloshing frequencies are already known, b_p/b is obtained and plotted in Figure 11.

Curves fitting is done to the data points in Figure 11 to obtain an equation that describes b_p/b as a function of the fill percentage and the tank cross-section. This equation is as follows:

$$\begin{aligned} \frac{b_p}{b} = & 1.089 + 0.726\Delta - 0.1379\Lambda - 0.953\Delta^2 - 1.216\Lambda\Delta \\ & + 0.05141\Lambda^2 - 0.06107\Delta^3 + 0.5739\Lambda\Delta^2 + 0.1632\Lambda^2\Delta. \end{aligned} \quad (4.6)$$

The curves specified by (4.6) are plotted in Figure 12. And the relative error of the curves fitting for b_p/b is plotted in Figure 13.

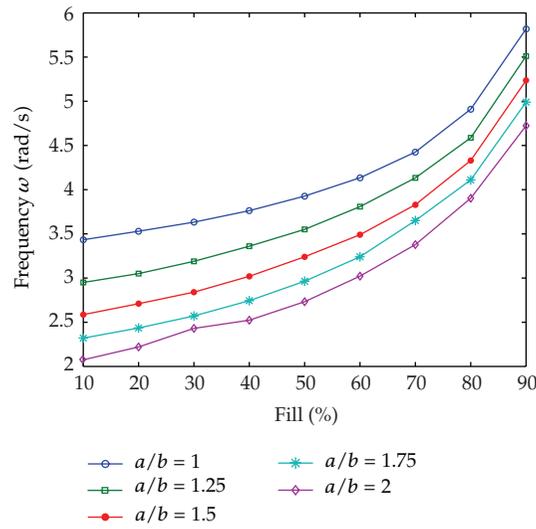


Figure 10: Natural frequencies of liquid sloshing.

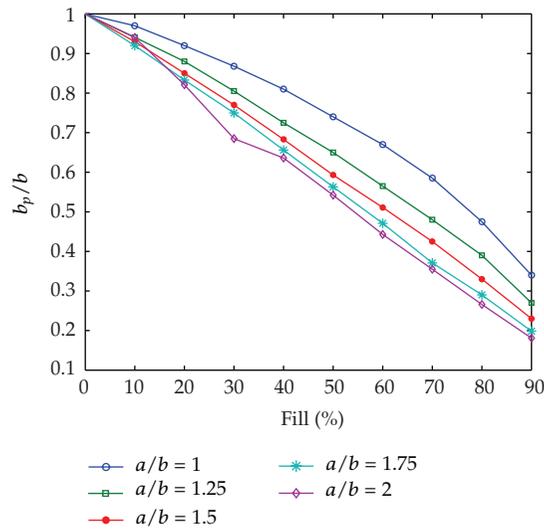


Figure 11: Values of b_p/b .

a_p can be obtained from (2.7) and (4.6). As a result, all of the pendulum arm length parameters have been obtained.

The maximum sloshing force during the liquid sloshing process can be obtained by monitoring the lateral force coefficient for the tank walls, and the results are presented in Figure 14. This shows that the maximum sloshing force is generated when the liquid fill percentage is close to 60%. Therefore, for tank vehicles, a liquid fill percentage close to 60% is the worst laden state.

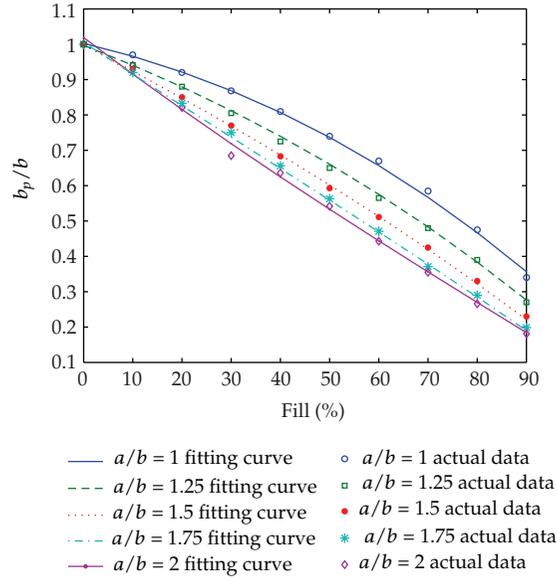


Figure 12: Fitting curves for b_p/b .

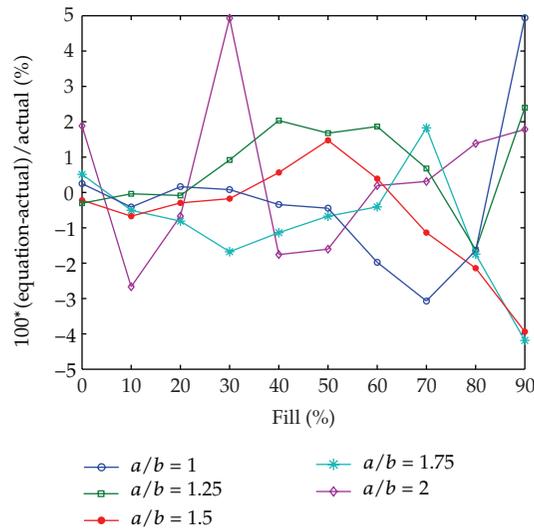


Figure 13: The relative error of the curves fitting for b_p/b .

The ratio of the sloshing force to the liquid mass is presented in Figure 15. This shows that the lower is the liquid fill percentage; the larger is the sloshing force generated by per unit of liquid mass.

To solve the sloshing force of the entire liquid mass, the maximum lateral acceleration is needed. According to (2.17), the pendulum amplitude is needed to be obtained at first.

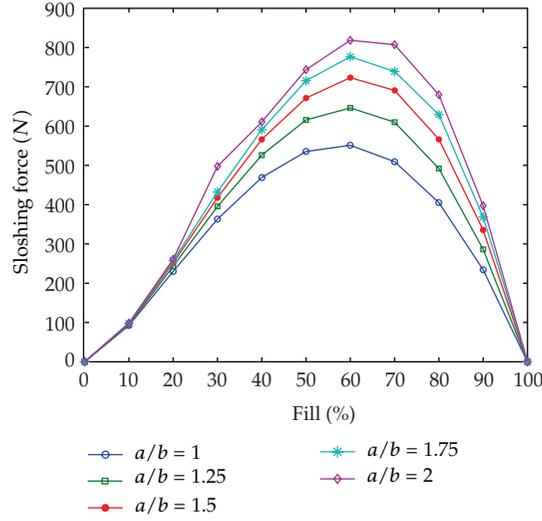


Figure 14: Maximum sloshing forces.

Table 2: Pendulum amplitudes in tanks with different cross-sections.

	$a/b = 1$	$a/b = 1.25$	$a/b = 1.5$	$a/b = 1.75$	$a/b = 2$
α	5	7.788	11.133	15	19.287
θ	5	6.244	7.474	8.705	9.925

However, the pendulum arm is not perpendicular to the liquid-free surface, except for tanks with a circular cross-section. According to Figure 4, the following equation can be obtained:

$$\theta = \tan^{-1}\left(\frac{b}{a} \tan \alpha\right). \tag{4.7}$$

α that exist in (4.7) can be obtained by solving the CG of the liquid bulk. For 5 degrees tilt angle of the liquid-free surface, α and θ are as listed in Table 2.

The maximum lateral acceleration can be obtained using (2.17), (3.15), and Table 2. Then, the lateral sloshing force for the entire liquid mass can be calculated using (3.14). Finally, m_p/m can be derived from (3.17) and these values are presented in Figure 16.

Curves fitting is done to the points in Figure 16 to obtain an equation that describes m_p/m as a function of the tank cross-section and the liquid fill percentage. The fitted equation is given by the following expression:

$$\begin{aligned} \frac{m_p}{m} = & 0.7844 - 1.729\Delta + 0.3351\Lambda + 1.156\Delta^2 + 0.7256\Lambda\Delta \\ & - 0.1254\Lambda^2 - 0.3219\Delta^3 - 0.9152\Lambda\Delta^2 + 0.08043\Lambda^2\Delta. \end{aligned} \tag{4.8}$$

The curves given by (4.8) are plotted in Figure 17. And the relative error of the curves fitting for m_p/m is plotted in Figure 18.

Given the pendulum mass, the fixed liquid mass can be obtained using (3.18).

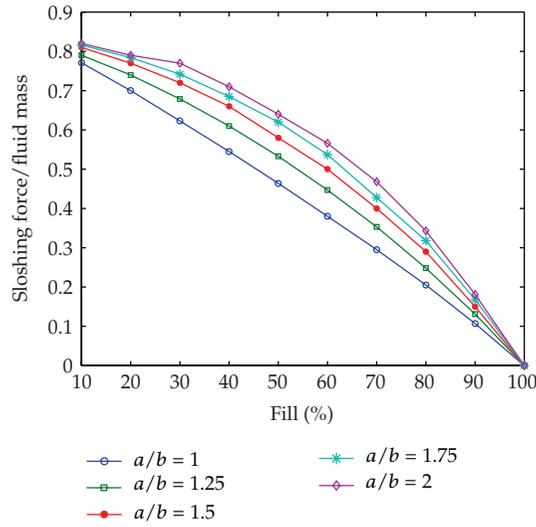


Figure 15: Sloshing force generated by per unit of liquid mass.

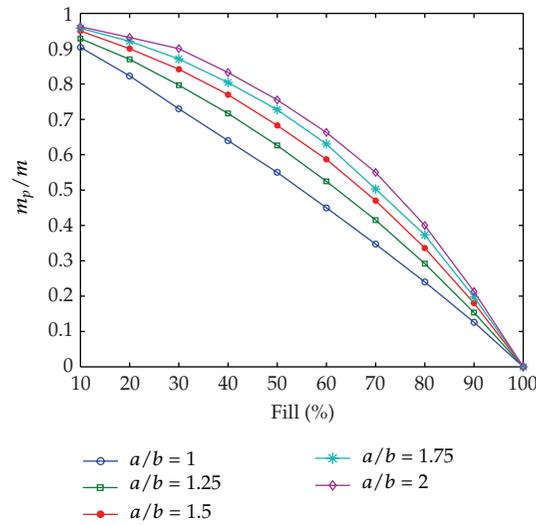


Figure 16: Values of m_p/m .

According to (3.19), the position of the fixed liquid mass is given by

$$\frac{b - b_f}{b} = \frac{m(b - b_{cg}) - m_f(b - b_p)}{m_f b}. \tag{4.9}$$

The curves given by (4.9) are presented in Figure 19. It shows that the position of the fixed liquid mass is close to the center of the ellipse, except when the liquid fill percentage is below 30%. And some points that are derived apparently from the equation curve are marked

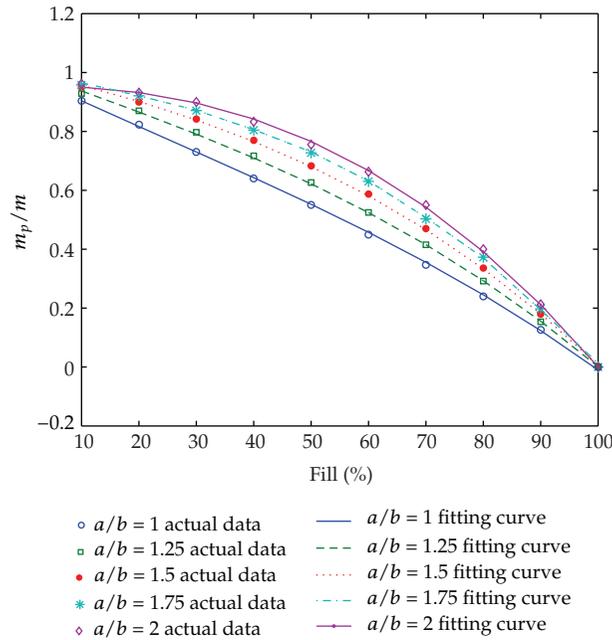


Figure 17: Fitting curves for m_p/m .

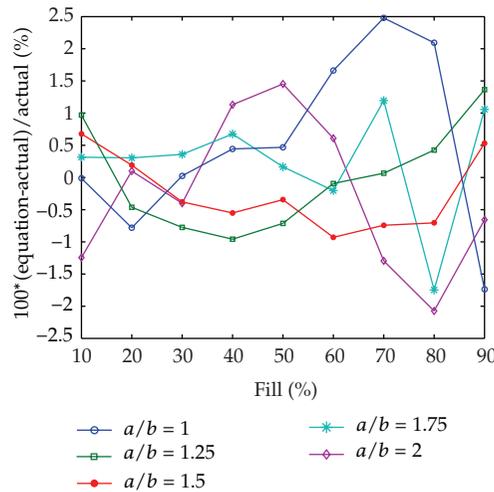


Figure 18: The relative error of the curves fitting for m_p/m .

in Figure 19. The reason why this phenomenon happens should be investigated in a future study.

4.3. Conclusions

To deal with the complexity of analyzing a liquid sloshing dynamic effect in partially filled tank vehicles, the paper uses equivalent mechanical model to simulate liquid sloshing.

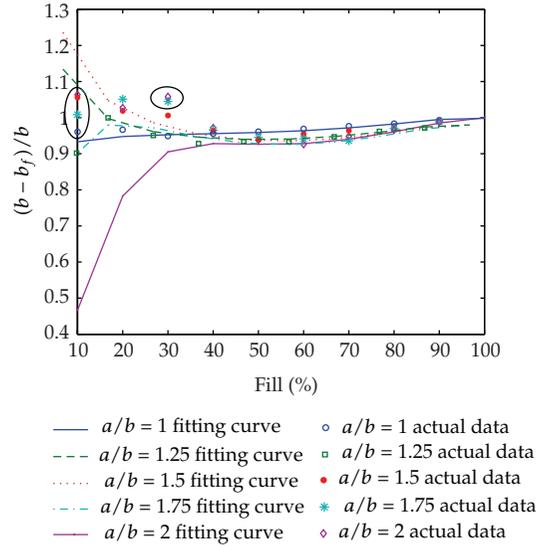


Figure 19: Location of the fixed liquid mass.

For tanks with circular or oval cross-sections, a trammel pendulum mechanical model is derived and parametric expressions for it are obtained through analogy analysis and FLUENT simulations. The establishment of the equivalent mechanical model for lateral liquid sloshing in partially filled tank vehicles has a great importance for accurately analyzing the roll stability of tank vehicles, as well as for developing active/passive roll control systems for them.

The following important discoveries were made from the FLUENT simulations.

- (1) For tanks with equal cross-sectional area and liquid fill percentages, tanks with a circular cross-section are subject to the lowest liquid sloshing forces.
- (2) For all of the tanks, the maximum liquid sloshing force is produced when the liquid fill percentage is close to 60%. Lower or higher fill percentages cause relatively less sloshing force.
- (3) The lower the liquid fill percentage is, the larger the liquid sloshing force produced by per unit of liquid mass is.

Since we make the assumption in deriving the equation of motion for the trammel pendulum that the pendulum amplitude is quite small and the motion of the pendulum is linear, the pendulum model is limited in analyzing liquid sloshing in tank vehicles when the vehicle subjects to gently lateral excitations only. Thus, an equivalent mechanical model for liquid sloshing which can describe nonlinear characteristics will be conducted in a future study.

Acknowledgment

This paper is supported by Project 20121099 supported by the Graduate Innovation Fund of Jilin University.

References

- [1] J. Woodrooffe, "Evaluation of dangerous goods vehicle safety performance," Report TP 13678-E, Transport Canada, 2000.
- [2] B. P. Douglas, H. Kate, M. Nancy et al., "Cargo tank roll stability study: final report," Tech. Rep. GS23-0011L, U.S. Department of Transportation, 2007.
- [3] W. H. Wang, Q. Cao, K. Ikeuchi, and H. Bubb, "Reliability and safety analysis methodology for identification of drivers' erroneous actions," *International Journal of Automotive Technology*, vol. 11, no. 6, pp. 873–881, 2010.
- [4] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research Part C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [5] R. Ranganathan, S. Rakheja, and S. Sankar, "Steady turning stability of partially filled tank vehicles with arbitrary tank geometry," *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, vol. 111, no. 3, pp. 481–489, 1989.
- [6] X. Kang, S. Rakheja, and I. Stiharu, "Optimal tank geometry to enhance static roll stability of partially filled tank vehicles," SAE Technical Paper 1999-01-3730, 1999.
- [7] X. Kang, S. Rakheja, and I. Stiharu, "Cargo load shift and its influence on tank vehicle dynamics under braking and turning," *International Journal of Heavy Vehicle Systems*, vol. 9, no. 3, pp. 173–203, 2002.
- [8] G. Popov, S. Sankar, T. S. Sankar, and G. H. Vatisstas, "Dynamics of liquid sloshing in horizontal cylindrical road containers," *Proceedings of the Institution of Mechanical Engineers, Part C*, vol. 207, no. 6, pp. 399–406, 1993.
- [9] H. Akyildiz, "A numerical study of the effects of the vertical baffle on liquid sloshing in two-dimensional rectangular tank," *Journal of Sound and Vibration*, vol. 331, pp. 41–52, 2012.
- [10] S. M. Hasheminejad and M. Aghabeigi, "Liquid sloshing in half-full horizontal elliptical tanks," *Journal of Sound and Vibration*, vol. 324, no. 1-2, pp. 332–349, 2009.
- [11] W. Rumold, "Modeling and simulation of vehicles carrying liquid cargo," *Multibody System Dynamics*, vol. 5, no. 4, pp. 351–374, 2001.
- [12] M. Toumi, M. Bouazara, and M. J. Richard, "Impact of liquid sloshing on the behaviour of vehicles carrying liquid cargo," *European Journal of Mechanics, A*, vol. 28, no. 5, pp. 1026–1034, 2009.
- [13] G. Yan, S. Rakheja, and K. Siddiqui, "Experimental study of liquid slosh dynamics in a partially-filled tank," *Journal of Fluids Engineering, Transactions of the ASME*, vol. 131, no. 7, Article ID 071303, 14 pages, 2009.
- [14] J. A. Romero, O. Ramírez, J. M. Fortanell, M. Martínez, and A. Lozano, "Analysis of lateral sloshing forces within road containers with high fill levels," *Proceedings of the Institution of Mechanical Engineers, Part D*, vol. 220, no. 3, pp. 302–312, 2006.
- [15] K. Modaresi-Tehrani, S. Rakheja, and R. Sedaghati, "Analysis of the overturning moment caused by transient liquid slosh inside a partly filled moving tank," *Proceedings of the Institution of Mechanical Engineers, Part D*, vol. 220, no. 3, pp. 289–301, 2006.
- [16] C. Y. Shang and J. C. Zhao, "Studies on liquid sloshing in rigid containers using FLUENT code," *Journal of Shanghai Jiaotong University*, vol. 42, no. 6, pp. 953–956, 2008.
- [17] F. T. Dodge, "Analytical representation of lateral sloshing by equivalent mechanical models," in *The Dynamic Behavior of Liquids in Moving Containers*, H. N. Abramson and S. Silverman, Eds., chapter 6, National Aeronautics and Space Administrator, Washington, DC, USA, 1966, NASA SP-106.
- [18] H. N. Abramson, W. H. Chu, and D. D. Kana, "Some studies of nonlinear lateral sloshing in rigid containers," *Journal of Applied Mechanics*, vol. 33, no. 4, 8 pages, 1966.
- [19] H. N. Abramson, "The dynamic behavior of liquids in moving containers," NASA SP-106, 1966.
- [20] Q. Li, X. Ma, and T. Wang, "Equivalent mechanical model for liquid sloshing during draining," *Acta Astronautica*, vol. 68, no. 1-2, pp. 91–100, 2011.
- [21] M. Utsumi, "A mechanical model for low-gravity sloshing in an axisymmetric tank," *Journal of Applied Mechanics, Transactions ASME*, vol. 71, no. 5, pp. 724–730, 2004.
- [22] J. S. Love and M. J. Tait, "Equivalent linearized mechanical model for tuned liquid dampers of arbitrary tank shape," *Journal of Fluids Engineering*, vol. 133, Article ID 061105, 7 pages, 2011.
- [23] R. Ranganathan, Y. Ying, and J. Miles, "Analysis of fluid slosh in partially filled tanks and their impact on the directional response of tank vehicles," *SAE Transactions*, vol. 102, pp. 502–505, 1993.
- [24] R. Ranganathan, S. Rakheja, and S. Sankar, "Directional response of a B-train vehicle combination carrying liquid cargo," *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, vol. 115, no. 1, pp. 133–139, 1993.

- [25] M. I. Salem, *Rollover stability of partially filled heavy-duty elliptical tankers using trammel pendulums to simulate fluid sloshing [Ph.D. thesis]*, West Virginia University, Department of Mechanical and Aerospace Engineering, 2000.
- [26] L. Dai, L. Xu, and B. Setiawan, "A new non-linear approach to analysing the dynamic behaviour of tank vehicles subjected to liquid sloshing," *Proceedings of the Institution of Mechanical Engineers, Part K*, vol. 219, no. 1, pp. 75–86, 2005.
- [27] <http://www.chinacar.com.cn/banguache/peixin.1182/XH9140G.140103.html>.
- [28] FLUNET 6. 3 User's Guide, FLUENT Inc., 2006.

Research Article

Train Stop Scheduling in a High-Speed Rail Network by Utilizing a Two-Stage Approach

Huiling Fu,^{1,2} Lei Nie,¹ Benjamin R. Sperry,² and Zhenhuan He¹

¹ School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

² Texas A&M Transportation Institute, College Station, TX 77843, USA

Correspondence should be addressed to Huiling Fu, fuhuiling00@163.com

Received 30 April 2012; Revised 28 July 2012; Accepted 1 October 2012

Academic Editor: Geert Wets

Copyright © 2012 Huiling Fu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Among the most commonly used methods of scheduling train stops are practical experience and various “one-step” optimal models. These methods face problems of direct transferability and computational complexity when considering a large-scale high-speed rail (HSR) network such as the one in China. This paper introduces a two-stage approach for train stop scheduling with a goal of efficiently organizing passenger traffic into a rational train stop pattern combination while retaining features of regularity, connectivity, and rapidity (RCR). Based on a three-level station classification definition, a mixed integer programming model and a train operating tactics descriptive model along with the computing algorithm are developed and presented for the two stages. A real-world numerical example is presented using the Chinese HSR network as the setting. The performance of the train stop schedule and the applicability of the proposed approach are evaluated from the perspective of maintaining RCR.

1. Introduction

A train stop schedule is one of the crucial parts in a train service plan (TSP). The train stop schedule specifies a set or subset of stations where individual trains will stop in order to satisfy passengers’ travel demand among stations in a rail network. From a system optimization point of view, passenger demand is heterogeneously distributed over space. Consequently, considering certain performance criteria, the presence of binary variables, and resource constraints such as restricted capacity, the train stop scheduling problem (TSSP) turns out to be NP-hard with uncontrollable computational complexity as scale of rail network or number of stations increases.

Practical experience has played an important role in understanding the TSSP and appreciating the complexity of the problem. To provide differential train services among

hierarchical stations and achieve connectivity over a large number of passenger origins and destinations (ODs) in a rail network, mature flexible combinations of typical train stop patterns exist in many established worldwide rail networks. Namely, express trains or trains with a few stops at stations of high classification (e.g., ICE train in Germany and IC train in The Netherlands), “skip-stop” or “zone-stop” trains stop at major stations to increase traveler’s alternatives (e.g., “Hikari” train in Japan), and “all-stop” pattern to service exchange passengers along train routes (e.g., AR train in The Netherlands and “Kodama” train in Japan), have been adopted in these countries for decades. Also adopted in these countries for many decades is the cyclic train operation mode, meaning that in every short time period (e.g., 1 hour), trains have the same operating frequencies, sequence, and speed, and trains with the same sequence in each period have identical stop pattern, departure and arrival time. By doing so, regularity in the train stop schedule is achieved, including high and fixed train frequencies between stations, short wait time, fast transfer, and flexible trip combinations for travelers.

In terms of applying system optimization theory, the challenging TSSP has attracted much attention in the literature, and there exist a number of studies showing different kinds of TSSP models or heuristic algorithms. Early studies applied diverse approaches to seek the best train stop schedules based on zoning [1], local versus express [2], and specific stop schemes [3]. Two optimization models were established with the objectives of covering more passenger demand with less train stops and saving more passengers’ travel time [4]; in the second model, a GA algorithm was introduced and tested on partial rail network in southern Germany. In the context of The Netherlands’ rail network, integer models combined with the multicommodity flow problem were developed to minimize the rail operator’s operating cost to generate multitype train stop schedule on the basis of the fixed train stop patterns in real world [5]. Additionally, in some literature sources, authors refer to approaches based on a prior given set of train stop patterns, which have been successfully applied, albeit for single rail line without branches. In the setting of the Taiwan HSR system, a multiobjective model was formulated to yield TSP with TSSP embedded in the model with the objective of minimizing the rail operator’s operating cost and the passenger’s travel time loss. The model was solved by a fuzzy mathematical programming approach [6]. A bilevel programming model was proposed and combined with network equilibrium analysis of passenger flow assignment on trains in a lower-level problem [7], and in a numerical study, there were only seven train stop patterns among five stations along the line could be selected to generate the final train stop schedule. Using a 46 km long, six-station transit line in the northeastern US as the background, a cost-efficient operation model that optimized all-stop, short-turn, and express transit services was developed [8], and a logit-based model was used to estimate the ridership for the seven candidate train stop patterns.

The TSSP has also attracted much attention of rail operators and researchers in China as the most extensive HSR system in the world is being built in that country. In practice, those mature train stop patterns and cyclic train operations employed in the above-mentioned countries are not directly transferable to this newly developed system because it is characterized by several large-scale, fully connected lines with many stations and uneven passenger demand distribution. In the literature studies, following closely with the Chinese HSR system’s specifics, the TSSP retains the challenge of combinatorial explosion yet now becomes more complex. This complexity is due in part because a great majority of passenger demands should be satisfied through train stops due to concentrated train ODs settings. It can be seen that numerical studies with various optimal models applications were still limited to simplified single line or downsizing network and largely dependent on adaptive

heuristic algorithms [9–11]. When handling this kind of complexity coupled with purposes of increasing connectivity, regularity without ignoring rapidity of some train services among major stations, the TSSP turns out to be extremely challenging which has not been well studied.

Motivated by above considerations, this study is different from “one-step” models commonly used in the literature. In this paper, we develop a computer-aided two-stage approach to efficiently solve the TSSP from RCR perspectives using the Chinese HSR network as a case study. As train ODs settings are concentrated at a few main stations and small percentages of major stations comprise a high proportion of passenger traffic, the station classification concept is followed. The first-stage of the approach utilizes mixed integer programming to organize passengers among scale-reduced higher classification stations; for the remaining passengers associated with lower classification stations, by developing a train operating tactics descriptive model, the second-stage of the approach computes additional train stops within the overall schedule frame determined in the first-stage. The remainder of this paper is organized as follows. Section 2 describes the formulation of the two-stage approach and related solution algorithm. Section 3 provides a numerical example from the Chinese HSR network. Section 4 concludes the study and brings forward the future work.

2. Model Formulation

2.1. Notations and Assumptions

The parameters and variables associated with the model development are summarized in Table 1. The basic assumptions are as follows.

- (a) Stations are divided into three classifications using a K -means cluster analysis considering issues such as the technical conditions of train-set maintenance, collecting and distributing of passenger traffic, political or economic factors.
- (b) Train ODs are pre-given determined by ridership and technical regulations, for example, between stations which can service as terminals (generally are stations of the first two classifications) and route length being within train-set maintenance kilometers.
- (c) Passengers among these ODs with ultralong travel distance are processed to be split into separate ODs which are disconnected at major transfer hubs such that a single passenger is not required to transfer more than two times during a single trip.
- (d) Optimizing train operating frequencies is out of scope of this study. For modeling needs, frequencies are estimated for trains among given train ODs according to potential attracted passenger traffic.
- (e) Passengers’ travel benefits in terms of RCR are mostly converted to train operating tactics during the two-stage train stop scheduling process, thus a seat reservation system is necessarily required to coordinate with the models practical application.

2.2. The First-Stage Modeling of TSSP for Higher-Classification Trains

Higher-classification trains are defined as the trains running and dwelling only among stations of the highest two classifications. Frequent train stops results in negative impacts

Table 1: Notation of model parameters and variables.

Parameter/variable	Description
$L(\text{gt})$	Set of trains among given train ODs
$L(\text{ht})$	Set of stop trains of higher classification
$L(\text{lt})$	Set of stop trains of lower classification
ℓ_k	Trains in set L indexed by k
$f(\ell_k)$	Estimated frequency of train $\ell_k (\in L(\text{gt}))$
$h(\ell_k)$	Route length of train ℓ_k (km)
$V(\text{fc})$	Set of the first classification stations in HSR network
$V(\text{sc})$	Set of the second classification stations in HSR network
$V(\text{tc})$	Set of the third classification stations in HSR network
v_i or v_j	Stations in set V indexed by i or j
E	Set of tracks in HSR network
e_m	Tracks in set E indexed by m , $e_m \in E$
$v(\ell_k, e_m)$	Travel speed of train ℓ_k on track e_m (km/h)
$L(\text{gt}, e_m)$	Set of trains in set $L(\text{gt})$ with their routes covering track e_m
D	Set of passenger ODs
$d(v_i, v_j)$	Passenger demand between station v_i and v_j that $D = \{(v_i, v_j) \in V \times V \mid d(v_i, v_j) > 0\}$
$\kappa(\ell_k)$	Seating capacity of train ℓ_k
$\theta(\ell_k)$	Loading coefficient of train ℓ_k
$N(\ell_k)$	The maximum times of stops can be added on train $\ell_k (\in L(\text{gt}))$ for trains in set $L(\text{ht})$
$\eta(v_i, \ell_k)$	Count parameter of whether stop being added at station v_i on train $\ell_k (\in L(\text{gt}))$
$L(\text{gt}, (v_i, v_j))$	Set of possible trains on which stop(s) can be added for passenger OD (v_i, v_j) , and $L(\text{gt}, (v_i, v_j)) = \{\ell_k \in L(\text{gt}) \mid (v_i, v_j) \in \ell_k\}$
$D'(\ell_k)$	Set of possible passenger ODs for which the stop(s) can be added on train $\ell_k (\in L(\text{gt}))$, and $D'(\ell_k) = \{(v_i, v_j) \in D \mid \ell_k \in (v_i, v_j)\}$
$x(\ell_k, (v_i, v_j))$	Binary variable, it is 1 only if for passenger OD (v_i, v_j) , stop(s) is(are) added on train $\ell_k (\in L(\text{gt}))$, else it equals 0
$y(\ell_k, (v_i, v_j))$	The flow of passenger OD (v_i, v_j) assigned on train $\ell_k (\in L(\text{gt}))$ with added stop(s) on it, $y(\ell_k, (v_i, v_j)) \geq 0$ and integer
$r(\ell_k)$	Proportion of through passengers out of total traffic attracted by train $\ell_k (\in L(\text{gt}))$
$\omega(\ell_k)$	Weights of influencing factor of adding stop(s) on train $\ell_k (\in L(\text{gt}))$
$\omega'(h(\ell_k), v(\ell_k, e_m))$, $\omega''(r(\ell_k))$	Weights functions of influencing factors of adding stop(s) on train $\ell_k (\in L(\text{gt}))$

of reducing train's rapidity, increasing passengers' total travel time and train operating cost. Hence, stops are not given priority to be added on trains in set $L(\text{gt})$ with properties as follows:

- (a) being operated between stations of the first classification,
- (b) a long travel time (depending on routes length and technical speed standards of running sections),

- (c) having high proportion of their dedicated origin and destination passengers (through passengers) out of total attracted traffic.

As a consequence of (a), high-quality train services between stations of the first classification are ensured, while (b) and (c) allow competitive travel times on long-distance routes to the benefit of a large number of through passengers using such routes.

By putting weights of the three influencing factors on trains in set $L(\text{gt})$, the negative impacts of adding stops on trains have to be minimized, giving rise to the objective of

$$\min \sum_{\substack{(v_i, v_j) \in D: \\ v_i, v_j \in V(\text{fc}), V(\text{sc})}} \sum_{\ell_k \in L(\text{gt}, (v_i, v_j))} (\omega(\ell_k) + \omega'(h(\ell_k), \nu(\ell_k, e_m)) + \omega''(r(\ell_k))) \cdot x(\ell_k, (v_i, v_j)), \quad (2.1)$$

where $\omega(\ell_k)$, $\omega'(h(\ell_k), \nu(\ell_k, e_m))$, and $\omega''(r(\ell_k))$ stand for weights of the three influencing factors, respectively. These weights are calibrated using a scoring approach, specifically, $\omega(\ell_k)$ is 0 if the first factor is not applicable to train ℓ_k , $\omega'(h(\ell_k), \nu(\ell_k, e_m))$, and $\omega''(r(\ell_k))$ are incremental with the increase of train travel time and proportion of through passengers for different trains.

Additionally, the maximum stop times on train ℓ_k need to be limited, thus,

$$\sum_{v_i \in V(\text{fc}), V(\text{sc})} \eta(v_i, \ell_k) \cdot x(\ell_k, (v_i, v_j)) \leq N(\ell_k) \quad \forall \ell_k \in L(\text{gt}). \quad (2.2)$$

The count parameter $\eta(v_i, \ell_k)$ equals 0 if a stop will not be added at station v_i and equals 1 if station v_i is to be added as a stop for more than one passenger OD. The condition of $v_i, v_j \in V(\text{fc}), V(\text{sc})$ in (2.1) and (2.2) holds for all the following constraints.

To ensure the total train stop frequencies at a given station are adequate to meet the passenger demand requirements, the supply-demand constraint is formulated as

$$\sum_{\ell_k \in L(\text{gt}, (v_i, v_j))} x(\ell_k, (v_i, v_j)) \cdot \kappa(\ell_k) \cdot f(\ell_k) \cdot \theta(\ell_k) \geq d(v_i, v_j) \quad \forall (v_i, v_j) \in D. \quad (2.3)$$

Furthermore, the flow of different passenger ODs assigned on a given train ℓ_k should not exceed that train's seating capacity

$$\sum_{\substack{(v_i, v_j) \in D'(\ell_k): \\ (v_i, v_j, \ell_k) \geq e_m}} y(\ell_k, (v_i, v_j)) \leq \kappa(\ell_k) \cdot f(\ell_k) \cdot \theta(\ell_k) \quad \forall e_m \in E, \ell_k \in L(\text{gt}, e_m). \quad (2.4)$$

in which $(v_i, v_j, \ell_k) \geq e_m$ means that only passenger OD(s) passing through track e_m is (are) taken into account.

In the assignment process, passenger flow conservation is given by

$$\sum_{\ell_k \in L(\text{gt}, (v_i, v_j))} y(\ell_k, (v_i, v_j)) = d(v_i, v_j) \quad \forall (v_i, v_j) \in D. \quad (2.5)$$

Following constraint (2.6) denotes that if stop(s) is (are) not added on train ℓ_k for passenger OD (v_i, v_j) , its flow assigned on train ℓ_k equals 0, \widehat{M} is a very large positive number

$$y(\ell_k, (v_i, v_j)) \leq \widehat{M} \cdot x(\ell_k, (v_i, v_j)) \quad \forall \ell_k \in L(\text{gt}), (v_i, v_j) \in D. \quad (2.6)$$

Aiming at achieving connectivity of train services, a binary variable of whether stop(s) is (are) added on train ℓ_k for passenger OD (v_i, v_j) rather than a stop being added at station v_i on train ℓ_k or not (as is typically included in most studies) is designed here. The approach of defining on which train(s) travelers of each passenger OD are assigned is still followed in the second-stage modeling of TSSP. Two types of decision variables are restricted as follows:

$$\begin{aligned} x(\ell_k, (v_i, v_j)) &\in \{0, 1\} \quad \forall \ell_k \in L(\text{gt}), (v_i, v_j) \in D, \\ y(\ell_k, (v_i, v_j)) &\geq 0 \quad \forall \ell_k \in L(\text{gt}), (v_i, v_j) \in D. \end{aligned} \quad (2.7)$$

Additionally, $x(\ell_k, (v_i, v_j))$ equals 1 only if for passenger OD (v_i, v_j) , stop(s) is (are) added on train $\ell_k (\in L(\text{gt}))$ in v_i (if v_j is origin or destination station of train ℓ_k) or v_j (if v_i is origin or destination station of train ℓ_k), or in both (if none of v_i and v_j is origin or destination station of train ℓ_k).

2.3. The Second-Stage Modeling of TSSP for Lower-Classification Trains

Lower-classification trains are defined as the trains dwelling at least once at stations of the third (i.e., lowest) classification. The train stop schedule generated in the first-stage (Section 2.2) gives the overall scheme; based on this initial scheme, additional train stops are further scheduled for the remaining passengers associated with the third classification stations. Before starting the descriptive model and computing algorithm development, tactics of sequencing trains in set $L(\text{ht})$ involved in modeling and algorithm implementing process are interpreted below.

- (a) Trains are categorized as in-line and cross-line trains corresponding to running within a single line and cross at least two different lines, respectively. In-line trains are given priority over cross-line trains to ensure a match between train route length and passenger travel distance.
- (b) From short trains to long trains considering equilibrium of train timetabling.
- (c) From trains with more stops to trains with fewer stops, leading to setting stops intensively on fewer trains aiming at increasing the proportion of trains with higher travel speed.

It is noted that the proposed algorithm is flexible such that it is still possible for rail operator to design other tactics and adjust preferred train sequence as passenger demand warrants. The descriptive model and computing algorithm are outlined in Algorithm 1 as follows.

In Algorithm 1, essentials including a train stop pattern enumeration technique, a train stop pattern decision-making criteria, and its corresponding passenger flow assignment procedure presented from Step 7 to Step 15 reveal the main objective of the descriptive

- (1) - Set sequence of the input trains in set $L(ht)$ (including selected non-stop direct trains (e.g., short-distance trains with running sections covering high passenger density) in set $L(gt)$)
- (2) **repeat**
- (3) **for** train ℓ_k ($k = 1$ to K ($K \leq |L(ht)|$)) **do**
- (4) - Enumerate all passenger OD pairs that can be covered
- (5) - Calculate the maximum remaining stop times that can be added on train ℓ_k in set $L(ht)$, notated as $N'(\ell_k)$
- (6) **Repeat**
- (7) **for** $N'(\ell_k)$ ($N'(\ell_k) \neq 0$) **do**
- (8) - Enumerate all station group(s) with combined $N'(\ell_k)$ station(s) of the third classification ($\in V(tc)$), notated as $\bar{V}(sg, n)$, where n is index symbol
- (9) - Calculate total passenger traffic of all combined passenger OD pairs among station(s) in $\bar{V}(sg, n)$ together with existing stop(s) and OD on train ℓ_k , notated as $TP(\bar{V}(sg, n)) = \sum_{(v_i, v_j)} d(v_i, v_j)$, where both v_i and v_j or at least one of them belong to $\bar{V}(sg, n)$
- (10) - Select the group $\bar{V}(sg, n)$ with the maximum $TP(\bar{V}(sg, n))$
- (11) **if** train ℓ_k has a *cost-efficient frequency** to be operated with assigned $TP(\bar{V}(sg, n))$ on it, **do**
- (12) - Add station(s) in the selected group on train ℓ_k as additional stop(s), and update passenger flow, go to Step 7
- (13) **else do**
- (14) - Update $N'(\ell_k) = N'(\ell_k) - 1$, go to Step 7
- (15) **end for**
- (16) **until** $N'(\ell_k) = 0$
- (17) **end for**
- (18) **until** assigning the remaining passengers on trains is completed, then set of $L(lt)$ is obtained

*From the perspective of the rail operator, a cost-efficient train operating frequency is defined as at least one train per day for a given train OD with a rational loading factor (e.g., ≥ 0.70).

Algorithm 1: Descriptive model/algorithm for computing train stop schedule for lower-classification trains.

model and computing algorithm: scheduling additional train stops to satisfy the remaining passenger demand with a minimum number of different train stop patterns.

3. Numerical Example

This section demonstrates the applicability of the developed models and algorithm using the Chinese HSR network (Figure 1) as a case study. The Chinese HSR network (planned for 2015, excluding intercity lines) consists of 21 lines with total mileage over 10,000 km. The center of the network is the Beijing-Shanghai (Jinghu) HSR line which is, to date, the longest HSR line in the network (1,318 km). The Jinghu HSR is built with the highest standards (the maximum speed is 350 km/h) and plays a significant role in entire network. In a selected study subnetwork (noted as dual lines) centered by the Jinghu HSR as shown in Figure 1, 38 train ODs associated with the Jinghu HSR among 27 stations of the first two classifications are

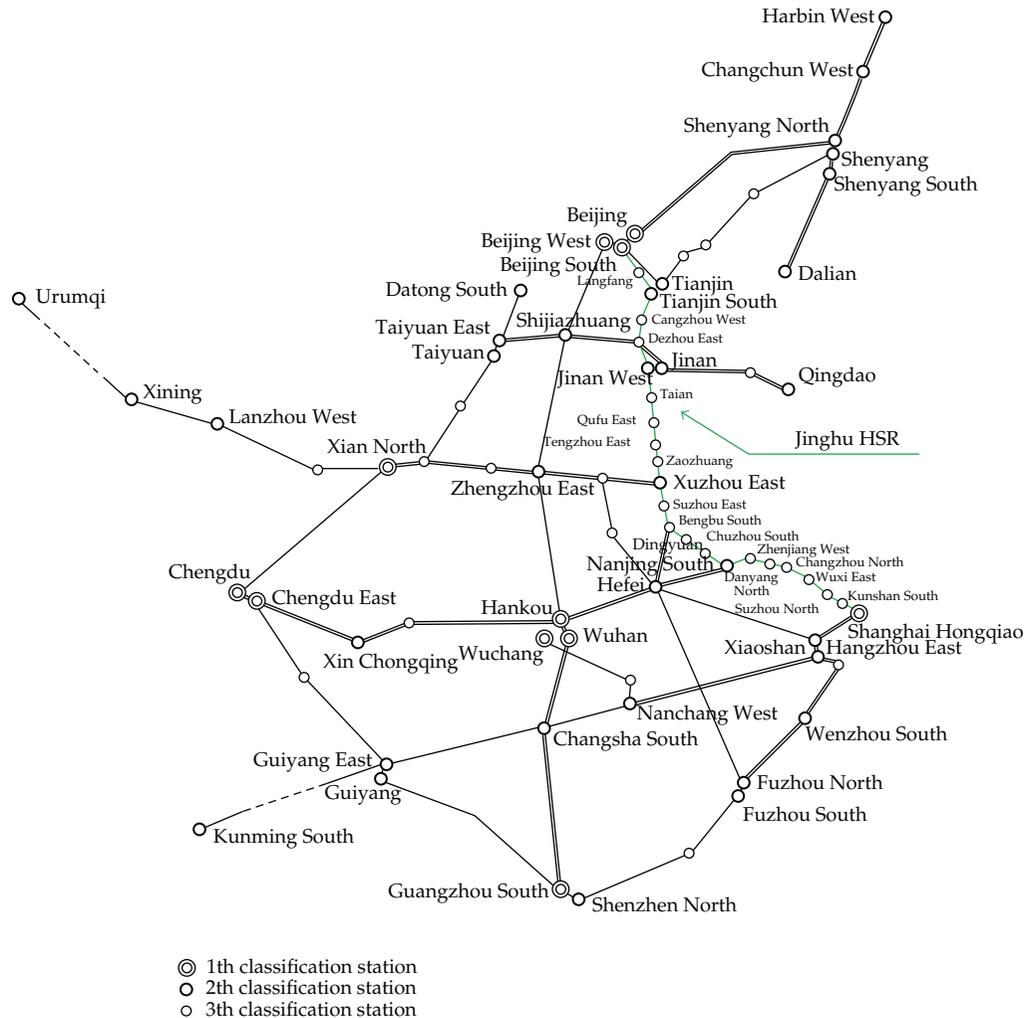


Figure 1: The Chinese HSR network with study Jinghu HSR and its associated network noted.

pregiven. A total of 481 passenger ODs relating to the Jinghu HSR having known forecasted daily traffic are within the scope of this numerical example.

The trains seating capacity and loading coefficients are identically set equal to 1,060 seats/train-set and 1.0 (at maximum), respectively. Within the running section on the Jinghu HSR, the maximum stop times of higher-classification in-line trains is 4 and 7 for lower-classification in-line trains; cross-line trains' maximum stop times are restricted as 6 and 9 for higher- and lower-classification trains, respectively. Due to data size, parameters below are not individually listed in detail: (a) passenger traffic of each OD, a total of 536,914 passengers (one direction, including passengers organized on nonstop direct trains) are considered in the scheduling process, for trains in set $L(gt)$, as are their (b) routes length, (c) travel time, (d) estimated frequencies, and (e) integrated weights of three influencing factors valued between 0.37 to 1.0.

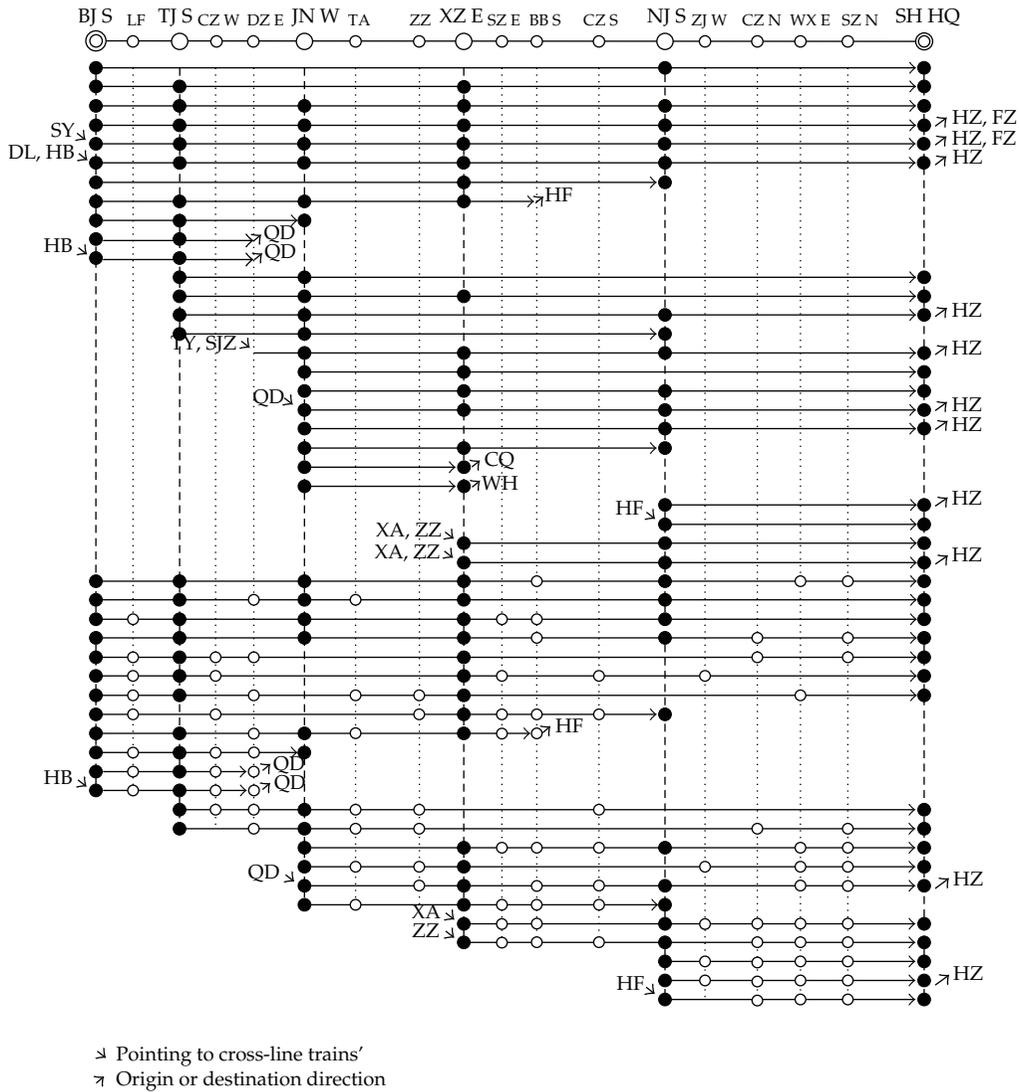


Figure 2: Train stop schedule estimated from two-stage approach.

The two-stage approach outlined here reduces the problem-solving scale, resulting in a total of 15 binary variables representing in-line passenger ODs and 118 binary variables for the other passenger ODs associated with the Jinghu HSR. Variables $y(\ell_k, (v_i, v_j))$, typically considered to be integer, are relaxed as real. Thus, the first-stage TSSP model is efficiently solved by a Branch-and-Bound algorithm coded in the LINGO 10.0 solver, which is embedded in a computer system written by C# programming language for computing the second-stage TSSP model. The computations are performed on an Intel i5 2.4GHZ with 2 GB RAM in the environment of Microsoft Windows XP, with a processing time less than 30 minutes. Figure 2 shows the two-stage TSSP model result for the train stop schedule on the Jinghu HSR.

It can be seen in Figure 2 that the two-stage approach results in a flexible combination of train stop patterns. An “all-stop” train stop pattern is not recommended for the long-distance sections for example, Beijing South-Shanghai Hongqiao. For the sake of servicing exchange passengers which are not related to the Jinghu HSR at all and with unknown detailed traffic (e.g., passengers between Xian North and Zhengzhou East, Harbin West, and Changchun West), in this numerical example, it was assumed that cross-line trains adopted an all-stop pattern at stations of the first two classifications on their running sections outside the Jinghu HSR. For example, the train “Xian North-Shanghai Hongqiao” stops at “Zhengzhou East” station on the route.

Since train frequencies optimization is not explicitly involved, regularity is not entirely the result of operating trains cyclically, but from restricting train stop times on different classification trains with different properties. Higher-classification trains’ average stop times is approximately 3, and the three trains with the highest integrated weights of influencing factors (“Beijing South-Shanghai Hongqiao,” “Beijing South-Wuhan,” and “Xian North-Shanghai Hongqiao”) stop an average of two times. Lower-classification trains stop an average of six times with a total of 23 stop patterns obtained. Stops are mostly allocated on in-line trains, and cross-line trains with short-medium travel distance constitute a low proportion totaling 8 stop patterns.

In the current TSP on the Jinghu HSR, there are approximately 16 percent of all in-line passenger ODs that have no train service or are serviced only by a few “all-stop, medium-speed” trains, rather than high-speed trains. Comparatively, connectivity is sufficiently ensured, as previously mentioned, by tracking on which trains passengers are assigned. Due to the “*cost-efficient frequency*” restriction (see Step 11 in Algorithm 1), only two in-line passenger ODs have no direct train service within this numerical example: the “Cangzhou West-Wuxi East” OD and the “Dezhou East-Zhenjiang West” OD. Moreover, greater connectivity is achieved by using only 23 lower-classification train stop patterns, a reduction of approximately 39 percent compared to the current TSP on the Jinghu HSR. On a more practical level, the rail operator benefits by significantly improving the homogeneity of a train timetable.

In terms of rapidity, considering in-line trains for statistics (at present a few cross-line trains only in two directions are in operation on the Jinghu HSR), the proportion of trains with 6 or 7 stops (the most commonly adopted train stop pattern) in the current TSP is about 66 percent. Comparatively, in this numerical example, the proportion of trains with 6 or 7 stops is reduced to approximately 46 percent of trains (36 percent of trains considering estimated frequencies). Furthermore, the number of trains making only 1 or 2 stops is nearly 38 percent in this numerical example (50 percent by considering estimated frequencies), in contrast to approximately 25 percent in the current TSP.

4. Conclusions

This paper presented a two-stage approach for solving the TSSP in a large-scale HSR network, such as the Chinese HSR network. A mixed integer programming model, a train operating tactics descriptive model, and algorithm were developed for the two stages, and a real-world numerical analysis demonstrated that the generated train stop schedule improved the combination of train stop patterns, while also improving the regularity, connectivity, and rapidity. In addition to ensuring optimization, to speed up the entire solving process, future research should consider using intelligent algorithms for the first-stage TSSP model. Another

area of future work comprises integrating the TSSP into the optimization of train frequencies by expanding the two-stage approach so as to capture a more comprehensive efficiency evaluation, incorporating indicators such as the capacity utilization ratio, rail operator's operating cost, and passenger's travel frequencies.

Acknowledgments

This work was supported in part by the project of Ministry of Railways (jointed support of the Ministry of Science and Technology) of China under Grant no. 2009BAG12A10 and the Fundamental Research Funds for the Central Universities (Beijing Jiaotong University) under Grant no. 11JBM323. The authors are grateful to the referees for their valuable comments and suggestions.

References

- [1] N. S. A. Ghoneim and S. C. Wirasinghe, "Optimum zone structure during peak periods for existing urban rail lines," *Transportation Research B*, vol. 20, no. 1, pp. 7–18, 1986.
- [2] G. L. Nemhauser, "Scheduling local and express trains," *Transportation Science*, vol. 3, no. 2, pp. 164–175, 1969.
- [3] S. Sone, "Novel train stopping patterns for high-frequency, high-speed train scheduling," in *Computers in Railways III, Technology*, T. K. S. Murthy, J. Allan, R. J. Hill, G. Sciotto, and S. Sone, Eds., vol. 2, pp. 107–118, Computational Mechanics Publications, 1992.
- [4] H. W. Hamacher, A. Liebers, A. Schöbel, D. Wagner, and F. Wagner, "Locating new stops in a railway network," *Electronic Notes in Theoretical Computer Science*, vol. 50, no. 1, pp. 13–23, 2001.
- [5] J.-W. Goossens, S. van Hoesel, and L. Kroon, "On solving multi-type railway line planning problems," *European Journal of Operational Research*, vol. 168, no. 2, pp. 403–424, 2006.
- [6] Y. H. Chang, C. H. Yeh, and C. C. Shen, "A multiobjective model for passenger train services planning: application to Taiwan's high-speed rail line," *Transportation Research B*, vol. 34, no. 2, pp. 91–106, 2000.
- [7] C. Lee and W. Hsieh, "A demand oriented service planning process," in *Proceedings of the World Congress on Railway Research*, Koln, Germany, 2001.
- [8] Y. Y. Ulusoy, S. Chien, and C. H. Wei, "Optimal all-stop, short-turn, and express transit services under heterogeneous demand," *Transportation Research Record*, no. 2197, pp. 8–18, 2010.
- [9] H. Fu and L. Nie, "Economy and efficiency oriented passenger flow organization: train stop scheduling for high-speed railways," *Journal of Beijing Institute of Technology*, vol. 20, pp. 79–83, 2011.
- [10] Y. Zhang, M. Ren, and W. Du, "Optimization of high speed train operation," *Journal of Southwest Jiaotong University*, vol. 33, no. 4, pp. 400–404, 1998 (Chinese).
- [11] L. Deng, F. Shi, and W. Zhou, "Stop schedule plan optimization for passenger train," *China Railway Science*, vol. 30, no. 4, pp. 102–107, 2009 (Chinese).

Research Article

Analysis of the Effect of the Length of Stop-Spacing on the Transport Efficiency of a Typically Formed Conventional Locomotive Hauled Passenger Train in China

Xuesong Feng,^{1,2} Keqi Wu,¹ Haidong Liu,² Qipeng Sun,³ and Hanxiao Zhang¹

¹ *Integrated Transport Research Center of China, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing 100044, China*

² *MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing 100044, China*

³ *School of Economics and Management, Chang'an University, Nan Er Huan Zhong Duan, Xi'an 710064, China*

Correspondence should be addressed to Xuesong Feng, charles.x.feng@gmail.com

Received 7 May 2012; Accepted 26 September 2012

Academic Editor: Huimin Niu

Copyright © 2012 Xuesong Feng 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.

By utilizing a computer-aided simulation approach, this research analyzes the detailed effect of the length of stop-spacing on the transport time per passenger-kilometer of a typically formed Chinese intercity passenger train hauled by, respectively, representative types of the locomotives utilized in China for different target speeds. It is empirically confirmed that a stop-spacing longer than approximately 20.00 km has little influence upon the transport efficiency of the train. In contrast, shortening the stop-spacing below about 20.00 km increases the transport time per passenger-kilometer of the train evidently, in particular for a target speed much higher than some 40.00 km/h. Therefore, the target speed of such a train providing transport service on a railway line whose transport capacity has not been much consumed ought to be changeable according to the length of each stop-spacing rather than consistent for the whole trip.

1. Introduction

Although the high-speed railway trains in China are playing more and more important roles in the work of intercity passenger transports, the Conventional Locomotive Hauled Passenger (CLHP) trains are still popularly preferred today by most of Chinese people for their intercity trips. According to the statistic data released by the Ministry of Railways of the People's Republic of China, around 80.00% of more than 1.67 billion [1] passenger trips

undertaken in China by railway trains in 2010 have been completed by the CLHP trains. Besides the advantages of punctuality, low fares, and so forth, the relative rapidity of the CLHP trains is also essential to have them favored for intercity trips in China.

It is commonly believed that a high target speed which a train strives to reach in its transport process between neighboring stops makes the train run fast and therefore transport passengers or freight efficiently. However, many other factors including the ramps, bends, and so forth of the rail line [2–5], the traction performance (such as the full traction power) of the train [6–8], the weights of different railway cars forming the train [9, 10], and the length of the stop-spacing [2, 11–14] all have strong influence upon the average speed of the train especially with a high target speed and accordingly its transport time per passenger-kilometer, that is, the transport efficiency. Many researchers and practitioners have been making continuous effort to interpret the relationship between the target speed of a train and its transport efficiency in consideration of various elements [15–17]. Nevertheless, the exact impact of the transport distance between neighboring stops on the time consumed per passenger-kilometer of a transport mission completed by a CLHP train for its different target speeds in view of other factors, for example, the traction performance of the applied locomotive, and so forth, still needs to be further explored in detail.

Based on the computer-aided simulations of the passenger transports by a typically formed Chinese CLHP train with different target speeds on a railway line with various distances between neighboring stops, this study attempts to clarify the quantificational effect of the length of stop-spacing on the passenger transport efficiency in an accurate manner. Additional regards are also paid to traction performances of various types of the applied locomotives of the studied CLHP train. Due to the inadequate support of the data of the alignments of actual railway lines, a hypothetical railway line which is straight and smooth is utilized for this research and the transport efficiencies of the studied CLHP train with a certain target speed and the same utilization ratio of its passenger carrying capacity for different lengths of all the stop-spacings of this railway line are analyzed in comparison. Moreover, because of the little influence of the weight of passengers together with their baggage upon the transport time of a train and usually high travel demands of the passengers of the CLHP trains in China, the utilization ratio of the seats and berths of the studied CLHP train for each stop-spacing is assumed to be the same here as 100.00%.

The latter parts of this paper are organized as follows. The studied CLHP train hauled, respectively, by different types of locomotives and the computer-aided simulation approach applied to compute the Passenger Transport Time (PTT) of the train are explained in Section 2. Next, Section 3 analyzes the detailed changes of the PTTs per 10,000 passenger-kilometers (p-km) of the train with the increase of its target speed for different transport distances between neighboring stops. Finally, Section 4 draws conclusions, makes some suggestions for the passenger transport work of the CLHP trains in China, and points out some future research issues.

2. The Studied CLHP Train and the Simulation Approach

As shown in Figure 1, the Locomotive (Lo) of the studied typically formed Chinese CLHP train in this research hauls 18 railway cars which are in sequence 1 car equipped with the Power (P) unit, 5 cars equipped with passenger Seats (S), 1 car providing Dining (D) services, 1 car equipped with Soft Berths (SB), 9 cars equipped with common Berths (B) and 1 car carrying checked Luggage (L) of the passengers. The type of the cars composing the studied CLHP train is 25 K which is one of the major types of the cars of the Chinese CLHP trains now.



Figure 1: Formation of the studied CLHP train.

The weights of the cars of 25 K for various purposes are, respectively, 60.50 tons (P), 48.80 tons (S), 48.00 tons (D), 66.00 tons (SB), 46.50 tons (B), and 42.40 tons (L) [18]. Therefore, the total weight of the 18 cars forming the studied CLHP train is 879.40 tons. A car of 25 K has 118 passenger seats (S), 36 soft berths (SB), or 66 common berths (B). As a result, the number of the total passenger seats and berths of this CLHP train is 1,220. Because the average weight of a railway passenger together with his/her hand baggage is 80.00 kg [19], the weight of all the passengers together with their baggage in the studied CLHP train is 97.60 tons if all the passenger seats and berths of the train are 100% utilized in the whole process of its passenger transport. This train is, respectively, hauled by two major types of the locomotives for the railway passenger transport work in China, that is, the SS3 and the SS8 which are all electric locomotives. The weights of the SS3 and the SS8 are correspondingly 138.00 tons and 88.00 tons [8]. Their designed top speeds are 100.00 km/h and 170.00 km/h, respectively [8].

By referring to the work of Mao et al. [7], the computer-aided simulation approach presented in Figure 2 is applied in this research to calculate the PTT of the studied CLHP train. The whole transport process of the train from one stop to the next is simulated for one calculation interval after another. The lengths of all the calculation intervals are set to be equal to 1.00 second (s) in this work. Only the traction force and operating condition (i.e., coasting, being in traction or braking) of the train are considered as unchanged values in one calculation interval. The train at a station is started up with its full traction power towards the target speed. With the first achievement of the target speed by the continuous acceleration of the train from its startup, the train commences to be coasted till the difference between its speed and the target speed reaches a predefined constant value and thereafter accelerated in its full traction power to the target speed alternately. In order to ensure the stop of the train in safety in the next station, the train begins to check whether brakes are necessary or not in a calculation interval when it arrives at a rail site where there is a certain distance away from the next stop. This is determined according to the speed (v_1) of the train and the permitted speed (v_2) which is decided based on the braking performance of the train and the transport distance from the location of the train at the beginning of this calculation interval to the next stop. If $v_1 \geq v_2$, the train brakes to decrease its speed as soon as possible to a small value which is able to absolutely ensure the safety of its stop in the next station; if $v_1 < v_2$, the train coasts. Such a decision is made for each latter calculation interval till the train stops in security in the next station according to the v_1 of the train in each of the latter calculation intervals and the v_2 which is determined based on not only the location of the train at the beginning of each latter calculation interval but also the braking performance of the train.

The traction force of a train utilizing a certain ratio, that is, $r\%$, of its full traction power in a calculation interval is determined by both the speed and the operating condition of this train, as explained by (2.1). When the train is coasting or braking, its traction force is 0 N:

$$f_k^r = \begin{cases} \frac{P_k^r}{v_{k-1}^r}, & (v - v_{k-1}^{\text{pr}}) > C^{\text{tm}} \text{ or } (v_k^{\text{ul}} - v_{k-1}^{\text{pr}}) > C^{\text{ul}} \\ 0, & (v - v_{k-1}^{\text{pr}}) \leq C^{\text{tm}} \text{ or } (v_k^{\text{ul}} - v_{k-1}^{\text{pr}}) \leq C^{\text{ul}}, \end{cases} \quad (2.1)$$

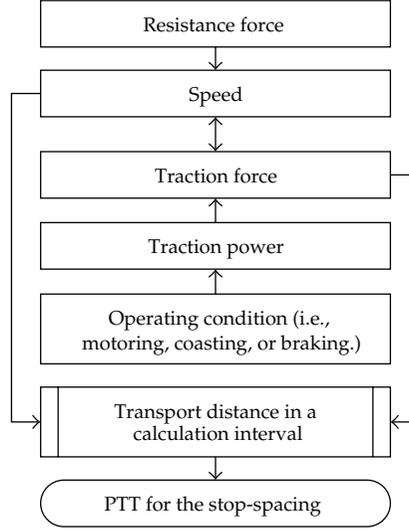


Figure 2: Simulation approach to calculate the PTT.

where f_k^r is the traction force of the train utilizing $r\%$ of its full traction power in the k th calculation interval, unit: N, P_k^r is the traction power of the train utilizing $r\%$ of its full traction power in the k th calculation interval, unit: W, v_{k-1}^{pr} is the speed of the train utilizing $pr\%$ of its full traction power at the end of the $(k-1)$ th calculation interval, unit: m/s, v is the target speed of the train, unit: m/s, v_k^{ul} is the upper limit speed in the k th calculation interval, which is equal to v^{tm} when there is no requirement by the rail line, unit: m/s, C^{tm} is the permitted maximum difference between speed of the train and target speed, unit: m/s, and C^{ul} is the permitted maximum difference between speed of the train and upper limit speed which is equal to C^{tm} when there is no requirement by the rail line, unit: m/s.

As illuminated by (2.1) and (2.2), the speed of the train in a calculation interval is decided by the speed of the train at the end of the previous calculation interval, the traction force for the utilized proportion of its full traction power in this calculation interval, the target speed, the upper limit speed required by the rail line in this calculation interval, the mass of the train, and the resistance force from air, rail line, and so forth in this calculation interval:

$$v_k^r = v_{k-1}^{pr} + \frac{f_k^r - f_k^L}{M} \times \Delta t, \quad (2.2)$$

where v_k^r is the speed of the train utilizing $r\%$ of its full traction power at the end of the k th calculation interval, unit: m/s, f_k^L is the resistance force in the k th calculation interval, unit: N, which is measured by (2.3), M is the mass of the train, unit: Kg, and Δt is the equivalent length of the calculation intervals, that is, 1.00 s, in this work:

$$f_k^L = \alpha_0 + \alpha_1 \times \left(v_{k-1}^{pr}\right) + \alpha_2 \times \left(v_{k-1}^{pr}\right)^2 + f_k^S, \quad (2.3)$$

where α_0 , α_1 , and α_2 are the coefficients determined by the body streamline design of the locomotive, the friction between the wheels and the rail, and so on, and f_k^S is the resistance

from the ramps, bends, and so forth, of the rail line in the k th calculation interval, unit: N. Because of the afore-explained assumption of a hypothetically straight and smooth rail line, the special resistance force is 0 N in this study.

Different types of locomotives have respective traction performances to overcome the resistance force in their traction processes on the same rail line for the same transport work and the same target speed at the expense of different transport time. The transport distance of the train in a calculation interval is interpreted as (2.4). The PTT of the train between two neighboring stops is computed by the summation of all the calculation intervals which make the train complete successive transport distances constituting this stop-spacing:

$$d_k^r = \frac{v_{k-1}^{\text{pr}} + v_k^r}{2} \times \Delta t, \quad (2.4)$$

where d_k^r means the transport distance of the train utilizing $r\%$ of its full traction power in the k th calculation interval, unit: m.

3. Analysis of Transport Efficiency

The PTT per 10,000 p-km of a train with the target speed of v between two stops is defined by (3.1) to evaluate its transport efficiency:

$$t_{ij}^v = \frac{T_{ij}^v}{\left(\sum_{q=1}^{\text{tn}} (P_{ij}^{v,q} \times R_{ij}^{v,q}) \right) \times D_{ij}^v}, \quad (3.1)$$

where t_{ij}^v is the PTT per 10,000 p-km of the train with the target speed of v from station i to station j , unit: hour (h)/10,000 p-km, T_{ij}^v is the PTT of the train with the target speed of v from station i to station j , unit: h, tn is the total number of the railway cars forming the studied train, $P_{ij}^{v,q}$ is the number of the passenger seats or berths of the q th railway car forming the train with the target speed of v from station i to station j , $R_{ij}^{v,q}$ is the utilization ratio of the passenger seats or berths of the q th railway car forming the train with the target speed of v from station i to station j , unit: %, and D_{ij}^v is the transport distance of the train with the target speed of v from station i to station j , unit: 10,000 km.

The transport distance (unit: 10,000 km) from the n th stop ($S(n)$) to the $(n+1)$ th stop ($S(n+1)$) ($n = 1, 2, \dots, 20$) of the hypothetically straight and smooth rail line in this research is interpreted by (3.2):

$$D_{S(n),S(n+1)} = 5.00 \times 10^{-4} \times n. \quad (3.2)$$

The changes of the PTTs per 10,000 p-km of the passenger transports of the CLHP train hauled, respectively, by the SS3 and the SS8 between different stops along this hypothetical rail line with the increase of the target speed of the train are presented in Figures 3 and 4 correspondingly. It is first observed in Figure 3 that when the target speed of the train is lower than about 40.00 km/h, the PTTs per 10,000 p-km of the train hauled by the SS3 for different transport distances between neighboring stops decrease relatively fast with the increase of the target speed and the length of stop-spacing has minor influence upon the

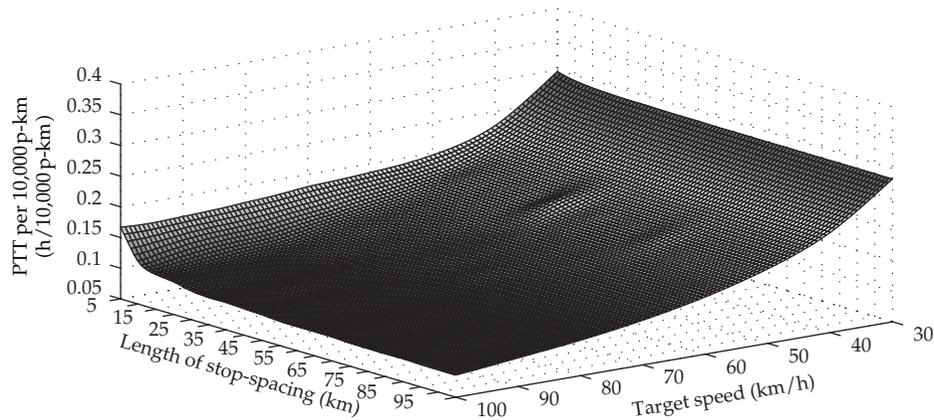


Figure 3: PTTs of the CLHP train hauled by the SS3.

PTTs per 10,000 p-km. If the target speed becomes higher than approximately 40.00 km/h, the decreases of the PTTs per 10,000 p-km with the improvement of the target speed slow down and start to further decelerate due to the decrease of the interstop transport distance especially below about 20.00 km. If the target speed is improved from 90.00 km/h, a comparatively very short stop-spacing (e.g., S01-S02) is able to cease the decrease of the PTT per 10,000 p-km with the increase of the target speed by stopping the acceleration of the train in this stop-spacing for its safe stop at the next station before its achievement of the target speed.

The changes of the PTTs per 10,000 p-km of the train hauled by the SS8 with the increase of the target speed for different interstop transport distances are shown in Figure 4. If the target speed is lower than around 40.00 km/h, the PTTs per 10,000 p-km of the train hauled by the SS8 decrease rapidly with the increase of the target speed and get little impact of the transport distances between stops. If the target speed is higher than some 40.00 km/h, the decreases of the PTTs per 10,000 p-km with the increase of the target speed of the train become slow. The interstop transport distances shorter than approximately 20.00 km at this time begin to evidently consume additional PTT per 10,000 p-km and such a trend becomes obvious with the increase of the target speed as well as the decrease of the length of stop-spacing. If the target speed increase from 90.00 km/h, the decreases of the PTTs per 10,000 p-km start to stop by following the ascending sequence of the transport distances between neighboring stops because of the previously explained stopped acceleration of the train before its achieving the target speed. Moreover, some target speeds (e.g., 90.00 km/h) for certain stop-spacings (e.g., S01-S02) make the speed of the train at the rail sites where the train begins to check the necessity of brakes before arriving at the next stop reach or exceed the permitted speeds by coincidence. As interpreted in Section 2, such a situation makes the speed of the train decrease as soon as possible to a small value to ensure absolute security of the transport. Thereafter, the train runs with this low speed till it arrives at the next stop, which causes the increases of some PTTs per 10,000 p-km especially for comparatively short interstop transport distances.

As to an unchanged utilization ratio of the passenger seats and berths of the studied CLHP train in China, it is found that the improvement of the target speed of the train below about 40.00 km/h decreases its PTT per 10,000 p-km rapidly and the length of stop-spacing at this time has very little effect on the PTT per 10,000 p-km. The decrease of the PTT per 10,000 p-km with the increase of the target speed starts to slow down if the target speed

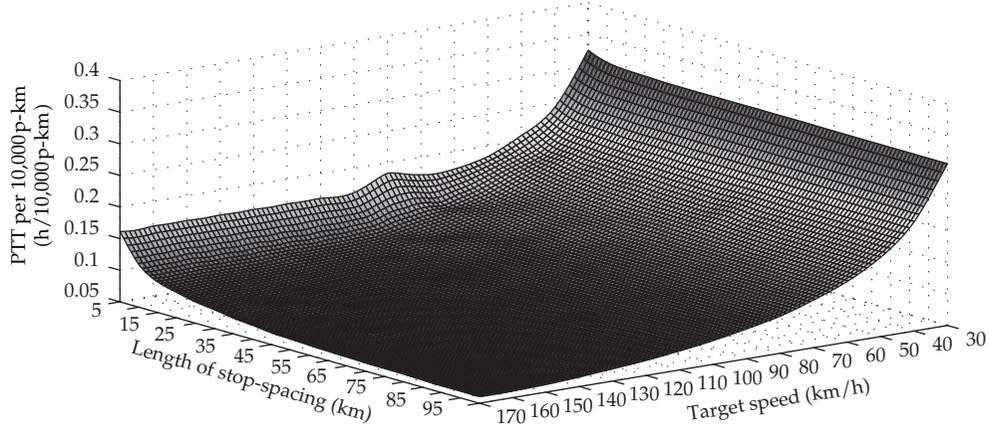


Figure 4: PTTs of the CLHP train hauled by the SS8.

becomes higher than around 40.00 km/h and begins to further decelerate with the decrease of the interstop transport distance especially below around 20.00 km. Such trends are able to be on the whole described by (3.3) according to curve fitting analyses:

$$t_{ij}^v = a \times v^b, \quad (a > 0), \quad (-1 < b < 0). \quad (3.3)$$

If v is smaller than approximately 40.00 km/h or D_{ij}^v is bigger than about 20.00 km, the parameters of a and b in (3.3) verge on constants whose values are determined mainly by the general traction performance of the applied locomotive. If v is over some 40.00 km/h and D_{ij}^v is below approximately 20.00 km, the absolute values of a and b are remarkably increased with the decrease of D_{ij}^v as interpreted by (3.4) and (3.5):

$$a = c_1^a \times \ln(D_{ij}^v) + c_2^a, \quad (c_1^a > 0), \quad (3.4)$$

$$b = c_1^b \times \ln(D_{ij}^v) + c_2^b, \quad (c_1^b < 0), \quad (3.5)$$

where c_1^a , c_2^a , c_1^b , and c_2^b are constant parameters and each type of locomotives has its own set of values for these parameters because of the specific traction performances of different types of locomotives.

In addition, when the target speed of a train exceeds 90.00 km/h, its reachable maximum speed may much differ from its target speed due to the restriction of the transport distance between neighboring stops, which will have the PTTs per 10,000 p-km of the train for different targets speeds over a certain value, that is, v^s , the same to each other for the same stop-spacing. As revealed in Figures 3 and 4, v^s is decreased with shortening the stop-spacing and also affected by the traction performance of the locomotive.

4. Conclusions

Now it is able to be empirically confirmed that decreasing the length of stop-spacing from about 20.00 km obviously increases the transport time per passenger-kilometer of a CLHP train especially for a relatively high target speed over approximately 40.00 km/h. In contrast, the interstop transport distances longer than some 20.00 km have almost no effect on the transport efficiency of the train. Such changes follow the trends generally interpreted by the power function of (3.3). As a result, it is clear that the target speed of a CLHP train is unnecessary to be kept as a consistent value which is especially much higher than about 40.00 km/h for the discounted efficiency of the train's transport between its originating and terminal stops due to the effect of some short interstop transport distances which also cause intensive traction energy consumption [9, 13]. In other words, the target speed of a CLHP train providing the transport service on a railway line whose transport capacity has not been much used should be flexibly decided according to the length of each stop-spacing.

Due to data limitation, only the transports of a typically formed CLHP train hauled by, respectively, two major types of the locomotives in China on a hypothetically straight and smooth rail line are studied in this research to explore the effect of the length of the stop-spacing on the PTT per 10,000 p-km for different target speeds of the train. The impacts of the ramps, bends, bridges, tunnels, and so forth of different rail lines on the transport time per passenger-kilometer of the transports completed by more kinds of trains which are formed by various numbers of different types of railway cars should be analyzed together with the effect of different transport distances between neighboring stops, traction performances of varied types of locomotives, and so forth for different target speeds from a more comprehensive viewpoint to further improve the conclusions of this study in the future.

Acknowledgments

This study is financially supported by the National Basic Research Program of China (2012CB725406) and the National Natural Science Foundation of China (71131001; 71201006).

References

- [1] National Bureau of Statistics of China, *China Statistical Yearbook*, China Statistics Press, Beijing, China, 2011.
- [2] D. N. Kim and P. M. Schonfeld, "Benefits of dipped vertical alignments for rail transit routes," *Journal of Transportation Engineering*, vol. 123, no. 1, pp. 20–27, 1997.
- [3] P. Lukaszewicz, *Energy consumption and running time for trains: Modelling of running resistance and driver behavior based on full scale testing [Doctoral Dissertation]*, Royal Institute of Technology, Stockholm, Sweden, 2001.
- [4] C. Chandra and M. M. Aqarwal, *Railway Engineering*, Oxford Higher Education, New York, NY, USA, 2008.
- [5] K. Kim and S. I. J. Chien, "Simulation-based analysis of train controls under various track alignments," *Journal of Transportation Engineering*, vol. 136, no. 11, pp. 937–948, 2010.
- [6] P. Martin, "Train performance and simulation," in *Proceedings of the Winter Simulation Conference (WSC '99)*, pp. 1287–1294, Phoenix, Ariz, USA, December 1999.
- [7] B. Mao, X. Li, and H. Niu, *Train Performance Calculation and Design*, China Communication Press, Beijing, China, 2008.
- [8] Z. Yang, X. Huang, S. Wu, and H. Peng, "Traction technology for Chinese railways," in *Proceedings of the International Power Electronics Conference (ECCE '10)*, pp. 2842–2848, Sapporo, Japan, June 2010.

- [9] IFEU (Institut für Energieund Umweltforschung Heidelberg GmbH), *Energy Savings by Light-Weighting (Final Report)*, International Aluminium Institute (IAI), Heidelberg, Germany, 2003.
- [10] H. Liu, B. Mao, Y. Ding, W. Jia, and S. Lai, "Train energy-saving scheme with evaluation in urban mass transit systems," *Journal of Transportation Systems Engineering and Information Technology*, vol. 7, no. 5, pp. 68–73, 2007.
- [11] B. van Wee, R. van den Brink, and H. Nijland, "Environmental impacts of high-speed rail links in cost-benefit analyses: A case study of the Dutch Zuider Zee line," *Transportation Research Part D*, vol. 8, no. 4, pp. 299–314, 2003.
- [12] E. Lindgreen and S. C. Sorenson, *Simulation of Energy Consumption and Emissions from Rail Traffic Evaluation*, Technical University of Denmark, Lyngby, Denmark, 2005.
- [13] X. Feng, "Optimization of target speeds of high-speed railway trains for traction energy saving and transport efficiency improvement," *Energy Policy*, vol. 39, no. 12, pp. 7658–7665, 2011.
- [14] X. Feng, B. Mao, X. J. Feng, and J. Feng, "Study on the maximum operation speeds of metro trains for energy saving as well as transport efficiency improvement," *Energy*, vol. 36, no. 11, pp. 6577–6582, 2011.
- [15] H. I. Andrews, *Railway Traction: The Principles of Mechanical and Electrical Railway Traction*, Elsevier, New York, NY, USA, 1986.
- [16] W. G. Wong, B. M. Han, L. Ferreira, X. N. Zhu, and Q. X. Sun, "Evaluation of management strategies for the operation of high-speed railways in China," *Transportation Research Part A*, vol. 36, no. 3, pp. 277–289, 2002.
- [17] C. W. Hsu, Y. Lee, and C. H. Liao, "Competition between high-speed and conventional rail systems: A game theoretical approach," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3162–3170, 2010.
- [18] J. Yan and M. Fu, *Automotive Engineering*, China Railway Publishing House, Beijing, China, 3rd edition, 2008.
- [19] International Union of Railways, *UIC Code 556 OR: Loadings of Coach Bodies and Their Components*, International Union of Railways, Paris, France, 3rd edition, 1990.

Research Article

Metropolis Parking Problems and Management Planning Solutions for Traffic Operation Effectiveness

Yuejun Liu, Wuhong Wang, Chengxi Ding, Hongwei Guo, Weiwei Guo, Liya Yao, Hui Xiong, and Huachun Tan

Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Wuhong Wang, wangwuhong@bit.edu.cn

Received 18 July 2012; Accepted 26 September 2012

Academic Editor: Huimin Niu

Copyright © 2012 Yuejun 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.

Advances in mobility are clearly illustrated by the rapid development of urbanization and motorization in developing countries. Following the dramatic incensement of traffic demand, the parking problem has been becoming much more seriously important in many metropolises. With the aim of seeking solutions as to how the parking system could operate more efficiently by using new technologies and new methodologies, this paper discusses the application of geographic information system into the parking planning and management for traffic operation effectiveness in metropolis. The concentration of this paper includes the characteristics of parking demand and the causations of parking problems, especially the basic parking principle and strategies for solving parking problems from the perspective of geographic information system are discussed in enough detail in this paper.

1. Introduction

Nowadays, the parking problems have been one of the most discussed topics by the general public [1, 2]. In many metropolises, the parking problems are becoming increasingly important. Following the rapid incensement of traffic demand, the imbalance between parking supply and parking demand has been considered as the main reason for metropolis parking problems. Moreover, the parking system plays a key role in the metropolitan traffic system, and the parking problems show closed relation with traffic congestion, traffic accident, and environment pollution.

The parking problems urge that the traffic professionals should seek more efficient solutions as to how the parking system could be used more efficiently and how parking

planning and management could be improved by using new technologies and new methodologies. Recently, the geographic information system (GIS) has received considerable attention regarding its potential as a powerful technique for solutions of parking problems. GIS is best utilized for integration of various data sets to obtain homogeneous composite land development units which helps in identifying the problem areas and suggests conservation measures [3, 4]. GIS technology offers extremely significant power in transport modeling. The spread of GIS use facilitates the efficient and portable spatial data storage, updating, and processing [5]. In all transportation planning cases, the satisfaction of total travel demand must be balanced with the provision of reliable transportation services and with minimizing the costs of externalities associated with road traffic [6]. With the development of the Internet technology, the web-based GIS has an applications into transportation environmental analysis for planning and management [7], and the Web, Java, and CORBA technologies were explored in implementing GIS-T systems [8]. GIS is desirable in researches of vehicle parking planning which is closely related to geographical location [9].

The improvement of GIS technology opens a new way for urban parking planning and policy-making. With the aim of using GIS technology to improve the planning and management of parking, this paper discussed the metropolis parking problems and corresponding solutions. The characteristics of parking demand and the reasons for parking problems are investigated in detail. Then, the basic parking principle and strategies for solving parking problems are suggested with the application of GIS through a case study of Beijing.

2. Key Issues of Parking Problems in Beijing

The main reasons for parking problems in Beijing can be concluded as the disparity between the supply of parking facilities and parking demand. The parking demand is very urgent which can be explained as high speed of growth, high intension of using, and high density of gathering [10]. The key issues of parking problems in Beijing are discussed as the following.

The disparity between the rapid increment of motor vehicle and limited new parking facilities results the parking difficulty directly. Since 2004, the number of private cars in Beijing grows 0.34 million annually. On the other hand, the supply of new parking facilities (including curbside parking and off-road parking) increases 0.10 million each year, let alone the lack of parking left over by history. In Beijing, the sharp gap between demand and supply of parking spaces is widening year by year. Therefore, the parking problems seem to be more and more serious. It must be mentioned that the inefficiency of facility usage aggravates the parking problem. For the reasons involving with economic, zone bit, and the characteristics of drivers, some of the parking facilities are used frequently while some others are used infrequently. Taking the commercial centers Wangfujing and Xidan for example, it is common that 20% to 30% of the underground and public parking lots are vacant while the curb side parking spaces are occupied by 120% to 150% of the top limit it is designed. Unfortunately, more curb side parking spaces distribute in the bypasses and Hutong system, and the capacity in these roads reduces very sharply. A new vicious cycle of narrow road and occupying road space by curb parking begins.

The lack of professional parking management is another issue in parking problems [11]. At present in Beijing, irregular parking that violates the transportation rules will be fined of 200 Yuan with no deduction of transportation credit points. It's also well known that there is very low chance that irregular parking is fined. All in all, the cost for penalty

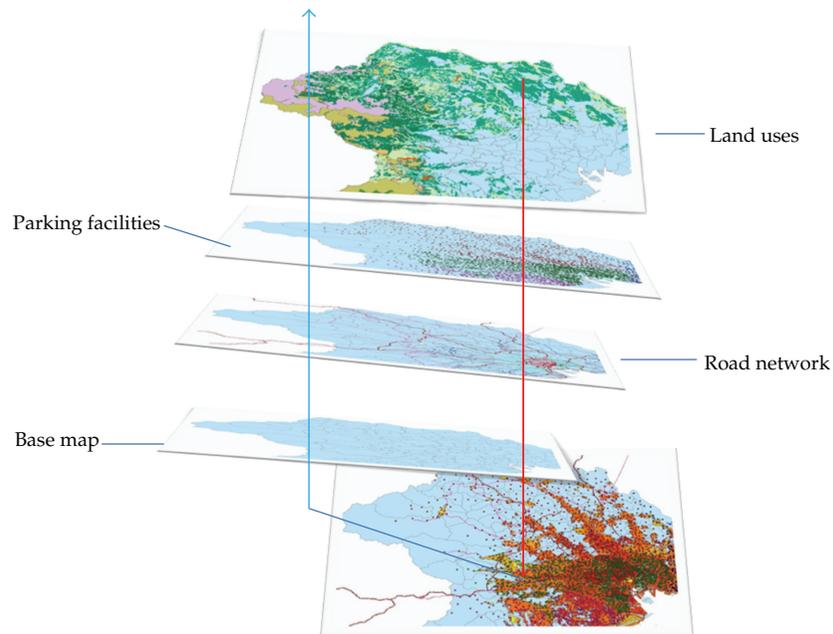


Figure 1: Map overlays.

of irregular parking is quite low and thus virtually increases the comparative opportunity cost. Comparatively in a metropolis such as those in Japan, irregular parking will be strictly punished and recorded. An interesting phenomenon is that the motor vehicles are used more intensive in Beijing, especially for the official car [12, 13]. The intensive using of motor vehicles leads to high demand of parking spaces while it is an uneven distribution in space and demand. During the period of commuting time, a large part of cars are driven out from garages and need parking lots afterwards, and thus, the accumulated time for parking is comparatively long, that is, demand for parking stays at specific time intervals in the morning and afternoon whereas parking supply is evenly distributed all day. The more intensive traffic demand facing the nonprofessional parking management would make a turn for a worse.

3. A Case Study for Parking Planning and Management

On parking planning, the need for new parking facilities should best be considered with due regard to the existing nearby parking facilities. With the powerful features typical of a GIS, it will allow decision-makers to visualize the need for the new parking facilities in relation to the existing nearby parking facilities not just in terms of their availability, but also the proximity through examining the topological relationships. Through the ArcGIS software platform, we provided decision support for a project of parking management and facilities in a city (Figure 1). Parking facility management and parking planning process in GIS is as follows.

Step 1. to determine input datasets. In this case (Figure 2), the input data included road network layout, traffic zone layout, and survey data (e.g., parking duration, parking type) [14].

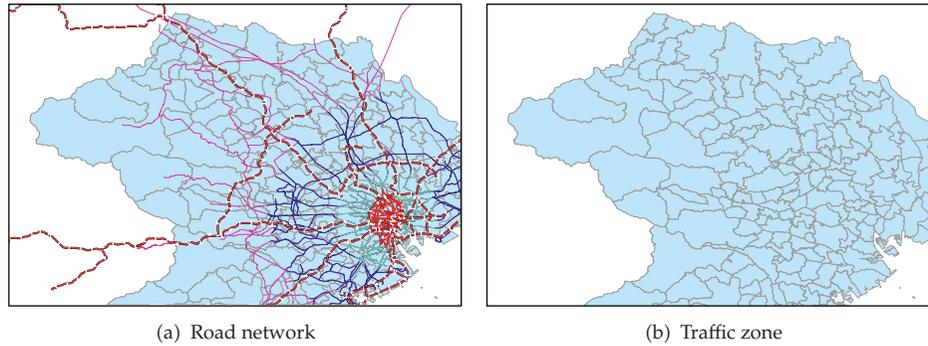


Figure 2: Basic datasets.

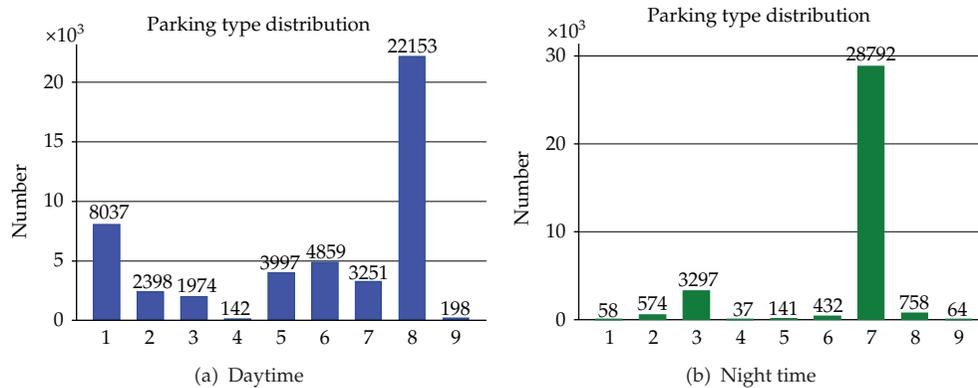


Figure 3: The distribution of parking types.

Step 2. derived datasets. In this case, derived data sets include parking distribution, parking times, parking fee, parking category, and so forth. As an example, the parking type distribution is shown in Figure 3. In the figure, the numbers from 1 to 9 represent different parking types, such as road-side parking, public off-road parking lot, residential parking facilities, compound units, and temporary parking area [15].

Step 3. to reclassify datasets. The datasets are reclassified according to the existing parking facilities, the distribution of parking, the type, and so forth (Figure 4).

Step 4. to weight and combine the datasets. Through GIS spatial analysis, different weights are assigned to the datasets, so as to determine the parking areas with the most serious parking problems. Thus it can help to decide to plan and build parking facilities.

Figure 5 will be able to be very intuitive reflect the city's parking situation, as a result, it can provide decision support for making related policies and measures for parking problems and management.

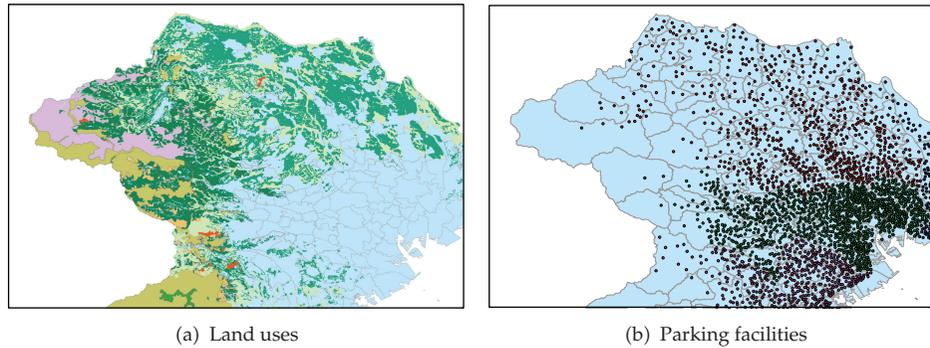


Figure 4: The existing parking facilities.

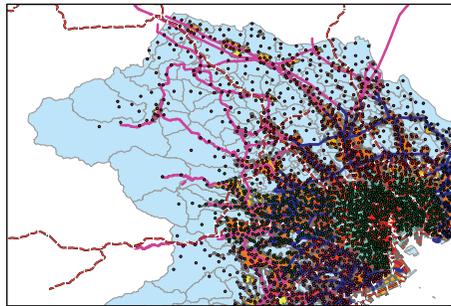


Figure 5: The distribution of parking problems.

4. Conclusions

Just like congestion, traffic accident, and environment pollution, the parking problems are the production of the urban urbanization and motorization; They are a concentrated expression of contradiction between car transitional needs and environmental conditions. New York, London, Tokyo, and other major cities in the process of urbanization, also face serious parking problems, which result in frequent and multiple traffic congestion and traffic accidents, and even gave a tremendous impact on the lives of the residents. Therefore, parking management is an important means of regulating motor process. To be aware of the parking problem is a reasonable solution to the parking problems, which makes technology and means of propaganda very necessary.

Even in the most efficient parking program, the communication is increasingly challenging. It is important for metropolis to communicate to the public that it is serious about regulations and enforcement. Booting and towing programs send the highly visible message to others that the metropolis is serious. The metropolis has to take booting and towing one step further through the use of the GIS application. The GIS web-based application data mines information stored in the parking management information system data warehouse to provide additional management insight. With GIS, a user can visually see characteristics, data, and information overlapping a base map. Through the features of GIS software, data comparisons can be made, trends can be identified, and action plans can be developed.

Acknowledgments

This research was supported in part by the Programme of Introducing Talents of Discipline to Universities under Grant B12022 and National Natural Science Foundation of China under Grant 51010305084, 51110305060, 51210305046.

References

- [1] T. K. Xia, M. Yang, R. Q. Yang, and C. Wang, "CyberC3: a prototype cybernetic transportation system for Urban applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 142–152, 2010.
- [2] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: a fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.
- [3] J. N. A. Murthy, P. R. K. Reddy, B. Pateriya, and M. N. Kumar, "Integrated study for sustainable development using remote sensing and GIS techniques- a case study for part of Musicatchment, Rangareddy District, A. P.," in *Remote Sensing Applications*, Narosa Publishing House, 2001.
- [4] G. Arampatzis, C. T. Kiranoudis, P. Scaloubacas, and D. Assimacopoulos, "A GIS-based decision support system for planning urban transportation policies," *European Journal of Operational Research*, vol. 152, no. 2, pp. 465–475, 2004.
- [5] P. Ferrari, "A model of urban transport management," *Transportation Research Part B*, vol. 33, no. 1, pp. 43–61, 1999.
- [6] P. D. Coddington, K. A. Hawick, and H. A. James, "Web-based access to distributed high-performance geographic information systems for decision support," in *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences*, pp. 6015–6016, Maui, Hawaii, USA, January 1999.
- [7] X. F. Li, S. R. Tu, X. F. He, and J. Ratcliff, "Web-based distribution of GIS metropolitan maps," in *Proceedings of the 5th International Conference on Information Visualisation*, pp. 419–424, London, England, 2001.
- [8] J. P. Hu, J. G. Wei, and H. M. Wang, "Traffic planning based on GIS," *Journal of the China Railway Society*, vol. 4, no. 22, pp. 13–15, 2000.
- [9] Beijing Transportation Research Center, "Beijing Transportation Development Annual Report," Beijing Transportation Research Center, Beijing, China, 2010.
- [10] P. C. Box, "Curb-parking problems: overview," *Journal of Transportation Engineering*, vol. 130, no. 1, pp. 1–5, 2004.
- [11] X. P. Zheng and H. W. Li, "Countries around The World Face "stop hard"," *Road Traffic Management*, vol. 3, pp. 24–25, 2010.
- [12] H. W. Guo, Z. Y. Gao, X. B. Yang, X. M. Zhao, and W. H. Wang, "Modeling travel time under influence of on-street parking," *Journal of Transportation Engineering*, vol. 138, no. 2, pp. 229–235, 2012.
- [13] W. E. Marshall, N. W. Garrick, and G. Hansen, "Reassessing on-street parking," *Transportation Research Record*, no. 2046, pp. 45–52, 2008.
- [14] D. Stevens, S. Dragicevic, and K. Rothley, "iCity: a GIS-CA modelling tool for urban planning and decision making," *Environmental Modelling and Software*, vol. 22, no. 6, pp. 761–773, 2007.
- [15] G. Spadoni, D. Egidi, and S. Contini, "Through ARIPAR-GIS the quantified area risk analysis supports land-use planning activities," *Journal of Hazardous Materials*, vol. 71, no. 1-3, pp. 423–437, 2000.

Research Article

A Cellular Automaton Traffic Flow Model with Advanced Decelerations

Yingdong Liu

College of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

Correspondence should be addressed to Yingdong Liu, liuyingdong@mail.lzjtu.cn

Received 1 June 2012; Revised 5 August 2012; Accepted 5 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Yingdong Liu. 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 one-dimensional cellular automaton traffic flow model, which considers the deceleration in advance, is addressed in this paper. The model reflects the situation in the real traffic that drivers usually adjust the current velocity by forecasting its velocities in a short time of future, in order to avoid the sharp deceleration. The fundamental diagram obtained by simulation shows the ability of this model to capture the essential features of traffic flow, for example, synchronized flow, meta-stable state, and phase separation at the high density. Contrasting with the simulation results of the VE model, this model shows a higher maximum flux closer to the measured data, more stability, more efficient dissolving blockage, lower vehicle deceleration, and more reasonable distribution of vehicles. The results indicate that advanced deceleration has an important impact on traffic flow, and this model has some practical significance as the result matching to the actual situation.

1. Introduction

With the drastic growth in the vehicle amount, traffic congestion has become more and more serious. The transportation problems have caused extensive attention to transportation industry, physics community, and mathematics community [1–13]. Owing to the properties of cellular automaton model (CA Model), such as the discreteness of space and time, the simplicity and flexibility of algorithm and the easiness to be simulated on computer, it could effectively simulate the vehicles movement in traffic and be widely used and developed in the study of traffic flow [1–8]. In 1992, Nagel and Schreckenberg proposed the famous NaSch model [1]. The model deals with single-lane traffic flow of cars moving in a one-dimensional cellular chain under periodic boundary conditions, which only considers vehicle acceleration, deceleration, random delays, and update of vehicle location. According to the above simple rules, this model can be used to reproduce the basic phenomena encountered in real traffic, for example, the occurrence of phantom traffic jams. However, some features have

not been considered in the NaSch model, such as the metastable state. To fix the omission, a number of improvements for the evolution rules have been proposed [2–8]. Literature [2] proposed the BJH model, which considers a possible delay before a car pulls away from being stationary due to the blocking of the leading vehicle. Literature [3] considered a velocity-dependent randomization delay probability and presented the VDR Model. Literature [4] took into account the velocity effect of a car on the successive and proposed the VE (velocity effect) model. Literature [5] presented a cellular automaton model for single-lane traffic flow, and the model considers the effect of headway distance between two successive cars on the randomization of the latter car. Literature [6] took into account the diversity of traffic behaviors under real traffic situations induced by various driving characters and habits to modify the weighted probabilistic cellular automaton model. Literature [7] modified the NaSch model by enabling the randomization probability to be adjusted on the bases of drivers' memory. Literature [8] proposed an improved cellular automaton model to describe the urban traffic flow with the consideration of traffic light and driving behavior effects. All of the CA Models adjust the current velocity only based on considering the velocity of vehicle itself, correlating velocity, vehicle distance, safe distance, and so forth in the perspective of space, and the adjustment of the vehicle velocity is often achieved by different random deceleration properties P . In the process of evolution, all the vehicles view the maximum velocity as the desired velocity, in other words, all the vehicles expect to reach the maximum velocity by gradual acceleration if the travel condition ahead permits. However, because the models do not take into account the deceleration of vehicles, it can lead to a terrible situation: if the following vehicle moves close to a still vehicle at the maximum velocity, it will have a sharp deceleration from the maximum velocity to zero in the next second. If that situation does occur, it is inevitably to get a rear-end collision in the actual traffic.

In the actual traffic, in order to avoid the happening of sudden deceleration and rear-end collision and to drive the vehicle steadily, the current vehicle velocity depends not only on the velocity of itself, the relative velocity with the leading vehicle, the distance between the vehicles, the safety distance, and so forth, but also on the velocity changing trend of the current vehicle and the leading vehicles. Therefore, at any time of driving, drivers must estimate the velocity in next few seconds and decide whether to accelerate, decelerate, or keep the velocity. For example, when meeting a red traffic light or the front vehicle moving slowly and not accelerating in next few seconds, the vehicle needs decelerate in advance. On the basis of the NaSch model of traffic flow, this paper proposes a one-dimensional cellular automaton traffic flow Model with advanced decelerations. Considering time factor, the model adjusts the current velocity by the forecasting velocity of the next several time steps to avoid the sharp deceleration to some extent, enhances the stable performance of traffic flow and keeping the smooth flow of traffic better. Our model shows the metastable state, phase separation, and hysteresis phenomenon by computer simulating, which exists in actual traffic.

2. The Establishment of Model

In the NaSch Model, vehicles randomly distributed in a one-dimensional discrete cellular chain, whose length is L . Each cell may either be empty or be occupied by one vehicle. All vehicles are assumed to move from the left to the right. The n th vehicle in the time-step t is located at the position $x_n(t)$, moving with an integral velocity $v_n(t) \in \{0, \dots, v_{\max}\}$, where $n \in \{1, 2, \dots, N\}$ and v_{\max} ($v_{\max} \geq 1$) are the maximum velocity which a vehicle can reach. The gap between consecutive vehicles is $d_n(t) = x_{n+1}(t) - x_n(t) - l_{n+1}$, which is the number of

empty cells in front of the n th vehicle in the time-step t . At each discrete time-step $t \rightarrow t+1$, an arbitrary arrangement of N vehicles is updated in parallel according to the following rules:

- (i) acceleration: $v_n(t+1) = v_n(t) + 1$;
- (ii) deceleration: $v_n(t+1) = \min(v_n(t+1), d_n)$;
- (iii) randomization: $v_n(t+1) = \max(v_n(t+1) - 1, 0)$ with the probability P ;
- (iv) update of location: $x_n(t+1) = x_n(t) + v_n(t+1)$.

In some expansion and amendments to the NS Model, a variety of ways to preestimate the velocity $v'_{n+1}(t+1)$ of the leading vehicle in the next time-step were used, such as the VE model [4], and deceleration rules were updated as follows:

$$v_n(t+1) = \min(d_n(t) + v'_{n+1}(t+1), v_{\max}). \quad (2.1)$$

Because the deceleration rules in NaSch and its expansive models are based on the greedy mechanism, the vehicles move forward by the currently allowed maximum velocity as much as possible. When the velocity difference is bigger, and the gap is smaller between consecutive vehicles, according to the given decelerating rules, it is inevitable that the following vehicle will abruptly decelerate. In the model of this paper, in order to avoid this condition, the velocities in the next *steps* time-steps are firstly estimated before determining the velocity in current time. If the abrupt deceleration exists in these next *steps* time-steps, the method that the vehicle gradually decelerates in advance is used to ensure the safe driving. Of course, as the result of the increasing of time-steps *steps*, there are more errors caused by estimation. So the paper sets $steps \leq 3$, equal to predicting velocity in 3 seconds. The vehicle velocity estimation is based on greedy mechanism.

(i) From a given configuration at time-step t , the forward effective distance $d_n^{\text{eff}}(t+1)$ at the next time-step $t+1$ can be obtained by

$$d_n^{\text{eff}}(t+1) = d_n + \max(v_{n+1}^{\text{anti}} - \text{gap}_{\text{safe}}, 0). \quad (2.2)$$

Among $v_{n+1}^{\text{anti}} = \min(d_{n+1}(t), v_{n+1}(t))$ is the anticipated velocity of the leading vehicle, gap_{safe} is the security gap. Then calculate estimated velocity $v_n^{\text{est}}(t+1)$ at the time-step $t+1$ by

$$v_n^{\text{est}}(t+1) = \min(d_n^{\text{eff}}(t+1), \min(v_n(t) + 1, v_{\max})). \quad (2.3)$$

At the same time, this step also confirms the permitted maximum velocity at the time-step $t+1$.

(ii) Using the same method, the vehicle velocity at the time-step $t+i+1$ can be calculated according to estimated velocity $v_n^{\text{est}}(t+i)$ and estimated effective distance $d_n^{\text{est}}(t+i)$ at the time-step $t+i$ ($i = 2 \cdots steps - 1$), until the estimated velocity $v_n^{\text{est}}(t+steps)$ at the time-step $t+steps$ is obtained.

(iii) Calculate the possible moving distance $l_{\text{steps}}^{\text{est}} = \sum_{i=1}^{\text{steps}} v_n^{\text{est}}(t+i)$ of the n th vehicle in the steps time-steps, if $l_{\text{steps}}^{\text{est}} \geq \text{steps} \times v_n(t)$, it means the n -th vehicle will accelerate or keep current velocity during the period from time-step $t+1$ to time-step $t+\text{steps}$. The velocity at the time-step $t+1$ determined as

$$v_n(t+1) = v_n^{\text{est}}(t+1). \quad (2.4)$$

Otherwise, it means the n th vehicle must abrupt deceleration during the period from time-step $t+1$ to time-step $t+\text{steps}$. In order to decelerate gradually, given the descent velocity difference of each time-step is Δv during the steps time-steps, the moving distance $l_{\text{steps}}^{\text{est}}$ can be expressed as

$$l_{\text{steps}}^{\text{est}} = \sum_{i=1}^{\text{steps}} (v_n(t) - i \times \Delta v). \quad (2.5)$$

Thus the descent velocity difference Δv is further attained as follows:

$$\Delta v = \frac{2 \times \text{steps} \times v_n(t) - 2 \times l_{\text{steps}}^{\text{est}}}{\text{steps} \times (\text{steps} + 1)}. \quad (2.6)$$

And the velocity at the time-step $t+1$ determined as

$$v_n(t+1) = \min(v_n(t) - \Delta v, v_n^{\text{est}}(t+1)). \quad (2.7)$$

In the process of the velocity estimated, the security gap gap_{safe} is important. Accidents can be avoided only if the security gap is ensured. Here we claim that the security gap is variable according to the velocity of the leading car. Because the gradually deceleration rule is used in this model, the velocity of the leading car may be less than anticipated velocity v_{n+1}^{anti} when it decelerate in advance. In order to drive safely, we always consider the velocity of the leading car as the worst condition and define the security gap as

$$\text{gap}_{\text{safe}} = \frac{v_{n+1}^{\text{anti}}}{\text{steps} + 1}. \quad (2.8)$$

In order to analyze the impact of driver's skill level and psychological conditions on the flow, the cars are divided into two types, AD and NAD. The cars in AD can judge the velocity at the next time-step depending on the rules in this model. While the cars in NAD always use the estimate velocity $v_n^{\text{est}}(t+1)$ as the velocity at the time-step $t+1$, r_{ad} is used to represent the ratio of drivers in AD and meanwhile $r_{\text{ad}} = 1$ means all of the drivers are decelerating in advance.

Based on the above, the rules of the traffic flow model based on Advanced Decelerations are set as follows.

(i) Acceleration: $v_n(t+1) = \min(v_n(t) + 1, v_{\max})$.

(ii) Estimated velocity from time-step $t+1$ to $t+steps$:

$$\begin{aligned} v_n^{\text{est}}(t+1) &= \min\left(d_n^{\text{eff}}(t+1), \min(v_n(t) + 1, v_{\max})\right) \\ v_n^{\text{est}}(t+i) &= \min\left(d_n^{\text{eff}}(t+i-1), \min(v_n^{\text{est}}(t+i-1) + 1, v_{\max})\right) \quad (i = 2 \cdots steps). \end{aligned} \quad (2.9)$$

(iii) Deceleration: calculate prediction driving distance $l_{\text{steps}}^{\text{est}}$ and descent velocity difference Δv in $steps$ time-steps, then

$$v_n(t+1) = \begin{cases} \min(v_n(t) - \Delta v, v_n^{\text{est}}(t+1)) & \text{if } n \in \text{AD}, l_{\text{steps}}^{\text{est}} < steps \times v_n(t), \\ v_n^{\text{est}}(t+1) & \text{otherwise.} \end{cases} \quad (2.10)$$

(iv) Randomization: with a certain probability p do

$$v_n(t+1) = \max(v_n(t+1) - 1, 0). \quad (2.11)$$

(v) Update of location

$$x_n(t+1) = x_n(t) + v_n(t+1). \quad (2.12)$$

In this model, when the current vehicle velocity is decided, the space factors, such as the security gap and the velocity of itself and the leading, are not only considered, but also the space factors, the velocities in the next $steps$ time-steps are estimated. When $steps = 1$, the model degenerate to the VE model [4].

3. Results and Analysis of Numerical Simulation

In the simulations, the length of each cell is given by 1.5 m, and $L = 5 \times 10^3$ cells are assumed. The length of each vehicle is 7.5 meters, which occupies $l = 5$ cells. The periodic boundary condition is assumed. One time-step is taken as 1 s. The maximum velocity is taken as $v_{\max} = 20$, which corresponds to the speed 108 km/h in real traffic. N is the total number of the vehicles distributed on the selected road and v_n is the velocity of the n th vehicle. The mean flow is calculated via the relation $q = \rho \times \bar{v}$, meanwhile the car density is $\rho = N \times l / L$ and the mean velocity is $\bar{v} = (1/N) \sum_{i=1}^N v_i$.

When we started to perform numerical simulation, all vehicles with a given density were initially arranged randomly on the whole lane. Each run was conducted 1×10^4 time-steps, in order to remove the transient effects, we discarded the data of the first 4×10^4 time-steps. The mean velocities were obtained by averaging over 30 runs. Because the flux from the

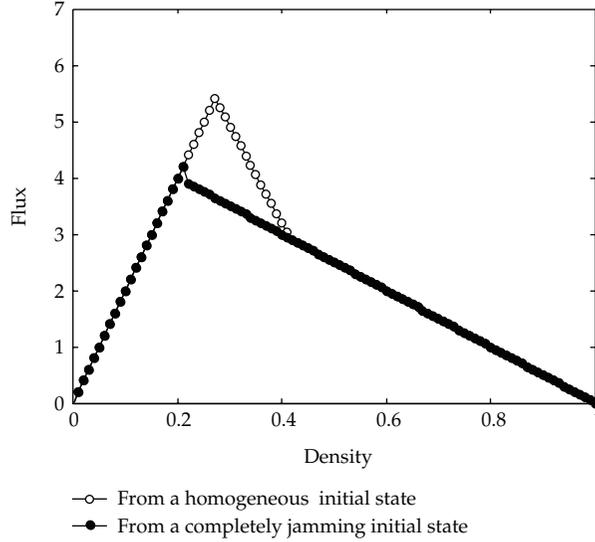


Figure 1: The fundamental diagram under $P = 0$, $r_{ad} = 1$, and $steps = 1, 2, 3$.

VE model are higher than those from the NS model and are much closer to the measurements results [4], the VE model is used as comparison.

Figure 1 depicts the fundamental diagram under the deceleration probability $P = 0$ and AD ratio $r_{ad} = 1$ together with the different estimated time-steps $steps$. In the diagram, the model curves under different estimated steps $steps$ coincide with each other, at the same time, the hysteresis phenomenon which is similar to VE Model [4] can be observed. The fundamental diagrams are obtained from two different types of initial states: the homogeneous state and the completely jamming state.

Figure 2 depicts the fundamental diagram under the deceleration probability $P = 0.3$ and AD ratio $r_{ad} = 1$ together with the different estimated time-steps $steps$ which are also obtained from two different types of initial states. In Figure 2, ρ_1 and ρ_2 show the positions of the two phase transition points, which divide the whole density scale into three phases having different macroscopic characters of traffic flow. A tiny difference in densities near the two phase-transition points can produce totally different steady states in the end. For example, Figures 3(b) and 3(c) show the phase-transition near ρ_1 ; Figures 3(d) and 3(e) show the phase transition near ρ_2 . In the region $\rho < \rho_1$, homogeneous states are steady as shown in Figures 3(a) and 4(b). Jamming states will dissolve soon, and the traffic will turn into a homogeneous free flow after evolvement. In the region $\rho > \rho_2$, jamming states are steady as shown in Figures 3(e) and 3(f). Little Jamming states will assemble to large ones, and the traffic will be steady on a phase-separated state in the end. There are one or more large jams and several free flow regimes in the final system. In the $\rho_1 < \rho < \rho_2$ region, the global prospect of the steady states is reflected in Figures 3(c), 3(d), 3(e), and 3(f). The hysteresis curves, which are characteristics similar to that of VE models [4], can be observed. The metastable states can stay long and finally evolve to different steady states determined by the initial conditions. For example, Figure 3(c) shows the diagram from a homogeneous initial state while the diagram in Figure 3(d) is from a jamming initial state. The former is

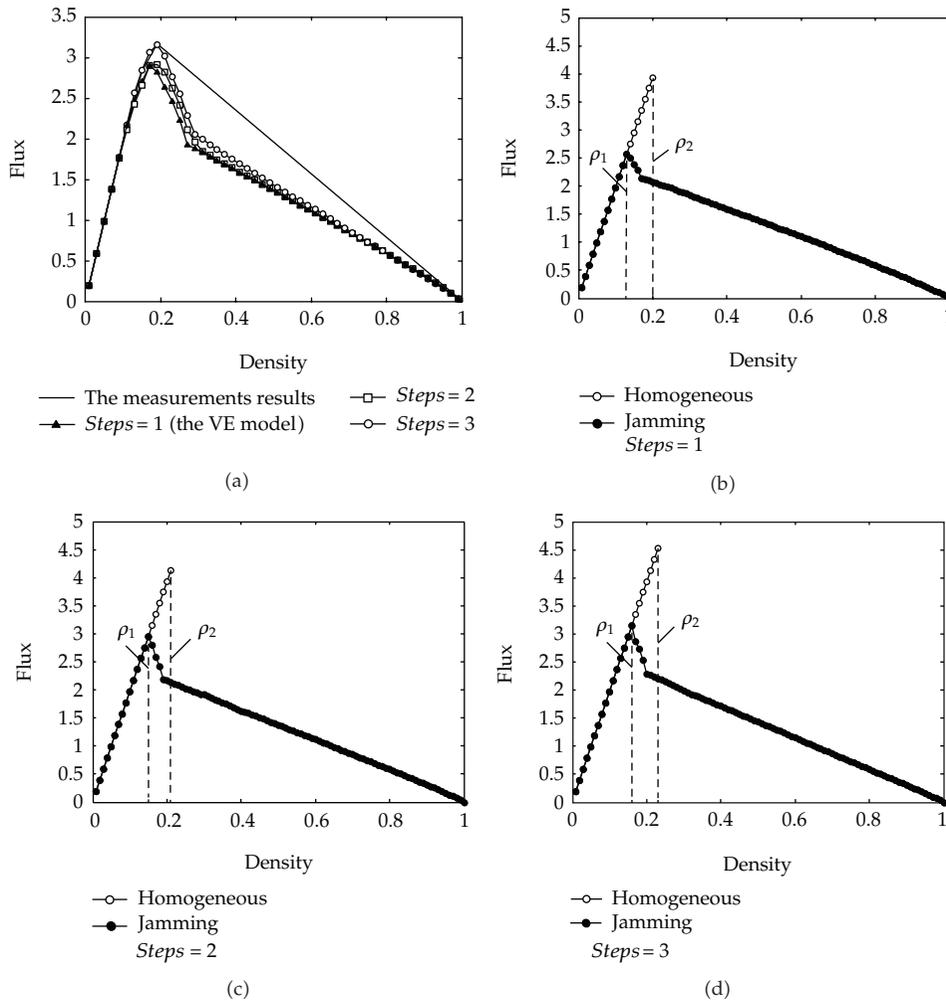


Figure 2: The fundamental diagram under $P = 0.3$ and $r_{ad} = 1$ (a) $steps = 1, 2, 3$; (b) $steps = 1$; (c) $steps = 2$; (d) $steps = 3$. “homogeneous” is obtained from a homogeneous initial state while the lower, “jamming” is from a completely jamming initial state.

symmetrical everywhere, while the following shows a road which consists of a jamming part and a homogeneous part. Both of the global densities are $\rho = 0.154$. Only from different initial states, different outcomes are totally obtained.

Although similarity can be found in the different estimated time-steps $steps$, there is essential distinction between them. In Figure 2(b), the critical points are $\rho_1 = 0.137$, $\rho_2 = 0.20$, and the maximum flux is 3.939 under $steps = 1$ (the VE model). In Figure 2(c), the critical points are $\rho_1 = 0.145$, $\rho_2 = 0.21$, and the maximum flux is 4.134 under $steps = 2$. In Figure 2(d), the critical points are $\rho_1 = 0.154$, $\rho_2 = 0.23$, and the maximum flux is 4.523 under $steps = 3$. One can see that with the increase of estimated time-steps, the critical points and the maximum flux become increasingly big; meanwhile, they are closer and closer to the real data.

Figure 4 depicts the maximum change of the velocity at each step versus the density under the different p . Figure 4(a) shows that the maximum changing quantity of the velocity

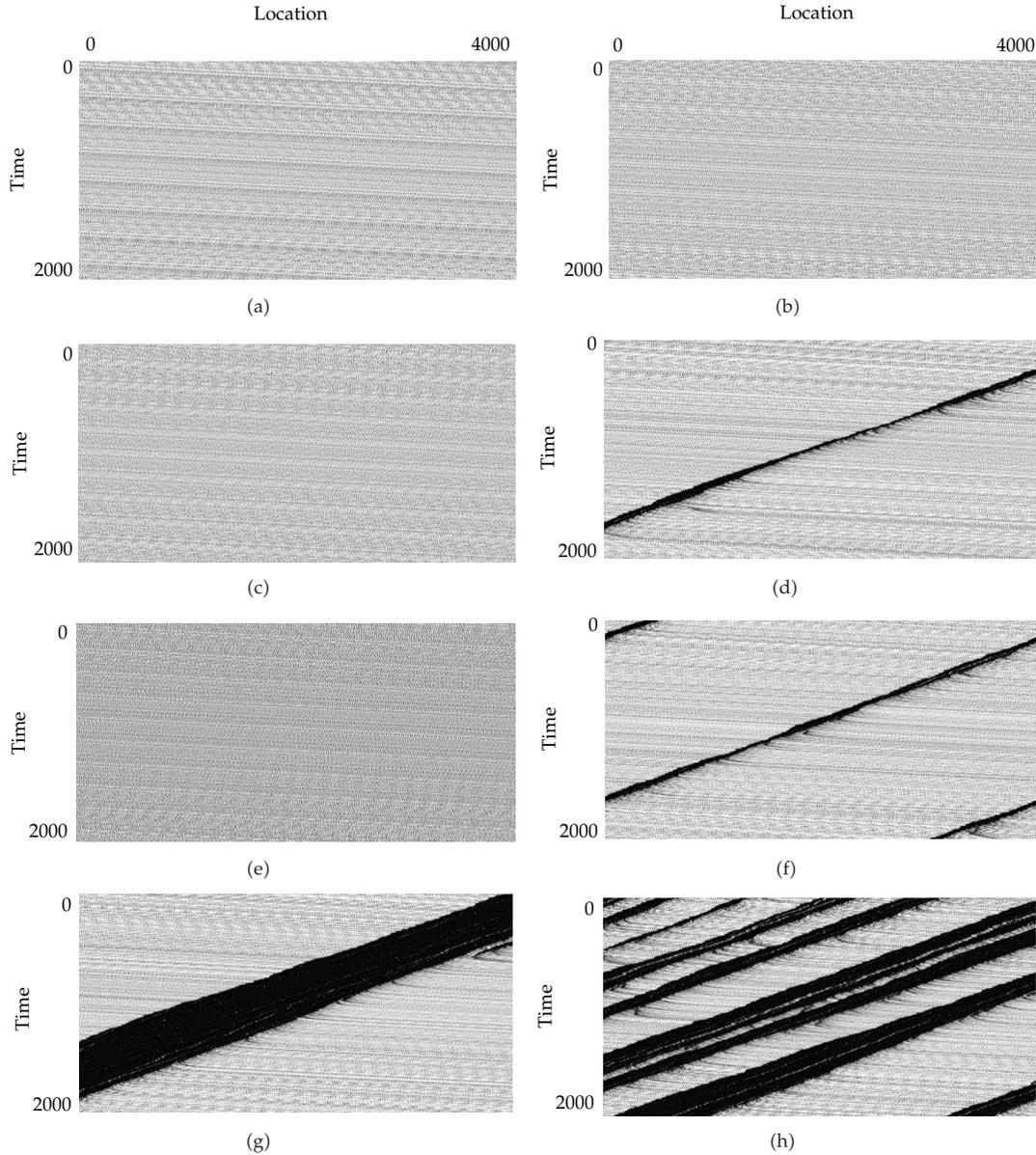


Figure 3: The spatial-temporal diagrams about phase transitions under $steps = 3$ (a), (b) $\rho = 0.153$; (c), (d) $\rho = 0.154$; (e), (f) $\rho = 0.23$; (g) $\rho = 0.24$; (h) $\rho = 0.4$. (a), (c), (e), and (g) are from homogeneous initial states, while (b), (d), (f), and (h) are from a jamming initial state.

under $p = 0$ will reach 19 cells when $steps = 1$, while it is 6 and 4, respectively, when $steps = 2$ and 3. Figure 4(b) shows that the maximum changing quantity of the velocity under $P = 0.3$ reach 20 cells where $steps = 1$, while it is 11 and 9, respectively, under $steps = 2$ and 3. These show that with the increase of estimated time-steps, the maximum changing quantity of each time-step velocity decreases obviously which is in accordance with the real conditions.

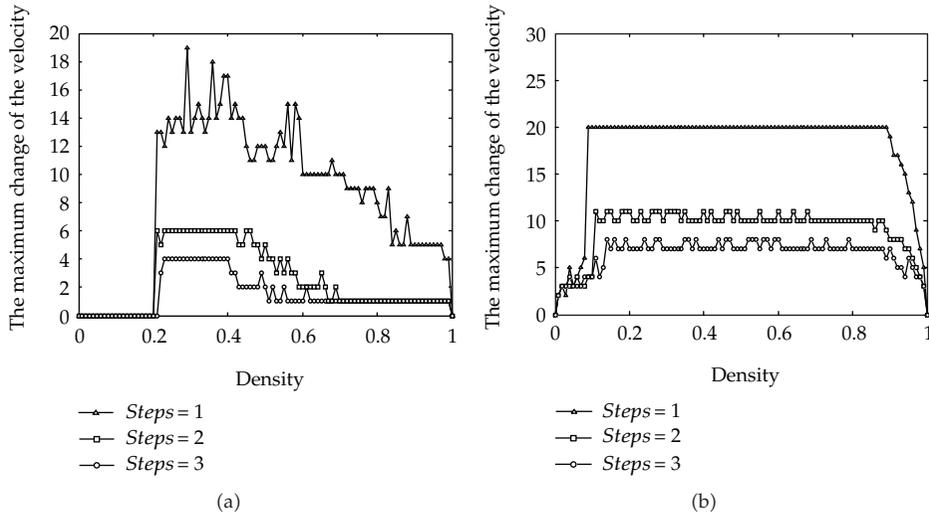


Figure 4: The maximum change of the velocity at each step versus the density under the different P (a) $P = 0$; (b) $P = 0.3$.

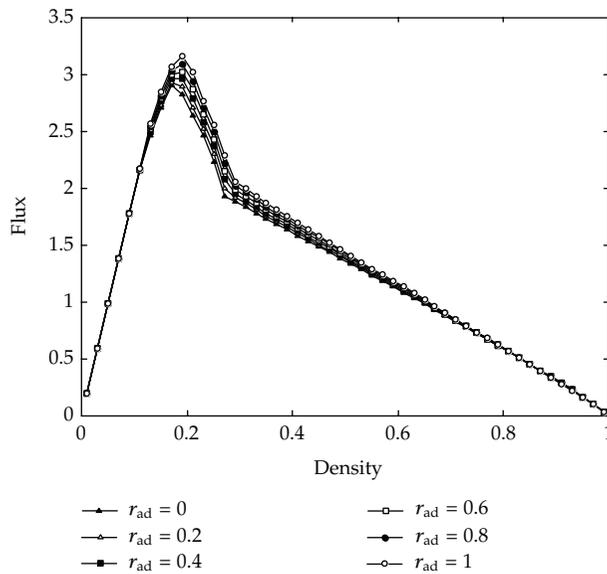


Figure 5: The fundamental diagram under $P = 0.3$ and $steps = 3$.

It means that according to estimated results to adjust vehicle velocity in advance can reduce the sudden deceleration and improve the driving safety.

Figure 5 shows the fundamental diagram under the deceleration probability $P = 0.3$ together with the different AD ratio r_{ad} . In Figure 5, the flux is almost the same as above when traffic is a homogeneous free flow. However, the flux gradually improves under the same estimated time-steps with the increase of the AD ratio r_{ad} while it also gradually

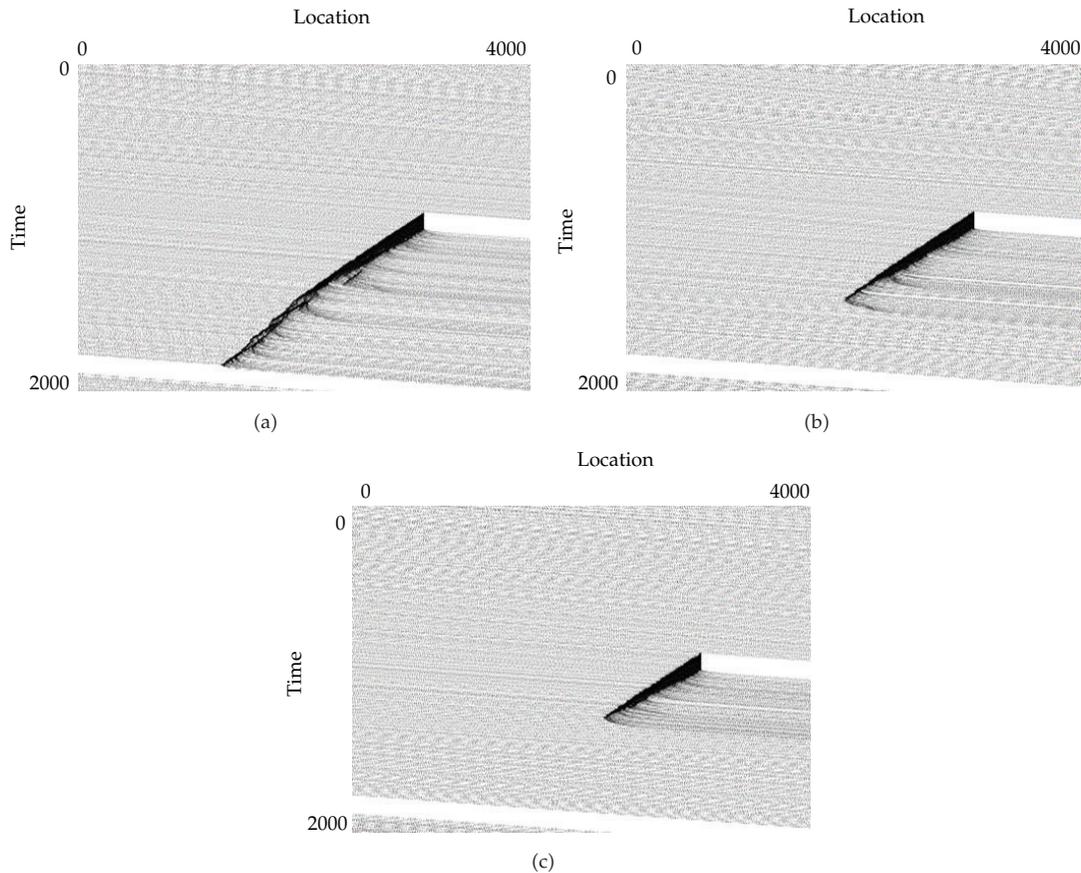


Figure 6: Blockage ablation simulation figure under $\rho = 0.1$, $P = 0.1$, $r_{ad} = 1$ (a) $steps = 1$; (b) $steps = 2$; (c) $steps = 3$.

improves under the same AD ratio r_{ad} with the increase of the estimated time-steps when the traffic density is higher. These illustrate that when the driver chooses the currently permitted maximum velocity as much as possible, traffic flow may not achieve maximum actually. However, it is beneficial to improve traffic flow by adjusting their driving behaviors in advance, which is in accordance with the real conditions.

A signal light is placed in the middle of the lane. When it evolves to 10000 time step, the green light turns to red light lasting for 100 steps, and then the green light on. The simulation results of Figure 6 show that the traffic jam caused by red light can be eliminated in a shorter time and be restored to the free flow state when $steps > 1$ (Figures 6(b), 6(c)). Although the traffic jam can be eliminated when $steps = 1$ (Figure 6(a)), it must go through a long time. Comparing with the simulation results, these illustrate that the driver can be better to respond to emergencies that may occur at anytime in traffic when he selects the appropriate speed forward. As a result, this will enhance the stability of the traffic flow and be better to maintain the traffic flow smooth. But the blindly following will worsen traffic congestion (see Figure 7).

Figure 7 shows the velocity fluctuations in evolution process of 1000 steps. In Figures 7(a), 7(b), and 7(c), when the density is lower ($\rho = 0.1$) under the state of $steps = 1, 2$, and 3 , vehicle can be maintained at a relatively smooth moving, and the velocity fluctuations stay

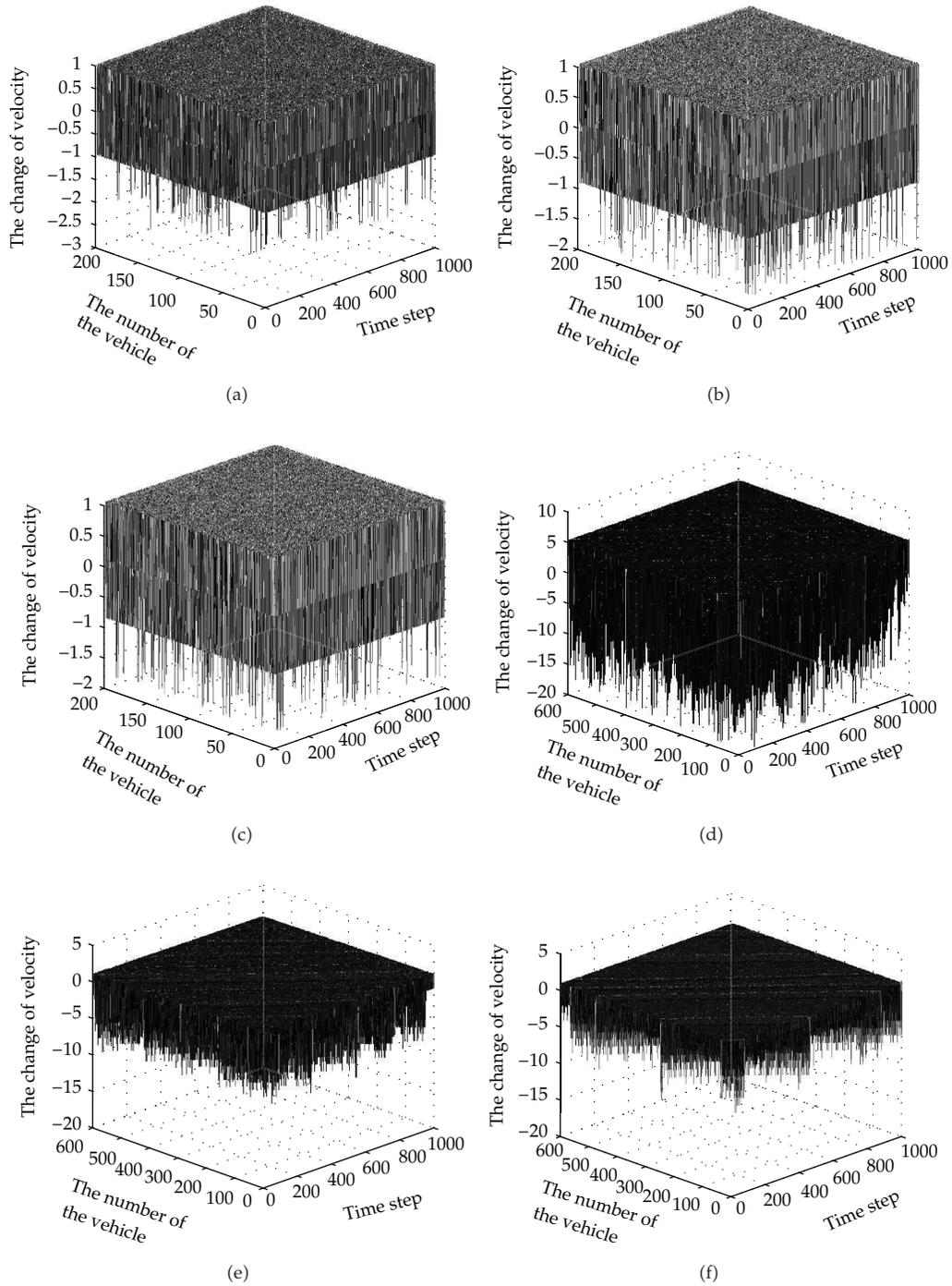


Figure 7: The velocity fluctuations map in 1000 time steps under $P = 0.1$, $r_{ad} = 1$ (a) $\rho = 0.1$, steps = 1; (b) $\rho = 0.1$, steps = 2; (c) $\rho = 0.1$, steps = 3; (d) $\rho = 0.3$, steps = 1; (e) $\rho = 0.3$, steps = 2; (f) $\rho = 0.3$, steps = 3.

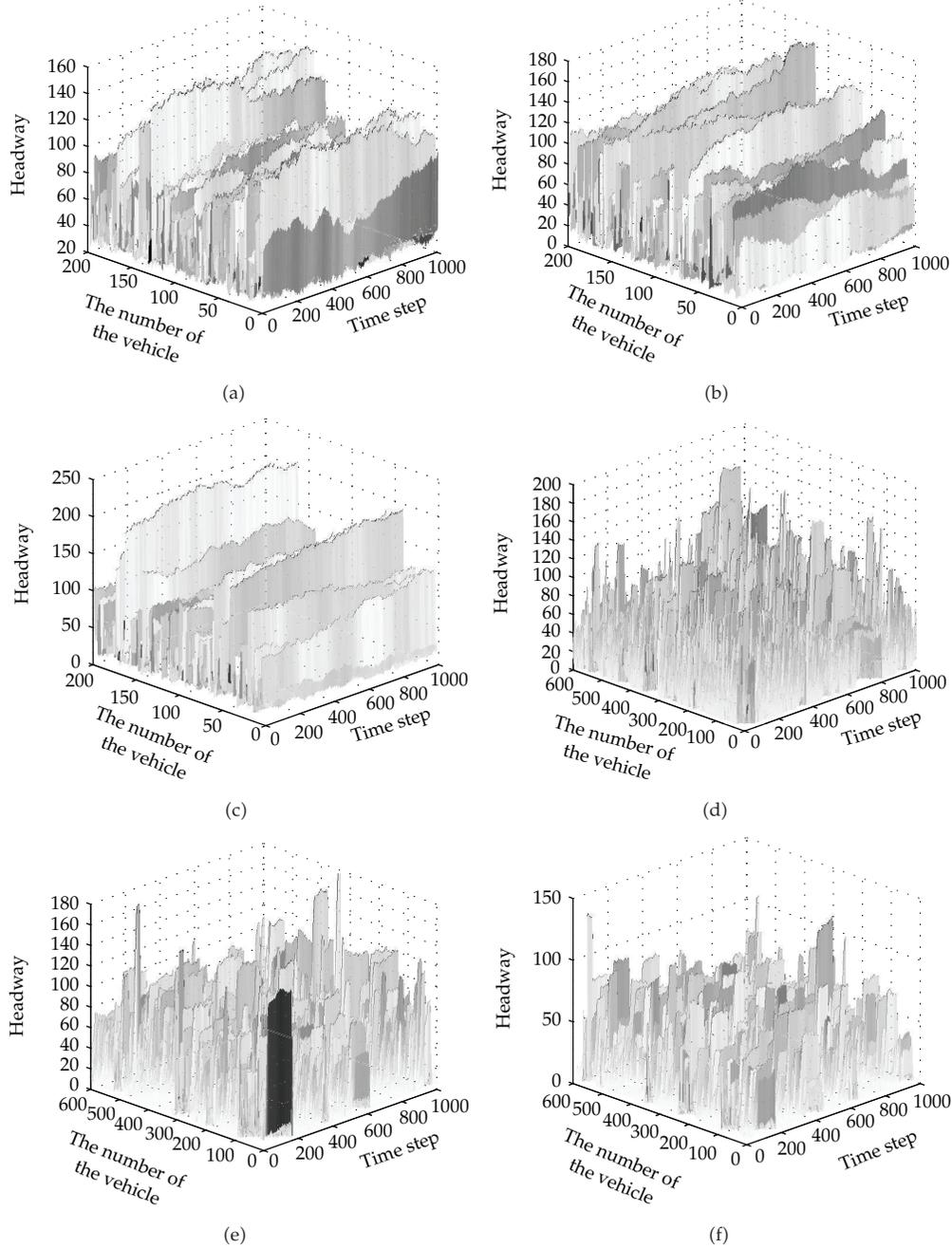


Figure 8: The headway fluctuations map in 1000 time steps under $P = 0.1$, $r_{ad} = 1$ (a) $\rho = 0.1$, $steps = 1$; (b) $\rho = 0.1$, $steps = 2$; (c) $\rho = 0.1$, $steps = 3$; (d) $\rho = 0.3$, $steps = 1$; (e) $\rho = 0.3$, $steps = 2$; (f) $\rho = 0.3$, $steps = 3$.

between -2 to 1 . In Figures 7(d), 7(e), and 7(f), under the density is higher ($\rho = 0.3$), when the state of $steps = 1$, the vehicle velocity fluctuations are serious and sometimes the fluctuation is up to -20 , that is to say, the velocity reduces from maximum 20 to 0 directly; at the same time, when the state of $steps > 1$, the vehicle can be maintained at a relatively smooth moving, and along with the increasing of $steps$, the vehicle velocity fluctuation amplitude decreases.

The above illustrate that the vehicle cannot decelerate abruptly if the drivers adjust their speeds in advance according to the size of the forecasted velocity, so that the traffic flow remains more stable, and the probability of traffic accident is smaller, which matches the actual traffic state.

Figure 8 shows the headway fluctuations in evolution process of 1000 steps. In Figures 8(a), 8(b), and 8(c), after a finite evolution, the results show that the headway is at a certain amplitude of the fluctuation and the fluctuation range is small when the traffic density is $\rho = 0.1$ and under the state of $steps = 1, 2,$ and 3 . In Figures 8(d), 8(e), and 8(f), after a finite evolution, the simulation results show that the headway is in the small range of relatively stable fluctuations when the traffic density is $\rho = 0.3$ and $steps > 1$, while the headway is in disorder substantial random fluctuations when the traffic density is $\rho = 0.3$ and $steps = 1$. All these illustrate that the vehicles can be more evenly distributed on the road if the drivers adjust their driving behavior in advance depending on the size of the predicted velocity, so that the traffic flow remains more stable and the probability of traffic accident is lower, which matches the actual traffic state.

4. Conclusion

From the views of time and space, this paper proposes the cellular automaton model based on the deceleration in advance. The model reflects the phenomenon in the actual traffic that drivers usually adjust the current velocity by forecasting its velocities in a short time of future. Computer simulations reproduce the metastable state, hysteresis, and phase separation phenomenon. After the observations of the density-flow relationship, the stability of traffic flow, the efficiency of blockage ablation, the velocity fluctuation, and the headway fluctuations, it is found that the vehicles will not suddenly decelerate and can be relatively even-distributed on the roads, the blockage caused by emergencies can be eliminated in a shorter time, the utilization rate of the road resources is higher, the traffic flow is greater; which better match the actual traffic state. All these support the view that the drivers who make the velocities prejudgment and fully respond to the actual traffic conditions in advance have an influence on traffic flow.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (Grant no. 50968009), the National Natural Science Foundation of China (Grant no. 71261014), and the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant no. 20096204110003).

References

- [1] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *Journal de Physique I*, vol. 2, no. 2, pp. 2221–2229, 1992.
- [2] S. C. Benjamin, N. F. Johnson, and P. M. Hui, "Cellular automata models of traffic flow along a highway containing a junction," *Journal of Physics A*, vol. 29, no. 12, pp. 3119–3127, 1996.
- [3] R. Barlovic, L. Santen, A. Schadschneider, and M. Schreckenberg, "Metastable states in cellular automata for traffic flow," *European Physical Journal B*, vol. 5, no. 3, pp. 793–800, 1998.
- [4] X. Li, Q. Wu, and R. Jiang, "Cellular automaton model considering the velocity effect of a car on the successive car," *Physical Review E*, vol. 64, no. 6, Article ID 066128, 2001.

- [5] S. X. Hu, K. Gao, B. H. Wang, and Y. F. Lu, "Cellular automaton model considering headway-distance effect," *Chinese Physics B*, vol. 17, no. 5, pp. 1863–1868, 2008.
- [6] Z. Qian, J. Bin, and L. Xin-Gang, "A modified weighted probabilistic cellular automaton traffic flow model," *Chinese Physics B*, vol. 18, no. 8, pp. 3271–3278, 2009.
- [7] J. X. Ding, H. J. Huang, and Q. Tian, "A traffic flow cellular automaton model to considering drivers' learning and forgetting behaviour," *Chinese Physics B*, vol. 20, no. 2, Article ID 028901, 11 pages, 2011.
- [8] H. D. He, W. Z. Lu, and L. Y. Dong, "An improved cellular automaton model considering the effect of traffic lights and driving behaviour," *Chinese Physics B*, vol. 20, no. 4, Article ID 040514, 7 pages, 2011.
- [9] M. Zhao, D. H. Sun, and C. Tian, "Density waves in a lattice hydrodynamic traffic flow model with the anticipation effect," *Chinese Physics B*, vol. 21, no. 4, Article ID 048901, 6 pages, 2012.
- [10] H. M. Niu, "Determination of the skip-stop scheduling for a congested transit line by bilevel genetic algorithm," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1158–1167, 2011.
- [11] E. I. Vlahogianni, J. C. Golias, and I. C. Ziomas, "Traffic flow evolution effects to nitrogen dioxides predictability in large metropolitan areas," *Transportation Research Part D*, vol. 16, no. 4, pp. 273–280, 2011.
- [12] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research Part C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [13] W. Wang, H. Guo, Z. Gao, and H. Bubb, "Individual differences of pedestrian behaviour in midblock crosswalk and intersection," *International Journal of Crashworthiness*, vol. 16, no. 1, pp. 1–9, 2011.

Research Article

Research on Intersection Signal Switching Model under Emergency Situation

Li Zongping, Bai Wei, and Gan Mi

School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, Sichuan 610031, China

Correspondence should be addressed to Gan Mi, yyganmi@gmail.com

Received 30 May 2012; Accepted 2 September 2012

Academic Editor: Huimin Niu

Copyright © 2012 Li Zongping 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 frequent occurrence of the city emergency leads to rapid development of emergency traffic management which is an important part in the Emergency Rescue System. In the intersection with heavy traffic, the emergency rescue vehicles often with increased delay, reduced safety, and sometimes even with a collision, which restrict the efficiency of the rescue. This paper established the intersection signal control optimization model based on detail analysis of emergency rescue vehicles traffic characteristics and traffic signal control. The models, on one hand, were able to guarantee the emergency vehicles through the intersection quickly and without delay; on the other hand, could ensure the minimum impact to other vehicles in the process of the emergency vehicles through the intersection. Finally, the model's practicality was verified by real cases.

1. Introduction

In modern life, the emergency events, natural disasters, and the non resistance events happen frequently. Moreover, these events are unpredictable and bring to several damages. Statistics show that every year our country has big loss because of the disasters (such as the natural disasters, accidents, the social and public events), and the loss caused by rescue that does not come in time increased year by year [1]. Therefore, the accurate and reliable traffic control measures are an effective way to reduce the response time of the emergency vehicle, improve the efficiency of the emergency rescue, and reduce the loss of lives and property. But in the process of the emergency response, the emergency vehicle often with increased delay, reduced safety, and even with a collision in intersections. According to the results of the emergency vehicle accident investigation in Virginia, about 31% accidents of the emergency vehicles' happened in the intersection. Moreover, the consequences are even more serious than happened in other places [2]. Therefore, the study on signal optimization control under emergency situations has an important significance in emergency rescue, traffic safety, and traffic management.

For the research on the right-of-way of the emergency vehicle, the scholars in transportation area in different countries focused on different aspects. Bachelder and Foster proposed the emergency vehicles preferred system based on the induction coils [3]. Mussa and Selekwa raised one conversion process of the traffic signal timing based on quadratic optimization method [4]. Nelson and Bullock pointed out that the signal control should back to normal operation rapid and safely after the emergency vehicles through the intersection [5], but the conversion mode of the signal control they proposed back to normal often used the ways of smooth, add only, or dwell, all of which are set according to the experience and are hard to adapt to the change of complex traffic condition.

In recent years, as the development of intelligent agent technology, many domestic scholars put forward the methods to solve the traffic problem based on this innovative, real-time and intelligent technology. Cao fulu (Chang'an University) proposed a kind of urban traffic signal optimization control system for the optimization of signal control through the coordination of the intelligence agents between adjacent intersections. Cheng Xiangjun (Beijing Jiaotong University) presented one distributed traffic signal coordination control method based on the multiagent technology, this method makes the vehicles through the intersection without delay under the higher utilization of the each phase in crossing traffic time.

We can conclude from the above analysis that the current study is mainly concentrated in the hardware and system design of the signal priority. The purpose of emergency signal optimization control can be expressed as one objective function, and this function subjects to some restrictions. We are able to obtain the optimal control strategy by the mathematical programming model and various kinds of optimization methods.

2. Signal Conversion Control

2.1. Control Process

This paper chooses the four phase signals cross-intersection as the research object. The signal optimization control measures under emergency situations are built according to the traffic conditions when the emergency vehicle's detector detects the request vehicles. The control process as in Figure 1.

2.2. Vehicle Detector Position Setting

2.2.1. General Vehicle Detector Position Setting

As shown in Figure 2, the position of the general vehicle detector can be expressed as

$$L_{VD} = L_{QD} + L_{DD}, \quad (2.1)$$

where, L_{VD} is the setting distance of other vehicle's detector; L_{QD} is the length of the other queuing vehicles under peak traffic condition; L_{DD} is the length of the slowly moving vehicles under the peak traffic condition.

To ensure the emergency vehicle could pass through the intersection safely and without delay, we must make sure the other vehicles in front of the emergency vehicle dissipated in time. L_{QD} can be determined by the average value of multiple times

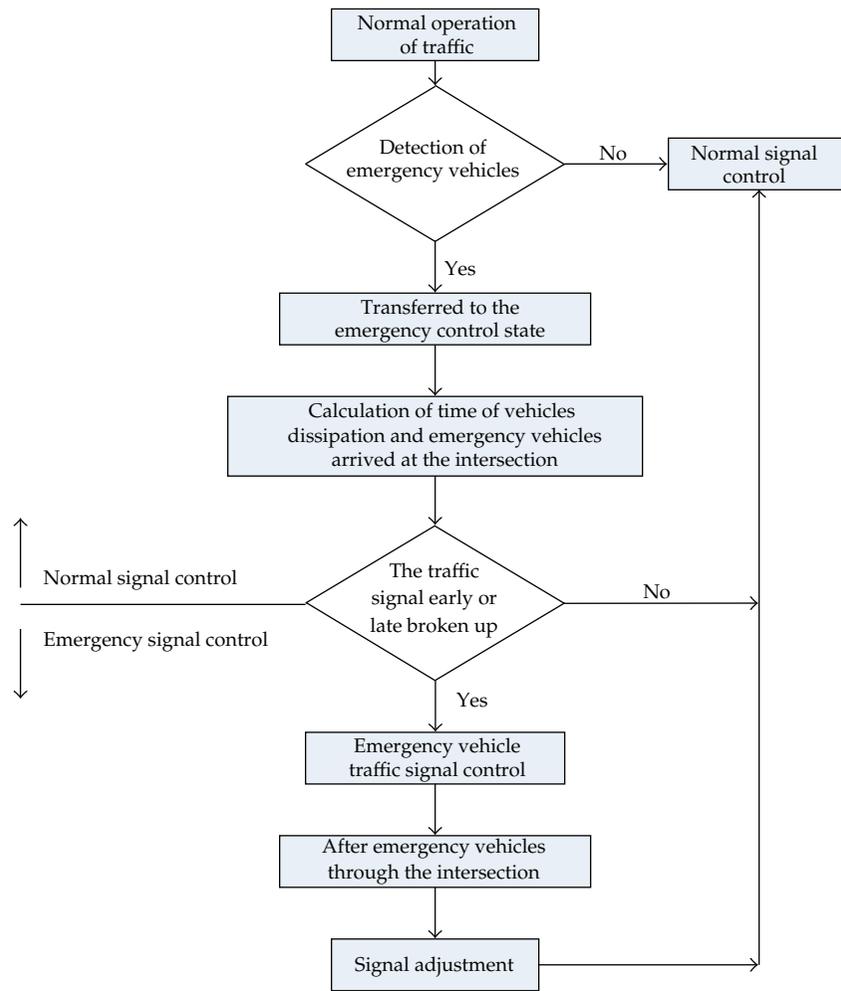


Figure 1: Intersection signal conversion control process under emergency conditions.

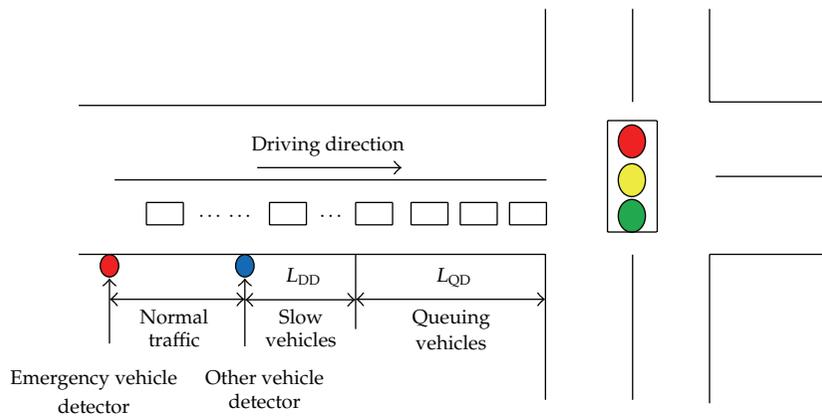


Figure 2: Classification of the traffic vehicle.

measurements. L_{DD} is difficult to detect, it can be replaced by the sum of the mean value and two times standard deviation of L_{QD} [6], as follows:

$$L_{QD} = \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i, \quad i = 1, 2, 3, \dots, n, \quad (2.2)$$

$$L_{DD} = L_{QD} + 2\sigma = \bar{X} + 2\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2}.$$

Then, x_i is the detected length of the other queuing vehicles under peak traffic condition; \bar{X} is the average queue length of the other vehicles; σ is the standard deviation of the other vehicle queue length.

2.2.2. Emergency Vehicle Detector Position Setting

The setting distance of the emergency vehicle detector should be able to guarantee that it has enough time to dissipate the vehicles before the emergency vehicle arrived at the intersection. So, there exists

$$t_{ED} = \frac{L}{v_{ED}}, \quad (2.3)$$

where t_{ED} is the ideal travel time of emergency vehicle from the detected position to the intersection parking line; L is the distance between the emergency vehicle detector and intersection; v_{ED} is the ideal average speed of the emergency vehicle.

In actual situation, the emergency vehicle could reach the intersection in any time. Assume the driving direction is already known, according to the different signal phase condition when the signal control machine receives the detected information from the emergency vehicle, the location of the emergency vehicle's detector can be set as follows.

(1) The signal phase of the emergency vehicle's driving direction is green:

Then, the disappearing time of the other vehicles is:

$$t_{CL} = \frac{(L_{QD} + L_{DD})}{v_{QD}}, \quad (2.4)$$

where t_{CL} is the disappearing time of the other vehicles before the emergency vehicle appears; v_{QD} is the dissipation speed of the other vehicle's queue in front of the emergency vehicle.

There should be a certain safety time interval between the last dissipated vehicle and the emergency vehicle. The formula is as follows:

$$t_{ED} - t_{CL} \geq t_{SI}, \quad (2.5)$$

where t_{SI} is the time interval between the last dissipated vehicle and the emergency vehicle.

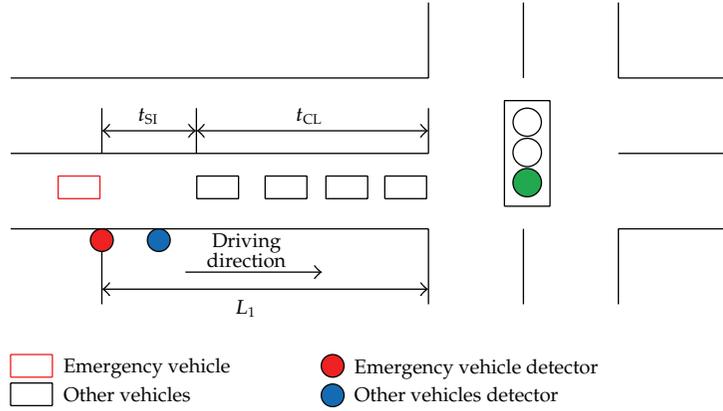


Figure 3: The location of emergency vehicle detector (green).

So the minimum distance of the emergency vehicle’s detector in this condition is

$$L_1 \geq \left[t_{SI} + \frac{(L_{QD} + L_{DD})}{v_{QD}} \right] \times v_{ED}. \tag{2.6}$$

(2) The signal phase of the emergency vehicle’s driving direction is yellow.

According to the conduct process we proposed above, the minimum distance of the emergency vehicle’s detector in this condition is

$$L_2 \geq \left[t_{SI} + t_{SL} + \frac{(L_{QD} + L_{DD})}{v_{QD}} \right] \times v_{ED}, \tag{2.7}$$

where t_{SL} is the loss of the start time (see Figure 4).

(3) The signal phase of the emergency vehicle’s driving direction is red.

The minimum distance of the emergency vehicle’s detector in this condition is in Figure 3

$$L'_3 \geq \left[t_{SI} + t_{SL} + \frac{(L_{QD} + L_{DD})}{v_{QD}} + t_{AR} - t'_{AR} \right] \times v_{ED}, \tag{2.8}$$

where t_{AR} is the all red light duration of this phase; t'_{AR} is the faded signal time of this red phase.

In practical situations, it is not easy to ensure t'_{AR} , which could be resolved by use the “real position” and the “virtual position”. The “real position” is the actual position of the emergency vehicle detector. The formula is as follows:

$$L_3 \geq \left[t_{SI} + t_{SL} + \frac{(L_{QD} + L_{DD})}{v_{QD}} + t_{AR} \right] \times v_{ED}. \tag{2.9}$$

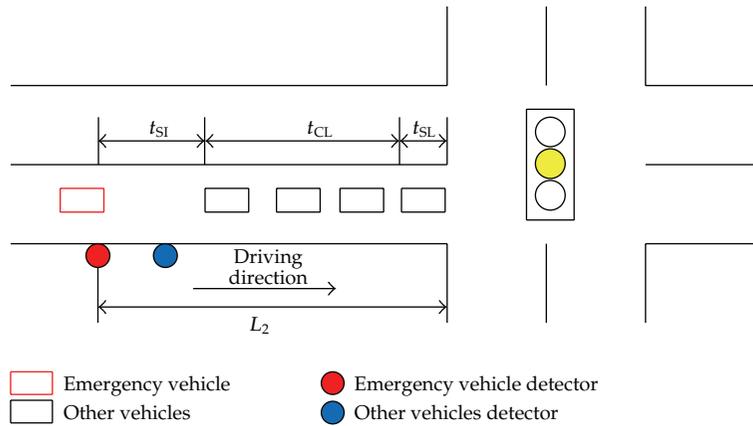


Figure 4: The location of emergency vehicle detector (yellow).

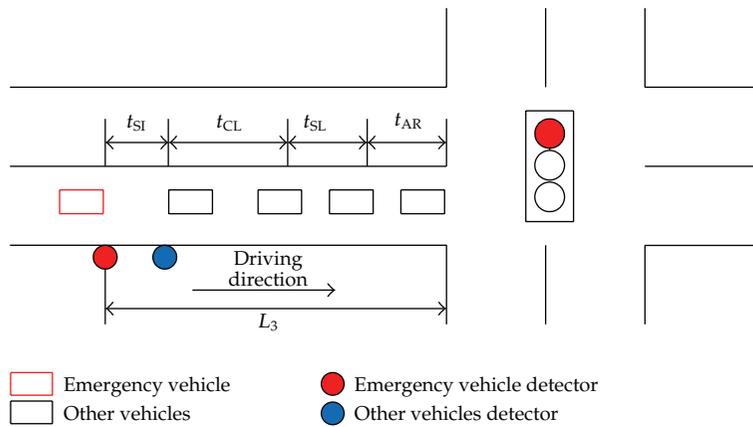


Figure 5: The location of emergency vehicle detector (red).

The “virtual position” is the position set by L_3' . In this condition, we should deduct the setting position of emergency rescue vehicle’s detector during the red signal phase. So there exists: $L_3 \geq L_3'$ (see Figure 5).

Therefore, the minimum distance of the emergency vehicle detector is:

$$L = \max\{L_1, L_2, L_3\} = L_3. \quad (2.10)$$

2.3. Optimal Control

At first, divide the entire process into three periods: detection period, before, and after the intersection period.

2.3.1. Detection Period

After the distance of the emergency vehicle detector is determined, the information about emergency vehicle, such as the location, speed, and the direction can be determined. The formula as follows [7–9]:

$$\begin{aligned}
 t_{HG} &= t'_{ED} - (t_{SW} + t'_{CL} + t_{SI}), \\
 t'_{ED} &= \frac{L}{v'_{ED}}, \\
 t'_{CL} &= t_{SL} + \frac{(L'_{QD} + L'_{DD})}{v_{QD}}, \\
 t'_{ED} &> t'_{CL},
 \end{aligned} \tag{2.11}$$

where t_{HG} is the green time that other phases obtained; t_{SW} is the loss time of conversion; t'_{CL} is the detected time of the dissipated vehicles in front of the emergency vehicle; t'_{ED} is the time from the position that emergency vehicle is detected to the stop line; v'_{ED} is the detected travel speed of the emergency vehicle; L'_{QD} is the actual queue length of other vehicles in emergency vehicle driving direction; L'_{DD} is the detected length of the slowly moving vehicles in emergency vehicle driving direction.

Once the control machine request received the request of the emergency vehicle, we could decide whether to carry out the signal priority control by the formulas we proposed above according to the signal phase and fixed signal timing cycle.

2.3.2. The Optimization before the Vehicles through the Intersection

The optimization purpose of this period is

- (1) to guarantee the emergency vehicle safety and without delay;
- (2) to minimize the influence to other vehicles.

According to the actual situation and with above optimization goals, we could implement optimization methods by different control measures, which can be specified as [10–13]

(1) The signal phase in emergency vehicle's driving direction is green when the detector detected emergency vehicle to ensure the safety when the emergency vehicles passes through the intersection. We do not consider the emergency vehicles through the intersection in the yellow signal phase.

- (i) Assume $t_{GR} > t'_{ED} > t'_{CL}$.

The signal control plan is still running as normal. t_{GR} is the remaining signal time of this green phase in emergency vehicle's driving direction when the detector detected emergency vehicle; t_{AG} is the full signal time of this green phase in emergency vehicle's driving direction.

- (ii) Assume $t'_{ED} \geq t_{GR} > t'_{CL}$.

The signal control plan is changed according to the difference between the surplus green time of this phase and the required time for emergency vehicles arrived.

(a) If $t'_{ED} - t_{GR} - t_{SW} - t_{SI} \geq T$, the signal control will be switched to normal, and the next phase can get the time for green is $t_{HG} = t'_{ED} - t_{GR} - t_{SW} - t_{SI}$. After t_{HG} , convert the green signal phase to the emergency vehicle's driving direction.

(b) If $t'_{ED} - t_{GR} - t_{SW} - t_{SI} < T$, extend the show time of the green light phase, there exist $t_{DG} = t'_{ED} - t_{GR} + T_{EI}$. Where, T is the time of the static car needed to through the intersection as usual. t_{DG} is the extended time of the green signal phase in the emergency vehicle's driving direction. T_{EI} is the time of the emergency vehicle needed to through the intersection.

(iii) If $t'_{ED} > t'_{CL} > t_{GR}$, the signal control plan is changed according to the difference between the surplus green time of this phase and the required time for emergency vehicles arrived also.

(a) If $t'_{ED} - t'_{CL} - t_{SW} - t_{SI} \geq T$, the signal control will be switched as normal, and the next phase can get the time for green is $t_{HG} = t'_{ED} - t'_{CL} - t_{SW} - t_{SI}$. After t_{HG} , convert the green signal phase to the emergency vehicle's driving direction.

(b) If $t'_{ED} - t'_{CL} - t_{SW} - t_{SI} < T$, extend the time of the green signal phase and the extended time is $t_{DG} = t'_{ED} - t_{GR} + T_{EI}$.

(2) The signal phase in emergency vehicle's driving direction is yellow when the detector detected emergency vehicle.

(i) Assume $t'_{ED} - t'_{CL} - t_{SW} - t_{SI} + t_{YR} \geq T$.

The signal control machine will be switched as normal, and the next phase can get the time for green is $t_{HG} = t'_{ED} - t'_{CL} - t_{SW} - t_{SI} - t_{YR}$. Where, t_{YR} is the remaining signal time of this yellow phase in emergency vehicle's driving direction when the detector detected emergency vehicle; t_{AY} is the full signal time of this yellow phase in emergency vehicle's driving direction, there exists $0 \leq t_{YR} \leq t_{AY}$. After t_{HG} , convert the green signal phase to the emergency vehicle's driving direction until the emergency vehicle passed the intersection.

(ii) Assume $t'_{ED} - t'_{CL} - t_{SW} - t_{SI} + t_{YR} < T$.

In this case, convert the green signal phase to the emergency vehicle's driving direction until the emergency vehicle passed the intersection.

(3) The signal phase in emergency vehicle's driving direction is red when the detector detected emergency vehicle.

(i) Assume $t'_{CL} < t'_{ED} < t_{RR}$.

The other phases can get the time for green is $t_{HG} = t'_{ED} - t'_{CL} - t_{SI}$. After that, convert the green signal phase to the emergency vehicle's driving direction and the break time of the red light in the emergency vehicle's driving direction is $t_{DT} = t_{RR} - (t'_{ED} - t'_{CL} - t_{SI}) + t_{SW}$. Where, t_{RR} is the remaining signal time of this red phase in emergency vehicle's driving direction when the detector detected emergency vehicle; t_{AR} is the full signal time of this red phase in emergency vehicle's driving direction.

(ii) Assume $t'_{ED} \geq t_{RR}$.

(a) If $t_{RR} \leq t'_{ED} - t'_{CL} - t'_{CL} \leq t_{RR} + t_{AG}$, the signal control machine will be switched as normal.

(b) If $t_{RR} + t_{AG} < t'_{ED} - t'_{CL} - t'_{CL}$, extend the time of the green signal phase and the extended time is $t_{DG} = t'_{ED} - t_{RR} - t_{AG} + T_{EI}$.

(c) If $t'_{ED} - t'_{CL} - t_{SI} < t_{RR}$, convert the green signal phase to the emergency vehicle's driving direction and the break time of the red light in the emergency vehicle's driving direction is $t_{DT} = t_{RR} - (t'_{ED} - t'_{CL} - t_{SI}) + t_{SW}$.

(4) Pedestrians scheme under the emergency situation.

Convert the pedestrian traffic light in emergency vehicle's driving direction to red until the emergency vehicle passed the intersection and make speakers or tips to notice to the crowd in the meantime.

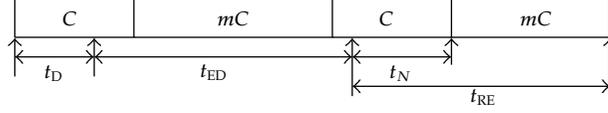


Figure 6: The priority control process of the emergency vehicle.

2.3.3. The Optimization after the Vehicles Passes through the Intersection

The control process of the emergency vehicle under the emergency situation can be described in Figure 6.

According to the Figure 6, the following relationships can be calculated:

$$t_N = C - \frac{(t_D + t_{ED})}{C}, \quad (2.12)$$

$$t_{RE} = t_N + nC - t_{SI},$$

where t_p is the time interval between the moment of the emergency vehicle detected and the beginning of next signal cycle; t_D is the time interval between the beginning of the signal cycle and the moment of the emergency vehicle detected; t_{RE} is the time length of the intersection signal optimization control after the emergency vehicles through the intersection; m, n is the signal cycle number.

In this period, establish the optimal model with the minimum delay of other vehicles' as the optimization target. Such that [14–18]

$$\begin{aligned} \min Z &= \min \sum_{l=1}^2 \sum_{j=1}^4 \sum_{k=1}^K \max \left\{ 0, Q_{jl}^i(k-1) + \sum_{i=1}^4 (q_{jl}^i(k) - \mu_{jl}^i \cdot \lambda_{jl}) \times g^i(k) \right\}, \\ \text{s.t.} \quad C(k) &= \sum_{i=1}^4 g^i(k), \\ g_{\min}^i &\leq g^i(k) \leq g_{\max}^i, \\ \sum_{k=1}^K C(k) &= t_{RE} = t_N + n \cdot C - t_{SI}, \\ Q(k) &= \sum_{l=1}^2 \sum_{j=1}^4 Q_{jl}^i(k) \leq Q_v, \\ Q_{jl}^i(k) &= \max \left\{ 0, Q_{jl}^{i-1} + VA_{jl}^i(k) - VD_{jl}^i(k) \right\}, \\ VA_{jl}^i(k) &= q_{jl}^i(k) \cdot g^i(k), \\ VD_{jl}^i &= \mu_{jl}^i \cdot \lambda_{jl} \cdot g^i(k), \\ Q_{jl}^0(k) &= Q_{jl}^4(k-1), \end{aligned} \quad (2.13)$$

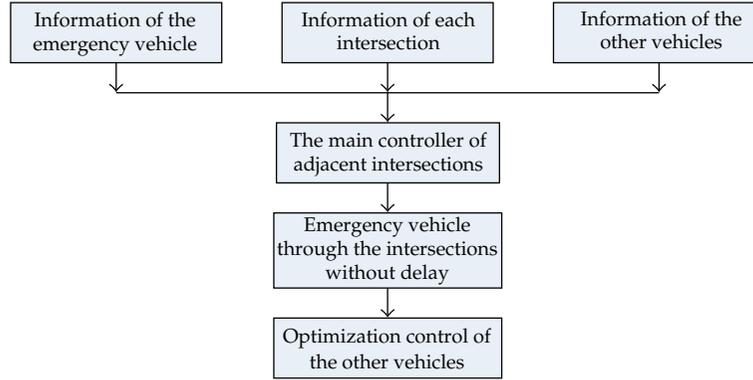


Figure 7: The process of the joint priority control.

where i is the number of the traffic signal phases, the maximum is 4; j is the driving direction of the intersection, 1, 2, 3, and 4 represent East, South, West and North; l is the direction of the lane, 1, 2 represent the direction of left and straight; k is the signal control cycle and K is the maximum number of it; $VA_{jl}^i(k)$ is the actual number of the reached vehicles in the i signal phase of k signal cycle and l lane of the j direction; $VD_{jl}^i(k)$ is the actual number of the past vehicles in the i signal phase of k signal cycle and l lane of the j direction; $Q_{jl}^i(k)$ is the number of queuing vehicles in the i signal phase of k signal cycle and l lane of the j direction; $q_{jl}^i(k)$ is the average arrival rate of the vehicles in the i signal phase of k signal cycle and l lane of the j direction; $g^i(k)$ is the duration of the effective green signal phase-in the i signal phase of k signal cycle; λ_{jl} is the saturated flow rate in the l lane of the j direction; T_{LS} is the loss time of each signal cycle; Q_v is the average queue length of each signal cycle under normal conditions; μ_{jl}^i is the traffic state of the l lane of the j direction in the i signal phase, if the car can move, $\mu_{jl}^i = 1$, else $\mu_{jl}^i = 0$.

The proposed optimization models could be solved with the genetic algorithm and the simulated annealing algorithm by applying the MATLAB tool.

2.3.4. The Control Strategy When through the Continuous Intersections

Take different control measures according to the traffic volume and the distance between the adjacent intersections.

If $L_N \geq L$, take the priority control independently for each intersection by the above method. Otherwise, take the joint priority control for the adjacent intersections. Where, L_N is the distance between the adjacent intersections. The process of the joint priority control is shown in Figure 7.

Set the controller of the largest vehicles' intersection as the main controller. The process of detection period and the optimization of the period before passed intersection are the same as analyzed in Section 2.3.2. After the emergency vehicle passed the intersections, take the optimization measures according to the sum of the vehicles in different directions in the adjacent intersections. Then, assign appropriate green signal time to the different directions of the adjacent intersections.

Table 1: Basic data of the experimental analysis.

DATA		Vehicles (vel/h)	Saturated flow rate per lane (vel/h)	Green time (s)
East	Straight	352	1450	31
	Left	228	1350	23
South	Straight	307	1400	28
	Left	231	1300	22
West	Straight	312	1450	31
	Left	219	1350	23
North	Straight	279	1400	28
	Left	201	1300	22

Table 2: The average length of the vehicles' queue.

		\bar{X} (m)	α (m)	$\bar{X} + \alpha$ (m)
East	Straight	99	20	139
	Left	64	17	98
South	Straight	87	35	157
	Left	62	26	114
West	Straight	81	27	135
	Left	70	14	98
North	Straight	71	35	141
	Left	53	21	95

Table 3: The setting distance of the detectors.

	Other vehicle's detector (m)	Emergency vehicle's detector (m)
East	139	618
South	157	679
West	135	606
North	141	623

Table 4: The second simulation of the optimization control (low traffic).

Average length of the queue (m)								The obtained time of green light (s)	The moment enter the intersection (s)
North		West		South		East			
Left	Straight	Left	Straight	Left	Straight	Left	Straight		
82	87	101	7	128	107	13	29	47	27
16	171	9	72	0	88	22	37	43	52
72	77	64	126	58	41	0	134	30	98
51	73	65	15	58	50	7	23	43	113

Table 5: The second simulation of the optimization control (high traffic).

Average length of the queue (m)								The obtained time of green light (s)	The moment enter the intersection (s)
North		West		South		East			
Left	Straight	Left	Straight	Left	Straight	Left	Straight		
21	136	87	58	7	116	7	122	44	41
23	73	57	72	27	58	0	80	53	67
35	14	81	117	56	23	7	71	42	86
57	63	0	101	92	95	14	95	35	101

Table 6: The optimization of the green light duration.

Phases	Phase 1	Phase 2	Phase 3	Phase 4
Green time (s)	33	57	21	28

Table 7: The length of the other vehicles' queue.

Length of the other vehicles' queue (m)							
North		West		South		East	
Left	Straight	Left	Straight	Left	Straight	Left	Straight
25	65	43	64	60	71	92	67

Table 8: The average delay of the other vehicles.

Traffic	Low				High	
Emergency signal priority control	Yes	No	Yes	No	Yes	No
Emergency vehicles' delay	0	21.2	0	51.9		
East						
Straight	63.4	43.9	71.9	55.7		
Left	30.7	37.1	35.6	45.8		
South						
Straight	51.2	40.6	75.9	48.1		
Left	59.4	45.5	63.8	54.9		
Average delay of other vehicles (m)						
West						
Straight	47.1	39.7	56.9	47.9		
Left	37.8	45.1	48.1	53.2		
North						
Straight	51.1	44.5	71.8	49.6		
Left	42.6	39.1	50.1	43.4		

3. Experimental Analysis

3.1. First Period

The selected experimental data and the average length of the vehicles' queue are shown in Tables 1 and 2.

The average speed of the dissipating vehicles is 3.3 m/s and the setting position of the detector are shown in Table 3.

3.2. Second Period

Select eight groups of data to simulate, as in Tables 4 and 5.

It is clearly shown in Tables 4 and 5 that all the emergency vehicles can pass through the intersection without delay and just cause little influence to other vehicles in most cases.

3.3. Third Period

Apply the first group data in Table 5 to calculate in MATLAB program, the results are shown in Table 6.

Apply the data in Table 6 to the simulation environment, the length of the other vehicles' queue are shown in Table 7.

Calculate the eight groups data, the average delay of emergency vehicle and other vehicles' are shown in Table 8.

4. Conclusion

This paper aims to establish the optimization model of the signal switching under emergency situations. On one hand, the models could guarantee the emergency vehicles are able to through the intersection safe and without delay; on the other hand, pass could ensure the minimum impact to other vehicles in the intersection during the processing of emergency vehicles pass through. Finally, real cases are employed to verify the practicality of the models. The models and algorithms can be also provided to some relevant management departments as the decision-making reference.

References

- [1] Y. Li and Yu. Zhang, "The emergency response law," *The Legislation. Administrative Law Research*, vol. 4, pp. 69–75, 2007.
- [2] National Highway Traffic Safety Administration, *Fatality Analysis Reporting System*, U.S. Department of Transportation, Washington. DC, USA, 2002.
- [3] A. Bachelder and C. Foster, System would Predictively Preempt Traffic Light for Emergency Vehicles, <http://www.nasatech.com/briefs/oct04/npo30573.html>.
- [4] R. Mussa and M. F. Selekwia, "Development of an optimal transition procedure for coordinated traffic signals," *Journal of Advances in Transportation Studies*, vol. 3, pp. 57–70, 2004.
- [5] E. J. Nelson and D. Bullock, "Impact of emergency vehicle preemption on signalized corridor operation: an evaluation," *Transportation Research Record*, no. 1727, pp. 1–11, 2000.
- [6] A. Mahdi, "Estimation of travel time reliability for freeways using mean and standard deviation of travel time," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, pp. 75–83, 2011.
- [7] G. Wei and G. Fang, "The bus priority signal control strategy research," *Journal of Systems Engineering*, vol. 16, no. 3, pp. 176–180, 2001.
- [8] Y. Xiaoguang, L. Yu, and H. Mingsheng, "signal control intersection bus priority signal determination method study," *China Highway*, vol. 14, no. 12, pp. 101–105, 2001.
- [9] Y. Xiao-Yong and Z. Tianwei, "Intersection light rail priority control strategy research," *City Public Traffic*, vol. 2, pp. 10–11, 2004.
- [10] H. C. Liu, J. Zhang, and D. X. Cheng, "Analytical approach to evaluating transit signal priority," *Journal of Transportation Systems Engineering and Information Technology*, vol. 8, no. 2, pp. 48–57, 2008.
- [11] P. Zhou, Z. K. Shi, and X. F. Chen, "Network control of urban traffic and its multi-objective optimal realization," *Control Theory and Applications*, vol. 19, no. 2, pp. 215–219, 2002.
- [12] Z. Zhongmin, *Road Traffic Organization Optimization*, People Traffic Press, 2004.
- [13] X. Hua and Z. Wenjun, "A consideration of internal state of emergency evacuation model," *The Computer Simulation*, vol. 22, no. 12, pp. 49–51, 2005.
- [14] H. Yanjun, *City public traffic road priority technologies and evaluation method [M.S. thesis]*, Southeast University, 2003.
- [15] W. H. Zhang, H. P. Lu, Q. Shi, and Q. Liu, "Optimal signal-planning method of intersections based on bus priority," *Journal of Traffic and Transportation Engineering*, vol. 4, no. 3, pp. 49–53, 2004.
- [16] X. F. Chen and Z. K. Shi, "Dynamic optimization method for traffic signal timings based on genetic algorithm," *Journal of System Simulation*, vol. 16, no. 6, pp. 1155–1157, 2004.

- [17] Y. Yiwu, D. Liyun D, and D. Shiqiang, "The intersections with left-turning traffic signal lights on vehicle delay effects," *Journal of Traffic and Transportation Engineering*, vol. 2, no. 4, pp. 80–85, 2002.
- [18] P. Qi, "Intersection signal timing optimization model research," *Journal of Shanghai Tiedao University*, vol. 20, no. 4, pp. 31–33, 1999.

Research Article

Study on the Road Network Connectivity Reliability of Valley City Based on Complex Network

**Yong-Sheng Qian,¹ Min Wang,¹ Hong-Xia Kang,²
Jun-Wei Zeng,¹ and Yu-Fei Liu¹**

¹ School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

² Taiyuan Railway Bureau Party School, Taiyuan 030013, China

Correspondence should be addressed to Yong-Sheng Qian, qianyongsheng@mail.lzjtu.cn

Received 8 May 2012; Revised 30 August 2012; Accepted 10 September 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Yong-Sheng Qian 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.

Based on the research progress in related fields and the distribution characteristics of road networks in valley cities, the complex network model of a city road network is established to study its connectivity reliability. Taking Lanzhou as the example, several parameters of the complex network abstracted from the road network are calculated and the practical meanings of them are described, respectively. On this basis, through computing the global efficiency and the relative size of the largest connecting subgraph under intentional attacks and random attacks, respectively, the curves of the above two parameters varying with the attacking times are drawn. The detailed investigation of connectivity reliability of Lanzhou road network is done by analyzing the curves' tendency. Finally, we find that the network of a valley city has a poor connection and has a lot of dead ends. Besides, the average length of the roads is very long and the holistic connectivity reliability is at a lower level; these are suitable to the group-type distribution of valley city's road network, and the connectivity reliability of the road network is stronger under random attacks than that under intentional attacks.

1. Introduction

Since the reform and opening up of China, the urbanization process has been accelerated greatly, which requires the urban road system to meet the requirements of the rapid development of cities. However, due to the limitations of historical and natural conditions, in many cities, especially the west valley cities, the phenomenon of urban road construction not meeting the requirements of urban development has appeared, which has brought inconveniences to the people's production and life travel and has led to a series of social

problems. Lanzhou, as the transportation hub and logistic center in northwest China, is suffering from particularly serious transport problems. Study on the reliability of urban road network will be able to provide theoretical supports for pertinently changing city traffic congestion.

2. Research Status in the Related Field

Complexity science is the further development, enrichment, and deepening of systematic science and nonlinear science, and is the latest and the most forward area of scientific research. The complex network method, developed in recent years, has provided a new perspective for the research of the complexity of a system. Two articles [1, 2] of complex network, published in *Nature* and *Science* in 1998 and 1999, respectively, pointed out that the connectivity distribution for many complex networks in real-world possesses exponential forms, and named such networks with scale-free property as scale-free networks. Since then, a boom of the study has been set off on complex network structure and its dynamical behavior, and the discussion on complex networks has penetrated into biology, physics, economics, computer science, as well as traffic transportation and other areas, which has improved the scientific understanding about the real world [3–9]. Jasny et al. used the complex theory to describe the connections in our life and science, he finds that the network analysis will help us to understand the world to a great degree [10]. Cho studied interactions between different people [11]. Bohannon studied the connections in some terrorist organizations, and helps the government to find out the leader of the terrorist organizations [12, 13]. With that, the characteristics and dynamic features of complexity networks are researched by many scholars [14–18].

Connectivity was first proposed by Mine and Kawai in 1982, which reflects the probability to maintain connectivity between nodes in transport networks. Asakura puts forward the concept of travel time reliability in 1991, which is another measure method of network reliability, fully taking into account the travel needs of the road network and traveler behavior [19]. In 1994, Nicholson and Du defined the recession reliability of the traffic flow [20], namely the probability for the decline of OD pair or the traffic flow in the network not exceeding a certain value, and pointed out that one or more road network OD pairs will be directly affected when recession occurs in one road section. Lam and Zhang proposed the conception of demanding satisfaction reliability in 2000, which reflects the ability of providing the potential travel demand for road network, and describes the probability for the travel demand satisfaction rate not less than a particular value [21]. In 2002, Bell and Schmoker expanded the travel time reliability to the travel cost reliability [22], including travel time, travel distance, vehicle charges, and public transport costs; if the cost for completing a trip is less than a given threshold value, the corresponding road section is believed to be reliable. In China, based on the growth rules of the simulating tree, Zhu et al. proposed the channel generation method [23], of which all channels are efficient and suitable for large-scale network, discussed the effective channels, sensitivity of road sections and the reliability; for the fallibility factors of the road traffic network, Jin explored the reliability of the evaluation system of urban road traffic networks, and researched the redundancy and the economy of time and space of road transport network from the viewpoint of dynamic degeneration [24]. In 2005, Lin et al. studied the reliability of transport networks in emergencies and gave out the three probability problems usually considered in unexpected situations: the destruction of properties of nodes and sections, repair characteristics of nodes and sections, and the connectivity reliability in emergencies

[25]. Fan et al. gave out the evaluation indexes of network reliability and the calculating model of reliability based on the network topology [26], and determined the arc, path, node, and the calculation method of reliability of the network structure; Yuan et al. considered the state of uncertainty of the transportation network under realistic traffic conditions [27], and proposed the change of choice of behavior from uncertain to deterministic network state and the route choice of model based on travel time reliability.

Through the analysis of the literature, we can see that the study of reliability of road network has transformed from the study only considering the physical structure of the road network to that of consideration of loading traffic flow and the mutual influence between travel demand and the capacity of road network, as well as the travel behavior of road users; the corresponding reliability indexes are connectivity reliability and travel time reliability. During this period, some scholars have proposed demand satisfaction reliability, the weak point reliability, performance index reliability, and suffered reliability. In China, research is conducted on the basis of foreign studies, and is the supplement or continuation of foreign study results; there is no significant breakthrough in the road network reliability indexes, and most of the researches are on connectivity reliability and travel time reliability. But domestic researches have explored the connectivity reliability and travel time reliability from different perspectives and conditions, which widen the thoughts and provide a foundation for further study on this area. There are also scholars that have studied the road network capacity reliability.

It can be seen from the results in the field that research on transport network reliability is increasing every year, which reflects the increasing travel demand in daily lives and higher requirements proposed by travelers on the travel system. In this paper, we studied the connectivity reliability of the valley city road networks using the theory of complex network.

3. Definition of Valley City and Its Urban Road Network Distribution Characteristics

From the perspective of geography, considering the location of the city and the relationship between the city and the valley, we define the valley city as follows: a city with the urban built-up area or the main or core part of the built-up area located in a valley, and the development of the main body of the city is strongly directly restricted by the valley terrain and the surrounding mountains or hills [27]. Because of the strong direct restriction by the valley terrain and the surrounding mountains or hills, the urban road network generally shows the banded distribution characteristic.

Compared with a plain city, the valley city road network presents its unique features. Firstly, with the limitation of the mountains and rivers, and in order to reduce the slope of the road, the valley city road lines are often built along the mountains or rivers, and road network form is always the freestyle. The valley city road lines are adapted to the natural terrain in order to save road engineering cost. Most of the road is not a straight line, and the road network is irregular. The valley city road network makes it difficult to form the ring road. Another big characteristic of a valley city is that the transportation is mainly concentrated on the main road of the urban axis. With the influence of the banded terrain and the problems of shortage of land, there is not enough land space to expand the urban road and share the traffic pressure when the urban trunk road is saturated. To research the connectivity reliability of valley city road network we will provide the reference to urban road network design and construction.

4. The Construction of the Complexity Network of Valley City Road Network

The complexity network can be described accurately using the language and symbols of the graph theory. The findings and techniques in graph theory about network have been widely transplanted into complexity networks. Based on this, the urban road network model has been established based on complexity network.

For the urban road network, intersections of the city's road network can be abstracted as nodes, and road sections connecting the intersections can be abstracted as edges in the network. Therefore, the city's road network can be abstracted as the complex network composed by intersection and road sections connecting the intersections.

The following assumptions have been made when constructing the complex network to simplify the issue.

(1) If from intersection A in the urban road network, we can get to intersection B through a certain road section, and through the same road section, we can get to intersection A from intersection B, we consider the above connections as the same one, namely, implementing undirected treatment on the network.

(2) Ignore the actual length of the road sections that connect nodes of the urban road network. Assume that length of road sections are all 1, then the distance between nodes can be expressed as the number of road sections between the nodes, which do not consider the weights of the connections, and abstract the urban road network as a nonweighted network.

(3) In this work, residential districts at road sections, river crossing transport facilities, and ends of roads are considered as network nodes, which are equivalent to intersections. At the same time, if there is more than one road section connecting two nodes, then consider these sections as one connection, which is only one side between the two nodes. Consider road sections connecting two intersections directly in a reality network as sides with length = 1 in the model; from any intersection in the network, we can get to any other intersection, namely, the entire network is connected.

5. Study on the Road Network Connectivity Reliability of Lanzhou

5.1. Definition of Urban Road Network Connectivity Reliability under Complex Network Theory

Based on the definition of the system reliability and the temporal and spatial imbalance of the urban transport travel, this paper presents the general definition of the urban road network reliability as: the statistical probability for the urban road network meeting a particular traffic travel demand under different spatial and temporal distribution.

Urban road network connectivity reliability based on the complex network is defined as follows: the ability for urban road network to maintain the connectivity state, after being attacked in different ways and the network suffering from a certain level of destruction.

5.2. Instructions of Parameters in Valley City Road Network Model

5.2.1. Degree and Degree Distribution

The degree k_i of intersection (node) i is the total number of the road sections connecting intersection i ; the average value of the degrees of all intersections is defined as the average

value of the urban road network degree; the degree distribution is indicated with function $p(k)$, reflecting the probability for a random intersection being connected with k road sections. The degree of the urban road network is the number of intersections connecting an intersection directly. The degree distribution reflects the maturity degree of the urban road network; the high degree distribution indicates that the urban road network has high network rate and less dead ends, and relatively high reliability.

5.2.2. Average Path Length

In complex networks, the distance d_{ij} between nodes i and j is defined as the minimum number of the sides connecting the two nodes; the network diameter is defined as the largest distance of two random nodes; the average value of the distances of all the node pairs of the complex networks is marked with L , which reflects the degree of separation between nodes and can be calculated by the following formula:

$$L = \frac{1}{1/2N(N+1)} \sum_{i \geq j} d_{ij}, \quad (5.1)$$

where N is the node number. If we ignore the distance between node i and itself because it is zero, the above equation becomes:

$$L = \frac{1}{1/2N(N-1)} \sum_{i > j} d_{ij}. \quad (5.2)$$

The path length of urban road network is the distance from one intersection to another and the average path length L refers to the average value of the shortest path length between all intersection pairs. The small L indicates a small distance between any two intersections in the road network, namely, good overall road network reachability.

5.2.3. Clustering Coefficient

The clustering coefficient is the characteristic parameter used to describe the tightness of the network, marked with C . A node with large clustering coefficient indicates that the tightness level of the adjacent nodes is high, and it is easy to form regional agglomeration for the subnetwork composed by this node and its neighboring nodes. It is calculated as follows: assume node i in the network is connected with k_i nodes through k_i sides (if the k_i nodes are connected to each other, then there should be $1/2k_i(k_i - 1)$ sides) and suppose that there are in fact only E_i sides connected to each other within k_i nodes, then the ratio between E_i and $1/2k_i(k_i - 1)$ is the Clustering coefficient, which is:

$$C_i = \frac{2E_i}{k_i(k_i - 1)}. \quad (5.3)$$

The entire network's clustering coefficient C is the average clustering coefficient of all nodes throughout the network, which is:

$$C = \frac{1}{N} \sum_i C_i. \quad (5.4)$$

The clustering coefficient of the urban road network reflects the aggregation of an intersection and its k_i adjacent intersections. The clustering coefficient is indicated with the ratio between the number of sections connecting the k_i intersections and the probable total number of those sections. Clustering coefficient reflects the aggregation between an intersection and its k_i adjacent intersections, and the great clustering coefficient represents the high level of short-range contact between intersections of the road network, namely, great density.

5.2.4. Efficiency of the Network

The efficiency of the network is qualified to analyze the network of small-world behavior instead of the average path length and clustering coefficient. The efficiency ε_{ij} between nodes i and j of the network is the reciprocal value of the distance d_{ij} between the two points, namely, $\varepsilon_{ij} = 1/d_{ij}$. If it is unreachable between nodes i and j , then d_{ij} tends to zero, and the corresponding $\varepsilon_{ij} = 0$. For the entire network, the average value of efficiencies between all node pairs is defined as the global efficiency, denoted with E_{glob} :

$$E_{\text{glob}}(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \varepsilon_{ij} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}. \quad (5.5)$$

Taking into account the situation of nonconnected graph, the local features of the network can be represented by the average efficiency of the partial subgraph G_i (all nodes connecting with node i). $E(G_i)$ can be used to analyze the effectiveness of the information transmission between node i 's adjacent nodes. E_{loc} is corresponding to the clustering coefficient C . Consider the following:

$$E(G_i) = \frac{1}{k_i(k_i-1)} \sum_{i \neq m \in G_i} \frac{1}{d_{im}}, \quad (5.6)$$

$$E_{\text{loc}}(G) = \frac{1}{N} \sum_{i \in G} E(G_i).$$

For the urban road network, G_i is the subgraph formed by the k_i intersections adjacent to intersection i , the maximum number of road sections is $1/2k_i(k_i-1)$, and d_{lm} is the length of the shortest path between nodes l and m of subgraph G_i . The efficiency of the urban road network reflects the connectivity, that is, the density of the intersections.

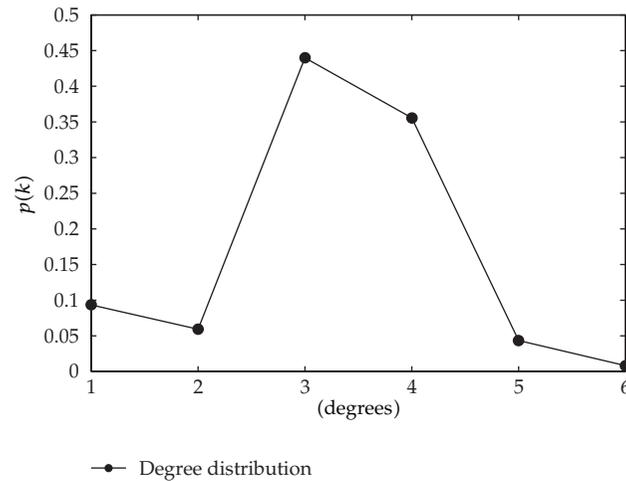


Figure 1: Degree distribution of Lanzhou City road network.

5.2.5. The Relative Size of the Largest Connected Subgraph

The largest connected subgraph of the network means the subgraph connecting all nodes with least sides. The relative size S of the largest connected subgraph is an important amount, weighing the capacity maintaining the original function of a network when being attacked continuously, which is the ratio of the total node number of the largest connected subgraph N' and total node number of the network N :

$$S = \frac{N'}{N}. \quad (5.7)$$

The relative size of the largest connected subgraph weighs the capacity maintaining the original function of a network after destruction.

5.3. Static Statistical Eigenvalues of the Road Network in Lanzhou City

In the study, the existing nodes and road sections of Lanzhou city were selected (a total of 349 selected intersections and 507 road sections). Calculating the degree and degree distribution of Lanzhou city road network, the average path length and the clustering coefficient is helpful to studying the topological properties of the Lanzhou city road network. Analyzing these statistical eigenvalues has a practical significance for learning the valley city road network and its connectivity reliability.

5.3.1. Degree and Degree Distribution

For the Lanzhou city road network model, node degree reflects the importance of the node in the network; its practical significance is the number of road sections connected to the intersection in different directions. Figure 1 shows the calculation results of degree distribution of Lanzhou City road network.

It can be seen from Figure 1 that the largest degree value is 6, followed by the value of 5, and these nodes share relatively smaller proportions. The reason why there are intersections connecting 5-6 road sections in different directions is that in the process of the network model handling, two close intersections are considered as one, including a number of important road junctions and overpasses. Combined with the actual situation of Lanzhou City, several overpasses and the Yellow River Bridges are connected closely, so the above treatment is done when handling nodes corresponding to these places. These nodes, corresponding to the Xiaoxihu overpass and the Jiefangmen overpass, have the greatest degree values, indicating that they are the most complex intersections in the road transport network in Lanzhou City.

Nodes with a degree = 3 in the Lanzhou city road network account for more than 40% of the total nodes, indicating that junctions of three roads account for a large proportion in the Lanzhou city road network, this is because of the limitations of topographical, the trunk line distribution of Lanzhou City is usually along the rivers and mountains, and the secondary roads, slip roads connected to the trunk roads, which is very easy to form junctions of three roads. Nodes with a degree = 4 account for the second largest proportion. No matter what type of urban road network, common road intersections are the convergence of four directions, which meets the actual situation; the grid distribution and the irrational square format local road network in Lanzhou result in a lot of crossroads, whose corresponding node degree value is 4. Nodes with degrees = 1, 2 account for about 20% of the total nodes, and the number of nodes with a degree = 1 is more than that with a degree = 2, which equates with the situations of Lanzhou. Because the road network is not perfect in Lanzhou, there are a lot of cul-de-sac with a corresponding degree = 1; in addition, city roads in Lanzhou show long strip distributions, and there are a lot of living areas distributed along the strip roads; the intersections of the living areas and the roads are considered as nodes with a degree = 2. The above analysis shows that the model results are consistent with the actual situation.

5.3.2. Average Path Length

The path length of the network represents the minimum number of road sections from a node to the specified node. Corresponding to the urban road network in Lanzhou, its practical significance is the number of road sections from one intersection to the specified intersection. Figure 2 is the statistical chart of the average path length of the Lanzhou road network. The figure shows that the proportion of intersections with the average node path length less than or equal to 8 is less than 36%, indicating that the average path length of the network is relatively large, and for most of the intersections, more than 8 road sections need to be passed to get to the other intersection in the network. The largest average path length of Lanzhou city road network is 27, indicating that moving from any intersection to another, the maximum average number of road sections to pass is 27. However, since the assumption of the length of all the road sections in the network model is 1, the calculation results are different from the actual length while the number of sections to go through matches the actual. In this article, the statistics data we applied is the average path length of the road network, but for the actual situation, the path length between the intersections at the edge of the network will exceed 27 because the intersections at the edge are isolated from other intersections. In addition, the average path length of the urban road network in Lanzhou is 21.9, that is, there are 21-22 sections from one intersection to another in Lanzhou city road network. The large average

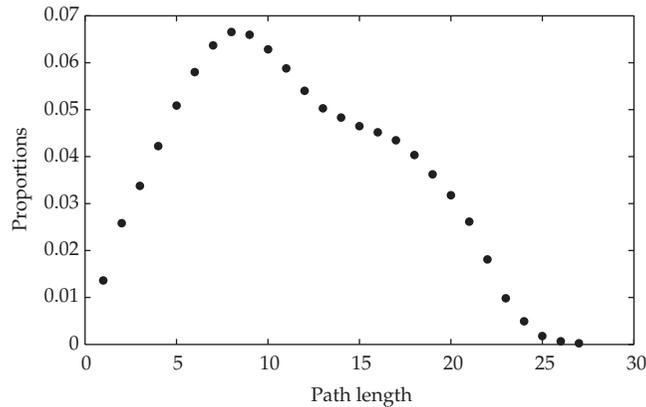


Figure 2: Statistical chart of the average path length of Lanzhou.

path length of the network has a direct relationship with the long strip and cluster layout of the Lanzhou road network, and the calculated results are consisted with the actual situation.

5.3.3. Cluster Coefficient

Another important statistical characteristic parameter of complex network is the clustering coefficient, which is a physical quantity measuring the density of the network between adjacent nodes. The practical significance of the clustering coefficient corresponding to the road network of Lanzhou is the aggregation level of intersections in Lanzhou. Figure 3 is the statistical chart of the clustering coefficient of the road network in Lanzhou. It can be seen from Figure 3 that intersections with clustering coefficient = 0 account for 75.5% of the total, indicating that 75.5% of the intersections of Lanzhou are in a star-shaped framework, which the connections between the adjacent intersections must go across it. The number of nodes with clustering coefficient = 0 is zero, indicating that no road section between adjacent nodes can achieve straight-through in Lanzhou road network. The overall clustering coefficient of the whole road network of Lanzhou is only 0.0582, indicating a poor aggregation, which is consistent with the cluster and long strip distribution of the Lanzhou road network.

5.4. Analysis of the Connectivity Reliability of Lanzhou Road Network

Complex networks usually face two kinds of attacks: random attacks and deliberate attacks. As a small world network, the urban road network in Lanzhou is also facing these two attacks. This paper studies the changes of the connectivity reliability of Lanzhou road network under the two attacks. According to the actual situation of the urban road network in Lanzhou, the following specific assumptions were made for the two attack patterns.

5.4.1. Assumptions for Random Attacks

The random attack refers to nodes in the network being deleted randomly at a certain probability. Taking into account the relatively large number of intersections, each attack will

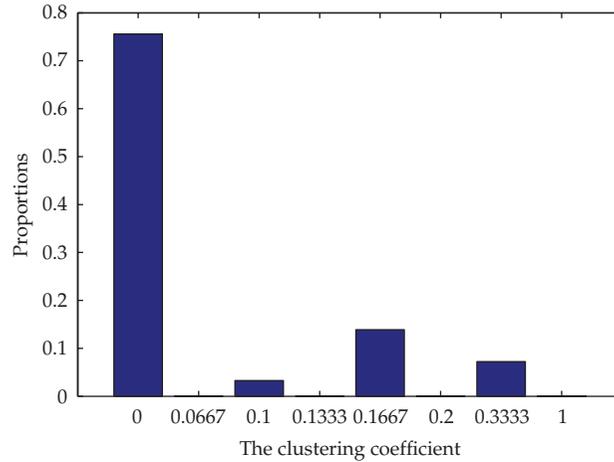


Figure 3: Statistical chart of the clustering coefficient of Lanzhou.

remove 5 nodes in the network, corresponding to 5 intersections of the urban road network in Lanzhou City.

5.4.2. Assumptions for Deliberate Attacks

The deliberate attack is the other attack mode the complex network suffered, which is a strategic attack on the nodes in the network. The purpose of this attack is to make the most severe damage with the least number of attacks. Therefore, this attack aims at and destroys the node with the greatest degree value of the network. In this work, the purpose of the deliberate attack is to make the most severe damage and make the road network of Lanzhou collapse as soon as possible. To make contrast with the random attack, 5 nodes with the greatest degree value will be removed from the network, corresponding to the 5 intersections with the most convergence.

5.4.3. Analysis of Results

Based on the above assumptions, the global efficiency E_{glob} and the relative size of the network's largest connected subgraph S under the two attacks are calculated in this paper. In order to more clearly describe the changes of the connectivity reliability of Lanzhou road network under different attack modes, the random attacks and deliberate attacks were compared, and the change curves of different indexes of the connectivity reliability based on the above data were formed. Specific analysis is as follows.

(1) Change curves of global efficiency E_{glob} of Lanzhou road network.

The global efficiency is an important indicator measuring the overall road network connectivity of Lanzhou. The calculated Lanzhou road network global efficiency is 0.118, indicating that the connectivity reliability of Lanzhou is poor.

Figure 4 shows the change curves of global efficiency under different attacks of Lanzhou road network. It can be seen from the figure that the global efficiency decreases faster under deliberate attacks compared with random attacks.

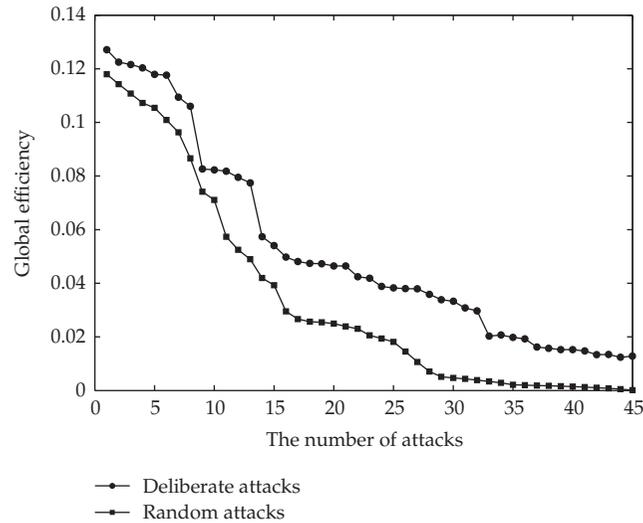


Figure 4: Change curves of global efficiency under different attacks.

Under deliberate attacks, the global efficiency will decrease to below 0.06, which is only half of the original value, after 11 attacks (the deleted nodes only accounting for 15.7% of the total nodes); the network global efficiency will decrease to 0.02 after 23 attacks (deleted nodes accounting for 33% of the total nodes), indicating that the connectivity of the network reliability has become very poor; after 45 attacks (the deleted nodes accounting for 64.5% of the total nodes), the global efficiency of the network will become 0.00013 (close to 0), which shows the road network has been near-collapse. Under random attacks, after 14 attacks (the deleted nodes only accounting for 20% of the total nodes), the global efficiency will decrease to below 0.06; after 35 attacks (the deleted nodes accounting for 50.1% of the total nodes), the global efficiency will decrease to 0.02, which means that the connectivity reliability will become poor when more than half of the nodes are attacked; after 45 attacks (the deleted node accounting for 64.5% of the total nodes), the global efficiency becomes 0.01279, which is much larger than that of the deliberate attack (0.00013), and until this time, the Lanzhou road network has not been crashed.

The reason for these results is that there are great differences in the deletion of nodes under the two attacks. The deliberate attack is carried out in accordance with the degree value of the nodes. After the nodes with relatively larger degree value of the network, which play a very important role in maintaining the road network connectivity reliability, are removed, dramatic changes will occur on the topology fabric of the road network, leading to a great number of isolated intersections, and ultimately the collapse of the entire network. However, for random attacks, the probability for removing nodes according to the order of deliberate attacks is very small, and the network will demonstrate stronger robustness than that of deliberate attacks when subjected to random attacks, so more isolated intersections are needed to make the same extent of damage as the random attack. The above analysis shows that the damage on the connectivity reliability of the Lanzhou road network subjected to random attacks is slighter than that of deliberate attacks.

(2) Change curves of the relative size S of the largest connected subgraph of Lanzhou road network.

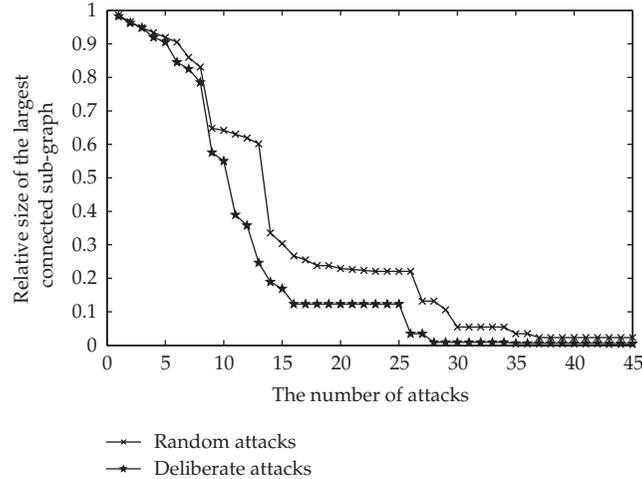


Figure 5: Change curves of relative size of the largest connected sub-graph under different attacks.

Figure 5 shows the change curves of the relative size of the largest connected subgraph of Lanzhou road network subjected to different attacks under different times. It can be seen from the figure that the relative size of the largest network connected subgraph of the Lanzhou road network decreases faster under deliberate attacks compared with random attacks.

Under deliberate attacks, the relative size S of the largest connected subgraph of the network decreases from 1 to 0.8252 with a small reducing extent and a slow speed after 7 attacks; S decreases sharply from 0.8252 to 0.1232 during the 7th to the 16th attack; in the subsequent attacks, the speed of S decrease is gradually slowing down, until the 45th attack when S is close to 0. The above process shows that during the first 7 attacks, little intersections are attacked and isolated, and most of the nodes are maintained in the largest connected subgraph, namely the connectivity reliability can be maintained at a relatively high level; in the next 7–16 attacks, the number of isolated intersections is gradually increasing, the number of nodes of the largest connected subgraph of the network is rapidly decreasing with the increasing number of attacks, the entire network has been divided into a number of relatively smaller networks with less nodes (maximum number of nodes in small networks is 12.32% of the original network), and the connectivity reliability has been very poor. In subsequent attacks, the subgraphs of Lanzhou road network are further divided, and the number of isolated intersections keeps increasing, until the entire network collapses. As for random attacks, the changes of the relative size of the largest connected subgraph show a similar trend as deliberate attacks; the difference is that the relative size of the largest connected subgraph under random attacks always maintains a higher level than that of deliberate attacks, and after the 45th attacks, it still remains at 0.0229, much larger than deliberate attacks of 0.0046. This process shows that, for random attacks, the connectivity reliability will degenerate as deliberate attacks; however, the network will maintain a certain level of connectivity reliability (the largest connected subgraph of the network maintains nearly 86% of the intersections of the original network) under less attacks (7 times); with the increase of the attack number, the intersections of the largest connected subgraph will be reduced greatly, and then will gradually level off.

The reason for these results is that when the network is subjected to deliberate attacks, the node removing is carried out in accordance with the degree value of the nodes; after the nodes with relatively larger degree value are removed during the early attacks, the network will be instantly differentiated into several subnets with sharply reduced nodes, so the relative size of the largest connected subgraph will change dramatically, the result of which is the maintenance of the degree of nodes left in the subgraphs at a very low level. When the network is subjected to random attacks, the deleting order is completely random, and the node with the largest degree value may remain in the largest connected subgraph until the end of the 45th attack; so in this case, the relative size of the largest connected subgraph may still have a relatively larger value compared with deliberate attacks. The above analysis shows that the connectivity reliability of the Lanzhou road network subjected to random attacks is better than that of deliberate attacks.

6. Conclusions

The paper constructs a complex network model of a valley city road network. Taking Lanzhou as an example of a typical valley city, the connectivity reliability of the road network was researched and the following conclusions are obtained.

(1) Valley city road network has poor connection and a lot of dead ends, and the number of road sections between intersections is too large, which makes it a poorly compacted network.

(2) Valley urban road network has long average path length and poor overall connectivity reliability, consistent with its cluster long strip layout; and the connectivity reliability of the urban road network subjected to random attacks is better than that of deliberate attacks, which is consistent with the characteristics of complex networks under attack.

Acknowledgments

This work is partly supported by the Humanities Social Sciences Programming Project of the Ministry of Education of China (no. 10YJA630126) and the State Social Science Fund Project (no. 11CJY067).

References

- [1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [2] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [3] D. Cui, Z. Y. Gao, and X. M. Zhao, "Cascades in small-world modular networks with 'CML's method," *Physica B*, vol. 17, pp. 1703–1710, 2007.
- [4] Z. Zhi, F. Zhong-Qian, and Y. Gang, "Synchronization speed of identical oscillators on community networks," *Chinese Physics B*, vol. 18, no. 6, pp. 2209–2212, 2009.
- [5] C. Di, G. Zi-You, and Z. Jian-Feng, "Properties of asymmetrically evolved community networks," *Chinese Physics B*, vol. 18, no. 2, pp. 516–521, 2009.
- [6] H. F. Du, S. Z. Li, W. F. Marcus, Z. S. Yue et al., "The topological railway network of China constrained by the geographic factors," *Acta Physica Sinica*, vol. 57, pp. 6771–6776, 2008 (Chinese).
- [7] J. Li, B. H. Wang, P. Q. Jiang, T. Zhou, and W. X. Wang, "Growing complex network model with acceleratively increasing number of nodes," *Acta Physica Sinica*, vol. 55, no. 8, pp. 4051–4057, 2006 (Chinese).

- [8] M. Zhao, B. H. Wang, P. Q. Jiang et al., "Recent advancement in research of synchronization of dynamical systems on complex networks," *Progress in Physics*, vol. 25, pp. 273–295, 2005 (Chinese).
- [9] Q. Luo, W. Wu, L. X. Li, Y. X. Yang, and H. P. Peng, "Adaptive synchronization research on the uncertain complex networks with time-delay," *Acta Physica Sinica*, vol. 57, no. 3, pp. 1529–1534, 2008 (Chinese).
- [10] B. R. Jasny, L. M. Zahn, and E. Marshall, "Connections," *Science*, vol. 325, no. 5939, p. 405, 2009.
- [11] A. Cho, "Ourselves and our interactions: the ultimate physics problem?" *Science*, vol. 325, no. 5939, pp. 406–408, 2009.
- [12] J. Bohannon, "Counterterrorism's new tool: 'Wletahetwork' analysis," *Science*, vol. 325, no. 5939, pp. 409–411, 2009.
- [13] J. Bohannon, "Investigating networks: the dark side," *Science*, pp. 410–411, 2009.
- [14] M. Kaiser and C. C. Hilgetag, "Optimal hierarchical modular topologies for producing limited sustained activation of neural networks," *Frontiers in Neuroinformatics*, vol. 4, p. 112, 2010.
- [15] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nature*, vol. 466, no. 7307, pp. 761–764, 2010.
- [16] M. Newman, "The physics of networks," *Physics Today*, vol. 61, no. 11, pp. 33–38, 2008.
- [17] A. L. Barabási, "Scale-free networks: a decade and beyond," *Science*, vol. 325, no. 5939, pp. 412–413, 2009.
- [18] D. Garlaschelli, A. Capocci, and G. Caldarelli, "Self-organized network evolution coupled to extremal dynamics," *Nature Physics*, vol. 3, no. 11, pp. 813–817, 2007.
- [19] Y. Asakura, "Reliability measures of an origin and destination pair in a deteriorated road network with variable flows," in *Proceedings of the 4th Meeting of the EURO Working Group in Transportation*, pp. 398–412, University of Newcastle, UK, 1996.
- [20] A. J. Nicholson and Z. P. Du, "Improving transportation system reliability: a framework," in *Proceedings of the 17th Australian Road Research Board Conference (ARRB '94)*, pp. 1–17, August 1994.
- [21] W. H. K. Lam and N. Zhang, "A new concept of travel demand satisfaction reliability for assessing road network performance," in *Proceedings of the Matsuyama Workshop on Transport Network Analysis*, no. 8, pp. 165–178, Matsuyama, Japan, 2000.
- [22] M. G. H. Bell and J.-D. Schmoker, "Public transport network reliability topological effects," in *Proceedings of the 3rd International Conference on Transportation and Traffic Studies*, Guanxi People's Press, Guilin, China, 2002.
- [23] S. Y. Zhu, W. Wang, W. Deng, Y. Tang, and B. Wang, "Research on traffic network reliability and access road algorithm," *China Journal of Highway and Transport*, vol. 13, no. 1, pp. 91–94 (Chinese).
- [24] J. Jin, "Reliability discussion on urban road network," *Journal of Transportation Systems Engineering and Information Technology*, vol. 2, p. 76, 2004 (Chinese).
- [25] H. Lin, C. Y. Wang, and L. Li, "Study on network travel time reliability under contingent events," *Journal of Engineering College of Armed Police Force*, vol. 21, p. 63, 2005 (Chinese).
- [26] H. Y. Fan, Z. Z. Wu, and X. G. Yang, "Research on the reliability of the topological structure for road transportation networks," *Journal of China University of Mining and Technology*, vol. 34, no. 4, pp. 482–485, 2005 (Chinese).
- [27] P. C. Yuan, Y. Han, and W. D. Ma, "Stochastic transportation network equilibrium model based on travel-time reliability," *Journal of University of Shanghai for Science and Technology*, vol. 30, no. 4, pp. 352–356, 2008 (Chinese).

Research Article

Modeling and Algorithms of the Crew Rostering Problem with Given Cycle on High-Speed Railway Lines

Zhiqiang Tian and Huimin Niu

School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

Correspondence should be addressed to Huimin Niu, hmniu@mail.lzjtu.cn

Received 18 June 2012; Revised 6 September 2012; Accepted 7 September 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Z. Tian and H. Niu. 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 studies the modeling and algorithms of crew roster problem with given cycle on high-speed railway lines. Two feasible compilation strategies for work out the crew rostering plan are discussed, and then an integrated compilation method is proposed in this paper to obtain a plan with relatively higher regularity in execution and lower crew members arranged. The process of plan making is divided into two subproblems which are decomposition of crew legs and adjustment of nonmaximum crew roster scheme. The decomposition subproblem is transformed to finding a Hamilton chain with the best objective function in network which was solved by an improved ant colony algorithm, whereas the adjustment of nonmaximum crew rostering scheme is finally presented as a set covering problem and solved by a two-stage algorithm. The effectiveness of the proposed models and algorithms are testified by a numerical example.

1. Introduction

Crew planning problem is known as crew scheduling and rostering in the transportation market-airlines, railways, mass transit, and buses. The common features for all these applications are: (i) both temporal and spatial features are involved, that is, each task is characterized by its starting time and location and its finishing time and location and (ii) all tasks to be performed by employees are determined according to a given timetable. Tasks are the smallest elements (or building blocks) and are obtained from the decomposition of flight, train or bus journey, or a trip between two or more consecutive stops in a bus line.

In recent decades, the research and applications of airline crew scheduling and rostering received extensive attention because of its enormous economic benefits and impact, and more articles have been devoted to methodologies and applications in this area than

to any other rostering application areas. The most popular approach for solving airline crew scheduling and rostering is the well-known decomposition technique, in which the overall problem is solved in three major stages: (i) crew pairing generation, (ii) crew pairing optimization, and (iii) crew rostering. Crew pairing generation is a process of constructing all or a large number of feasible pairings/duties from the given timetable. Some of the pairings generated in the first stage will be selected so that all flight legs/trips are covered at a minimum cost. In the last stage, the pairings selected in the second stage are sequenced into rosters that will be assigned to individual crew. For these three stages, the first two are usually called as crew scheduling problem, while the last one is called as crew rostering problem.

Crew rostering problem is combinatorial according to its features. The aim of crew rostering problem is to assign individual crew members to the planned rotations once the set of rotations has been determined (a rotation is understood to be a sequence of crew legs/trips on consecutive days made by a crew that leaves from and returns to the same base). The crew rostering problem is most often treated as a zero-one integer programming problem with a defined objective function and specific set of constraints. Buhr [1] proposed minimizing the discrepancy between the average monthly flight time per crew member and the monthly flight time of individual crew members. When the scale of the problem is large, it is often not possible to find an optimal solution, so the "day-by-day" heuristic algorithm [2, 3] and the "crew-by-crew" heuristic algorithm [4–6] are widely used for generating an initial (or satisfactory) solution. The main drawback of "day-by-day" algorithm is that it usually leads to no available crew members on following days, while the drawback of "crew-by-crew" method very often leads to uneven workloads for the crew members. Furthermore, other techniques including branch-and-bound [7], genetic algorithm [8], and simulated annealing [9] are all used to find the optimal solution.

Ryan [10], Gamache, and Soumis [11] modeled the crew rostering problem as a generalized set partitioning problem. Ryan [10] used linear relaxation and the branch-and-bound method. Gamache and Soumis [11] solved the generalized set partitioning problem by column generation, and the numerical tests on problem from Air France show that the algorithm was capable of solving very large-scale problems.

As discussed above, many articles deal with airline crew planning problem, but the railways applications of crew scheduling and rostering have appeared only recently in the public transport sector literature. Since the development of high-speed railway lines is so fast, the need for efficient methods for generating computerized crew roster schemes is growing because the complexity involved in this issue is too much to be managed manually.

The remainder of this paper is organized as follows: the overall statement of the problem is first presented in Section 2 where the crew planning problem is divided into two subproblems. The formulations of the two subproblems are given in Sections 3 and 4, respectively, while a numerical example with generated data is represented in Section 5.

2. Problem Statement

Crew rostering plan with given cycle is a specific form in which the crew routes (or crew legs) are formed into crew rosters under the premise that the plan cycle T , the work time T_w , and the rest time T_r are all given (where T equals the sum of T_w and T_r). Then the crew routes are assigned to crew members as a continuous work plan. The crew members arranged in crew rostering plan is usually set as the primary objective since it matters greatly in the operation costs.

The determination of the given cycle of crew rostering plan is mainly dependent on the operating characteristics and the working regulations of crew members of high speed railway lines, which should also guarantee that the working time of crew members is in accordance with the working standard. Because the crew routes (or crew legs) got by the phase of crew scheduling plan generally have different working time, the crew roster schemes consisted by crew routes (or crew legs) usually have variance in working time. In the long run, the unbalance of working time between crew members can be relieved by letting crew members work under different crew roster schemes in several cycles. However, from the viewpoint of plan-making, for all crew roster schemes, minimizing the variance of working time is helpful for keeping the workload balance of all crew members.

In the process of working out the crew scheduling plan, crew legs are further combined into crew routes. However, when it comes to the crew rostering plan, both crew routes and crew legs can be used as the fundamental data. In contrast, using crew legs to make up the crew rostering plan is more flexible, and actually includes the scenarios of making plan with the input of crew routes. Therefore, this paper only focuses on the implementation method of crew rostering plan with the input of crew legs. For the crew rostering plan with given cycles, we have two specific forms.

- (i) Dividing all crew legs into several subsets which could form a feasible crew roster scheme subject to all constraints, and then the crew roster scheme is assigned to T crew members (or T crew member groups) to accomplish sequentially.

Crew rostering plan with this pattern has strong regularity, which is easy to be worked out and helpful for increasing the possibility to manage the crew members by groups. However, when applying crew rostering plan with this pattern, the crew roster scheme executed sequentially by T crew members may sometime use more crew members than the lower bound we really need. In order to explain this scenario, a term named as maximum crew roster scheme is defined at the first stage.

Maximum Crew Roster Scheme

For a feasible crew roster scheme, if the total working time exceeds the working time standard T_w while connecting any other crew leg, this scheme will be regarded as a maximum crew roster scheme, and then the crew members needed to accomplish this scheme are equal to the cycle of the crew rostering plan (T).

In crew rostering plan worked out by the first mode, the number of crew members used just equals kT (where k is the number of crew roster schemes) if all crew roster schemes are maximum crew roster scheme. If there is a nonmaximum crew roster scheme, some crews will be unnecessarily used if we still assign this scheme to T crew members.

- (ii) Determining crew roster scheme for each crew member directly in crew rostering plan. In this circumstance, the matching problem between crew members and crew legs instead of crew routes is directly considered, which is beneficial to reduce the crew member arranged in crew rostering plan.

A feasible crew rostering plan worked out by this method is shown in Table 1, where the cycle of the plan is 9 days, of 7 days for work and 2 days for rest, 3 crew legs are accomplished by 4 crew members.

Table 1: A crew rostering plan with given cycle of 9 days.

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9
Crew member 1	1	2	3	1	2	3	1	Rest	Rest
Crew member 2	Rest	Rest	1	2	3	1	2	3	1
Crew member 3	3	1	Rest	Rest	1	2	3	1	2
Crew member 4	2	3	2	3	Rest	Rest	Rest	2	3

Table 1 shows that the crew roster schemes for each crew member are not exactly the same. While the scale of the problem is large, the compilation and implementation of the plan will be relatively complicated.

In order to work out the crew rostering plan more easily and more convenient, an implementation strategy by integrating the advantages of the two methods discussed above is put forward. In this strategy, the first method is adapted to work out a crew rostering plan which has maximum crew roster schemes as much as possible. In summary, the implementation method of crew rostering plan with given cycle studied in this paper contains two steps: the first one realizes the decomposition of all crew legs, while the second one gives further adjustment to the nonmaximum crew roster scheme which may be obtained by the first step. For convenience, the two steps above are called as decomposition of crew legs and adjustment of nonmaximum crew roster scheme, and then the optimization model and algorithm for the two subproblems are studied.

3. Model and Algorithm of Decomposition of Crew Legs

3.1. Description

The decomposition problem divides all crew legs into several subsets, and each set could form a feasible crew roster scheme. Because the number of crew members used affects directly the total cost of daily operation, the first target of this problem is usually to minimize the crew members arranged. Meanwhile, for the working time in each crew roster scheme could be different, and consequently, the working load between all crew roster schemes should also be considered when dividing the crew legs. As mentioned above, the crew members needed to accomplish a maximum crew roster scheme is exactly equal to the cycle of crew rostering plan. But for a nonmaximum crew roster scheme with m crew legs, if a crew member could only undertake one crew leg in one day, then the lower bound of crew members needed to accomplish the nonmaximum crew roster scheme is $\lceil (m \times T) / T_w \rceil$.

The decomposition problem is essentially a set partitioning problem, which could be described by the representation according to feasible schemes. While generate feasible crew roster schemes, the following regulations and constraints should be considered.

- (i) One crew member could undertake at most one crew leg in one day.
- (ii) Within the given cycle, the accumulated working time of crew members should not exceed the working time standard T_w .
- (iii) The connection time between two adjacent crew legs should satisfy the corresponding standard.
- (iv) Within the given cycle, each crew member used should at least assign a continuous rest T_r .

Through the above analysis, the optimization objectives put forward in this paper for crew legs partitioning problem are as follows.

- (i) Minimizing the number of feasible crew roster schemes, which means maximize the number of maximum crew roster schemes at the same time.
- (ii) Minimizing the number of crew legs in the nonmaximum crew roster scheme if there is one in the crew rostering plan.
- (iii) Realizing the most equalized working time distribution of all maximum crew roster schemes.

3.2. Crew Legs Partitioning Model

$$\min \sum_{i=1}^m x_i, \quad (3.1)$$

$$\min \sum_{i=1}^m x_i(1 - \sigma_i)k_i, \quad (3.2)$$

$$\min \sum_{i=1}^m x_i \sigma_i \left(t_i - \frac{\sum_{i=1}^m x_i t_i}{\sum_{i=1}^m x_i} \right)^\alpha, \quad (3.3)$$

$$\text{subject to } \sum_{i=1}^m x_i a_{ij} = 1, \quad j = 1, 2, \dots, n, \quad (3.4)$$

$$x_i \in \{0, 1\}, \quad i = 1, 2, \dots, m, \quad (3.5)$$

where m is the total number of feasible crew roster schemes, x_i is a 0-1 decision variable equals 1 if feasible crew roster scheme i is chosen in the final solution, and 0 otherwise, σ_i is a 0-1 variable equals 1 if feasible crew roster scheme i is a maximum crew roster scheme, and 0 otherwise, t_i is the total working time of all crew legs in feasible crew roster scheme i , k_i is the number of crew legs contained in feasible crew roster scheme i , α is the control parameter for calculate the proportionality of crew roster schemes, a_{ij} is a 0-1 variable that equals 1 if feasible crew roster scheme i contains crew leg j , and 0 otherwise, n is the total number of crew legs.

Objective functions (3.1)–(3.3) correspond to the optimization objectives in Section 3.1, respectively. Constraint (3.4) represents that each crew leg belongs to one crew roster scheme; constraint (3.5) is the 0-1 restriction of decision variable.

3.3. Ant Colony Algorithm with Biphormone and Biheuristic Information

The model founded above is a set partitioning model with multiobjective, for which finding a feasible solution is NP-hard. In order to solve the model easily, the problem is converted into a network optimization problem by treating the crew legs as nodes, and treating the connections between crew legs as arcs, then the new problem needed to be solved is to find a Hamilton chain, which covered all crew legs. The Hamilton chain is composed by some subchains, each of which represents a feasible crew roster scheme. The issue of finding

Hamilton chain in a network could be executed by algorithms based on construction strategy. Ant colony algorithm is a kind of bionic algorithm simulating the mode of finding route in the real world by ants, which has already performed well when applied in problems that can be solved by construction strategy [5, 12]. For this reason, the ant colony algorithm is chosen to solve the converted problem. The critical operations of ant colony algorithm which are different from the common use are as follows.

(i) *The Construction of the Solution*

Ant k starts from the crew leg with the earliest ending time and continues the construction process by adding new crew legs one by one. If the working time of the sequence of crew legs does not exceed the working standard, then the next crew leg will be chosen by the probability P_{ij}^k , otherwise, it should make clear a new feasible crew route scheme has already been found, after recording the scheme and indexes, setting the value of variable that record the working time of crew leg sequence to 0, and now, the ant chooses a crew leg as the start node for a new crew roster scheme by the probability P_i^k . While all crew legs have been selected, the key issue is to determine whether the last crew roster scheme is a maximum crew roster scheme or not, calculate all indexes, and finish the construction process of ant k .

(ii) *Representation of Phenomena*

The ant colony algorithm in this paper has bi-phenomena representation methods that both record the phenomena of all arcs τ_{ij} and all nodes τ_i in the solution construction graph. τ_{ij} is the degree of the expectation for crew leg i connecting crew leg j , while τ_i is the expectation for selecting crew leg i as the start node for the next crew roster scheme.

(iii) *Selection Strategy*

During the selection process, the probability of choosing next crew leg for ant k located at node i is related with its current state.

When the node i where ant k is located is not an "origin" node, the formula of probability for selecting the next crew leg is as follows:

$$P_{ij}^k = \begin{cases} \frac{[\tau_{ij}]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}]^\alpha [\eta_{il}]^\beta}, & \text{if } j \in N_i^k, \\ 0, & \text{otherwise,} \end{cases} \quad (3.6)$$

where η_{ij} is the heuristic information calculated by the following formula:

$$\eta_{ij} = \frac{1}{t_{ij}} \times \frac{1}{|((T_i + t_j)/m) - t^*| + 1}, \quad (3.7)$$

where T_i is the accumulated working time of all crew legs in current sequence, t_{ij} is the connection time between crew leg i and crew leg j , t_j is the working time of crew leg j , m is the number of all crew legs after connecting crew leg j , and t^* is the average working time of all crew legs.

When the node i where ant k is located is an “origin” node, the formula of probability for selecting the next crew leg as the start node of a new crew roster scheme is as follows:

$$P_i^k = \begin{cases} \frac{[\tau_i]^\alpha [\eta_i]^\beta}{\sum_{\text{Selected}_i \neq 1} [\tau_i]^\alpha [\eta_i]^\beta}, & \text{if Selected}_i \neq 1, \\ 0, & \text{otherwise.} \end{cases} \quad (3.8)$$

The heuristic information η_j is calculated by the following formula, in which t_j^e is the end time of crew leg j :

$$\eta_j = \frac{1}{t_j^e}. \quad (3.9)$$

(iv) Evaluation Function

It is inconvenient to evaluate the solutions constructed by the ants since the model contains three optimization objectives. Thus, these objectives are considered in one formula by weight when establishing the evaluation function

$$\min \left\{ \alpha \sum_{i=1}^m x_i + \beta \frac{\sum_{i=1}^m x_i (1 - \sigma_i) k_i}{T_w} + \gamma \frac{\sum_{i=1}^m x_i \sigma_i (t_i - (\sum_{i=1}^m x_i t_i / \sum_{i=1}^m x_i))^2}{(t_{\max} - (\sum_{i=1}^m x_i t_i / \sum_{i=1}^m x_i))^2 \sum_{i=1}^m x_i \sigma_i} \right\}, \quad (3.10)$$

where α , β , γ , are the weigh factors of each optimization objective, respectively, t_{\max} is T_w times of the working time of crew leg which has longest working time. It is obvious that the first part of formula (3.10) is an integer and the values of other parts are within the interval $[0,1)$. Since the number of crew member arranged is the objective that most concerned by the management organization, and the importance of other two objectives reduce, gradually, the values of α , β , γ , are set to 1, 1, and 0.1, and the value of each part belongs to $[0,m]$, $[0,1)$, and $[0,0.1)$, respectively.

4. Model and Algorithm of Adjustment of Nonmaximum Crew Roster Scheme

4.1. Description

The essence of the adjustment of nonmaximum crew roster scheme is to work out a new crew rostering plan with the crew legs contained by the nonmaximum crew roster scheme. At this time, the second strategy for the implementation of crew rostering plan is used and its main objectives are still minimizing the crew members arranged and equalizing the working time distribution of all crew roster schemes. While adjusting the nonmaximum crew roster scheme, the constraints in Section 3.1 should be satisfied. Moreover, any crew leg must be assigned to one and only one crew member.

4.2. Multiobjective Model

The optimization model founded for the second subproblem in this paper is as follows

$$\min M, \quad (4.1)$$

$$\min \sum_{i=1}^M \left(\sum_{k=1}^T \sum_{j=1}^N x_{ij}^k t_j - \frac{T \sum_{j=1}^N t_j}{M} \right)^2, \quad (4.2)$$

$$\text{subject to } \sum_{i=1}^M x_{ij}^k = 1, \quad j = 1, 2, \dots, N; \quad k = 1, 2, \dots, T, \quad (4.3)$$

$$\sum_{j=1}^N x_{ij}^k \leq 1, \quad i = 1, 2, \dots, M; \quad k = 1, 2, \dots, T, \quad (4.4)$$

$$\sum_{k=1}^T \sum_{j=1}^N x_{ij}^k \leq T_w, \quad i = 1, 2, \dots, M, \quad (4.5)$$

$$\sum_{j=1}^N x_{ij}^{f(k+1,T)} (t_j^s + 1440) - \sum_{j=1}^N x_{ij}^k t_j^e - T_{\text{con}} + \theta \left(2 - \sum_{j=1}^N x_{ij}^{f(k+1,T)} - \sum_{j=1}^N x_{ij}^k \right) \geq 0, \quad (4.6)$$

$$i = 1, 2, \dots, M; \quad k = 1, 2, \dots, T$$

$$\sum_{k=1}^T \prod_{r=0}^{T_r-1} \left(1 - \sum_{j=1}^N x_{ij}^{f(k+r,T)} \right) \geq 1, \quad i = 1, 2, \dots, M, \quad (4.7)$$

$$x_{ij}^k \in \{0, 1\}, \quad i = 1, 2, \dots, M; \quad j = 1, 2, \dots, N; \quad k = 1, 2, \dots, T, \quad (4.8)$$

where M is the crew members needed to finish all the task; N is the total number of crew legs in nonmaximum crew roster scheme; x_{ij}^k is a 0-1 decision variable that equates 1 if crew member i undertakes crew leg j on day k , and 0 otherwise; T is the given cycle of crew rostering plan; T_w is the working time while T_r is the continuous rest time for crew members in cycle T ; T_{con} is the standard for the connection of crew legs; t_j is the work time of crew leg j ; t_j^s is the start time while t_j^e is the end time of crew leg j ; θ is an infinite positive number.

The feasible crew roster schemes in crew rostering plan are all loops consisting of crew legs, so the connection time between each two adjacent crew legs should satisfy the standard. Hence, the following function is designed to reflect this characteristic:

$$f(k, T) = \begin{cases} k\%T, & \text{if } k < T, \\ T, & \text{otherwise.} \end{cases} \quad (4.9)$$

Objective function (4.1) is designed to minimize the crew members arranged in the crew rostering plan, while objective function (4.2) means the most equalized working load among all crew roster schemes. Constraints (4.3)–(4.7) correspond to the restrictions in Section 3.1, respectively. The $\theta(2 - \sum_{j=1}^N x_{ij}^{f(k+1,T)} - \sum_{j=1}^N x_{ij}^k)$ in constraint (4.6) guarantees that

the formula is tenable while one of the adjacent jobs is not the crew leg. Constraint (4.8) is the restriction of 0-1 decision variable.

4.3. Two-Phase Algorithm

It is hard to solve the model founded above by traditional optimization algorithms because it includes both linear constraints and constraints based on the description. Therefore, this paper designs an algorithm with two phases. The first phase represents the converted problem by set covering model, and then the initial feasible crew rostering plan with minimal crew members arranged is obtained. On the basis of the result of the first phase, an algorithm with the ideology of simulated annealing is used in the second phase to implement further improvement until a better crew rostering plan is obtained.

Phase 1. Initial crew rostering plan with minimal crew members arranged. The adjustment of nonmaximum crew roster scheme is converted into a standard set covering problem by representing the feasible crew roster scheme with 0-1 string, and then all feasible crew roster schemes are found. Finally, the initial crew rostering plan with minimal crew members is obtained by solving the set covering model. The generation of crew roster scheme, the conversion of feasible scheme, and other issues are as follows.

(i) The Generation and Conversion of Feasible Crew Roster Scheme

The crew roster schemes obtained by the decomposition of crew legs contains at least one nonmaximum crew roster scheme, in which the crew legs contained must be less than T_w . Therefore, all the feasible crew roster schemes can be easily found by a deep search method. The generated crew route schemes are represented by the real index of crew legs, thus a 0-1 string with the length of M is used to represent the crew leg for one day in the scheme. The crew leg i is undertaken by crew members when the i th element of the string equals 1, and the crew member is assigned a rest while all elements in the string equal 0. Therefore, a 0-1 string with the length of MT can be used to represent the situation of all the schemes with T days. Thanks to the conversion from feasible crew roster schemes, there does not exist the situation that more than 2 elements in the substring equal 1 in one day.

(ii) The Generation of the Set of All Feasible Crew Roster Schemes

Because the execution process of crew roster scheme is a circulation, the feasible crew roster scheme got by the deep search method can just represent a specific situation in the cycle. In order to generate all feasible crew roster schemes, a duplication process is used to copy the schemes found in step (i). Each crew roster scheme can be duplicated T schemes for the cycle of the scheme equals to T . For a crew rostering plan with the cycle of 4 days, in which 3 days are working and 1 day is resting. Since a feasible crew roster scheme is 2-1-3, the schemes duplicated are shown in Table 2.

All feasible crew roster schemes could be obtained after the duplication to the searched schemes, but some of them are exactly the same ones (e.g., a scheme 1010 will become two 1010 and two 0101 after the duplicate). In order to keep the uniqueness of all schemes, a procedure that checks the repetition and deletes the repeated schemes is then applied. After the above operation, the problem of adjusting nonmaximum crew roster scheme becomes

Table 2: Duplication of feasible crew roster scheme.

Schemes	Before conversion	After conversion
Initial scheme	2-1-3-0	0100-1000-0010-0000
Duplicated scheme 1	1-3-0-2	1000-0010-0000-0100
Duplicated scheme 2	3-0-2-1	0010-0000-0100-1000
Duplicated scheme 3	0-2-1-3	0000-0100-1000-0010

a standard set covering problem with the objective of covering all crew legs in the plan cycle with minimal number of crew roster schemes

$$\begin{aligned} \min \sum_{u=1}^U y_u \\ \text{subject to } \sum_{u=1}^U y_u a_{uv} \geq 1, \quad v = 1, 2, \dots, MT, \end{aligned} \quad (4.10)$$

where y_u is a 0-1 decision variable that equals 1 if crew roster scheme u is chosen in final solution and 0 otherwise, a_{uv} is a 0-1 variable that represents whether crew roster scheme u covers crew leg v .

(iii) *The Generation of Initial Crew Rostering Plan*

The set covering problem above can be solved by various mature algorithms, such as genetic algorithm, ant colony algorithm, and algorithms based on the application of column generation principle, and interested readers are referred to the related references [13–15].

By solving the set covering model, an initial crew roosting plan with minimal crew members arranged can be obtained accompanied by the scenario where some of the crew legs may be covered more than once. We assume that crew leg i is covered k_i times in the solution of set covering problem, then the initial feasible crew roosting plan of this problem will be obtained by deleting k_i-1 extra covers of crew leg i randomly. Using the opposite process of the duplication while all crew legs appear only one time during the solution, then, the crew roosting plan based on the original description is obtained with minimal crew members.

Phase 2. Adjustment for equalized workload based on the ideology of simulated annealing.

When designing the algorithm for most equalized workload of crew roster schemes, two concepts are defined firstly.

Feasible Swap

In the k th day of crew roosting plan, while the crew roster schemes obtained by swapping the jobs (work or rest) of crew member i and j still satisfy all the restrictions, the operation adopted is called as a feasible swap.

Positive Swap

A feasible swap is called a positive swap while the equilibrium of workload has improved after the adoption of this swap (the variation of equilibrium Δ_i is a positive number). On the contrary, a feasible swap, which reduces the equilibrium, is a negative swap.

According to the definitions, the essence of the adjustment of nonmaximum crew roster scheme is to improve the equilibrium of workload by finding and adopting positive swaps, until a crew rostering plan with high equilibrium is found. For providing a relative larger searching space, ideology of simulated annealing is merged in the designed algorithm, which means that the selection of negative swaps is allowed in the operation.

The details of the algorithm are as follows.

Step 1. Let the initial temperature t equal t_0 , the start day k for searching feasible swap equals 1, calculate the working time T_i of each crew roster scheme in initial crew rostering plan, the average working time T^* of all crew roster schemes, and the equilibrium of initial plan.

Step 2. Search all feasible swaps from the first crew roster scheme at day k .

Step 3. Let $k = k + 1$, if $k > T$, which means that all feasible swaps has already been found and turn to the next step, otherwise, turn to Step 2.

Step 4. Calculate the variation Δ_i of each feasible swap, and then calculate the probability P_i for each swap to be selected.

Step 5. Select a feasible swap by probability P_i , the probability for accepting the swap selected equals 1 when the variation Δ_i is a positive number, and equals P_{accept} when the variation Δ_i is a negative number. Record the swap used and update the equilibrium of crew rostering plan, and then turn to next step. This step is repeated once the selected swap is refused because the probability P_i is lower than P_{accept} .

Step 6. Implement the update of temperature by the update function and determine whether the updated temperature reaches the lowest point t_{min} . Turn to next step if the temperature equals t_{min} , otherwise, let k equals 1 and return to Step 2.

Step 7. Once the adjusting process of the equilibrium of all crew schemes is finished, the best crew rostering plan is outputted.

The parameters in the simulated annealing algorithm above are similar to the common usage except the selection probability of feasible swaps which is described here only. Assume that there are n feasible swaps at one status in Step 5 of the algorithm above, then the probability P_i for choosing feasible swap i is represented as follows:

$$P_i = \frac{\Delta_i - \Delta_{\text{min}} + \varepsilon}{\sum_{i=1}^n (\Delta_i - \Delta_{\text{min}} + \varepsilon)}, \quad (4.11)$$

where Δ_{min} is the minimal variation of workload equilibrium of all feasible swaps, which is usually a negative number. Δ_i is the variation of workload equilibrium of feasible swap i . ε is a small positive number used to ensure that the probability of the feasible swap with minimal variation does not equal zero.

Table 3: Details of generated crew legs.

Crew legs	Start time	End time	Working time
1	868	1240	372
2	708	1109	401
3	652	1116	464
4	829	1195	366
5	984	1360	376
6	775	1146	371
7	701	1179	478
8	999	1432	433
9	812	1239	427
10	450	826	376
11	665	1064	399
12	569	1020	451
13	414	817	403
14	931	1393	462
15	543	978	435
16	784	1153	369
17	604	1035	431
18	491	897	406
19	852	1275	423
20	392	859	467
21	730	1152	422
22	971	1354	383
23	512	949	437
24	574	996	422

In the algorithm designed above, the probability for the acceptance of negative swap is relatively high when at a high temperature and reduced gradually with the decreasing of temperature. When the temperature reaches the lowest level, this probability tends to zero and the algorithm converges finally.

5. Numerical Example

According to the operational characteristics of China high-speed railway currently, this paper generates several crew legs as the basic data to work out the crew rostering plan with given cycle. Assume that the earliest starting time and the latest ending time of all crew legs are 6:00 and 24:00, respectively, which will be 360 and 1440 when represented by minute. The working time of each crew leg equals the time span between start time and end time. The connection time standard between two crew legs is 960 min, and all crew legs start and end at the same crew base. The details of the 24 generated crew legs are shown in Table 3, in which the start time, end time, and working time of all crew legs are represented by minute.

First, with the data generated above, this paper works out a crew rostering plan with the cycle of 8 days, in which 6 days are for working and 2 days for rest, the results are shown in Table 4.

Table 4: Crew rostering plan with the cycle of 8 days.

Roster schemes	Days						Crew members arranged	Total crew member arranged	Lower bound of crew members	Operation time(s)
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6				
Crew roster scheme 1	13	20	10	18	23	15	8			
Crew roster scheme 2	24	12	17	3	11	7	8			
Crew roster scheme 3	2	21	6	16	9	4	8	32	32	3.5172
Crew roster scheme 4	1	19	14	22	5	8	8			

Table 5: Crew rostering plan with the cycle of 7 days.

Roster schemes	Days					Crew members arranged	Total crew member arranged	Lower bound of crew members	Operation time(s)
	Day 1	Day 2	Day 3	Day 4	Day 5				
Crew roster scheme 1	13	20	10	18	23	7			
Crew roster scheme 2	15	12	24	17	3	7			
Crew roster scheme 3	11	7	2	21	6	7	34	34	2.4618
Crew roster scheme 4	16	9	4	19	1	7			
Crew roster scheme 5	22	14	5	8		6			

It is easily obtained from Table 4 that the 4 crew roster schemes in the crew rostering plan are all maximum crew roster schemes which means each scheme contains 6 crew legs that equals the working time criterion. Under this circumstance, the adjustment procedure is not needed and the number of crew members arranged in the crew rostering plan equals the product of the plan cycle and the total number of crew schemes, which has already been minimal and the plan can be executed with high regularity.

Moreover, the crew legs above are used to work out a crew rostering plan with the cycle of 7 days in which 5 days for work and 2 days for rest, the outcome obtained by the decomposition of crew legs is shown in Table 5.

It can be concluded from Table 5 that the previous 4 crew roster schemes are all maximum schemes, and the crew members used in each scheme are equal exactly to the cycle of the crew rostering plan. The 5th crew roster scheme (see the last row in Table 5) is a nonmaximum crew roster scheme. If still using 7 crew members to undertake this scheme at this time, one or more crew members will be not required actually. Actually, the lower bound of crew members needed for the last crew roster scheme is $\lceil (4 \times 7) / 5 \rceil = 6$. Therefore, the second phase is needed to adjust the 5th crew roster scheme. The initial crew rostering plan is worked out with the crew legs in the 5th crew roster scheme by the algorithm designed in Section 4.3. The initial crew rostering plan of adjustment of nonmaximum crew roster scheme is shown in Table 6, in which the number 0 represents no crew leg is assigned to this crew member on this day.

Table 6: Initial crew rostering plan of adjustment of non-maximum crew roster scheme.

Crew	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Accumulated working time (min)	Proportionality (min ²)
1	22	14	8	22	8	0	0	2094	529799.33
2	0	0	22	14	14	8	22	2123	
3	5	22	0	0	22	14	8	2037	
4	14	5	14	5	0	0	5	2052	
5	0	8	5	8	5	22	0	2001	
6	8	0	0	0	0	5	14	1271	

Table 7: Optimal crew rostering plan of adjustment of nonmaximum crew roster scheme.

Crew	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Accumulated working time (min)	Proportionality (min ²)
1	14	14	14	14	0	0	0	1848	29749.33
2	0	0	22	5	22	8	22	1958	
3	5	5	0	0	8	5	8	1994	
4	22	22	8	22	0	0	5	1958	
5	0	8	5	8	5	22	0	2001	
6	8	0	0	0	14	14	14	1819	

It is obvious that the crew member arranged to the 5th crew roster scheme is 6 which is equal to the lower bound of the crew member needed. However, the accumulated working time varies from 1271 min to 2123 min, and the proportionality is not optimal. In order to improve the proportionality, the algorithm based on the ideology of simulated annealing is used and the improved crew rostering plan is shown in Table 7.

After the adjustment of the initial crew rostering plan, the variations of accumulated working time among all crew roster schemes are relatively small, and the proportionality of the plan is much better than before.

6. Conclusion

This paper studies the compilation method of crew rostering plan with given cycle of high-speed railway lines. By merging two different compilation principles, the problem is divided into two subproblems which are the crew legs partitioning problem and the adjustment of nonmaximum crew roster scheme. The division of the problem gives consideration to both of the complexity for compilation and the convenience for execution, and the algorithm designed is effective for finding an optimal solution. By studying this problem, we found that under the assumption that each crew member undertakes at most one crew leg in a day, the lower bounds of crew members needed are $\lceil m/T_w \rceil \times T$ and $\lceil (m \times T)/T_w \rceil$ when work out the crew rostering plan with the two forms discussed in the beginning separately, respectively. While compiling the crew rostering plan by merging the two strategies, which is used in this paper, the lower bound is $\lceil m/T_w \rceil \times T + \lceil (m - \lceil m/T_w \rceil) \times T/T_w \rceil$. When m/T_w is an integer, the value of all lower bound are equal no matter what compiling pattern we use, and the minimal crew members arranged may be still equal. The compiling method of crew rostering plan put forward in this paper has its own advantages, but unexpectedly, sacrifices both objectives to

some extent. The algorithms applied in this paper can also be used separately when working out the crew rostering plan with single requirement. When compile crew rostering plan by set covering model, it should be aware that the scale of the problem will affect both the difficulty of the algorithm design and the execution efficiency.

Acknowledgment

This research work was supported by the National Natural Science Foundation of China (no. 50968009 and no. 71261014).

References

- [1] J. Buhr, "Four methods for monthly crew assignment—a comparison of efficiency," in *Proceedings of the AGIFORS Symposium*, vol. 18, pp. 403–430, 1978.
- [2] B. Nicoletti, "Automatic crew rostering," *Transportation Science*, vol. 9, no. 1, pp. 33–42, 1975.
- [3] D. Teodorović and P. Lučić, "A fuzzy set theory approach to the aircrew rostering problem," *Fuzzy Sets and Systems*, vol. 95, no. 3, pp. 261–271, 1998.
- [4] R. Moore, J. Evans, and H. Noo, "Computerized tailored blocking," in *Proceedings of the AGIFORS Symposium*, vol. 18, pp. 343–361, 1978.
- [5] J. Byrne, "A preferential bidding system for technical aircrew," *Proceedings of the AGIFORS Symposium*, vol. 28, pp. 87–99, 1988.
- [6] P. Lučić and D. Teodorović, "Simulated annealing for the multi-objective aircrew rostering problem," *Transportation Research A*, vol. 33, no. 1, pp. 19–45, 1999.
- [7] H. Dawid, J. König, and C. Strauss, "An enhanced rostering model for airline crews," *Computers and Operations Research*, vol. 28, no. 7, pp. 671–688, 2001.
- [8] A. Monfroglio, "Hybrid genetic algorithms for a rostering problem," *Software: Practice and Experience*, vol. 26, no. 7, pp. 851–862, 1996.
- [9] A. Ernst, M. Krishnamoorthy, and D. Dowling, "Train crew rostering using simulated annealing," in *Proceedings of the International Conference on Optimization: Techniques and Applications (ICOTA '98)*, Perth, Australia, 1998.
- [10] D. M. Ryan, "The solution of massive generalized set partitioning problems in air crew rostering," *Journal of the Operational Research Society*, vol. 43, pp. 459–467, 1992.
- [11] M. Gamache and F. Soumis, "A method for optimally solving the rostering problem," *Cahier Du GERAD, G-90-40 H3T 1V6*, Ecole des Hautes Etudes Commerciales, Montreal, Canada, 1993.
- [12] M. Dorigo and L. M. Gambardella, "Ant colonies for the travelling salesman problem," *BioSystems*, vol. 43, no. 2, pp. 73–81, 1997.
- [13] F. D. Croce, R. Tadei, and G. Volta, "A genetic algorithm for the job shop problem," *Computers and Operations Research*, vol. 22, no. 1, pp. 15–24, 1995.
- [14] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 26, no. 1, pp. 29–41, 1996.
- [15] M. Desrochers and F. Soumis, "A column generation approach to the urban transit crew scheduling problem," *Transportation Science*, vol. 23, no. 1, pp. 1–14, 1989.

Research Article

A Method for Queue Length Estimation in an Urban Street Network Based on Roll Time Occupancy Data

**Dongfang Ma,¹ Dianhai Wang,¹ Yiming Bie,²
Feng Sun,³ and Sheng Jin¹**

¹ College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China

² School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150091, China

³ College of Traffic and Transportation, Jilin University, Changchun 130022, China

Correspondence should be addressed to Dianhai Wang, wangdianhai@zju.edu.cn

Received 15 May 2012; Revised 19 August 2012; Accepted 26 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Dongfang Ma 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 method estimating the queue length in city street networks was proposed using the data of roll time occupancy. The key idea of this paper is that when the queue length in front of the queue detector becomes longer, the speeds of the following vehicles to pass through the detector will become smaller, resulting in higher occupancy with constant traffic intensity. Considering the relationship between queue lengths and roll time occupancy affected by many factors, such as link length, lane width, lane number, and bus ratio, twelve different conditions were designed, and the traffic data under different conditions was obtained using VISSIM simulation. Based on the analysis of simulation data, an S-type logistic model was decided to develop for the relationship between queue lengths and roll time occupancy, and the fitting equations were obtained under the twelve simulation situations. The average model for the relationship between queue lengths and roll time occupancy was presented by successive multiple linear regression with the fitting equation parameters and simulation parameters, and the estimation model for queue length was presented through analyzing the equation of the average relation model.

1. Introduction

With traffic congestion worsening in urban areas, a growing number of signalized intersections are being operated in oversaturated conditions, and the queue lengths at some roads are even approaching the link lengths during peak hours, leading to spillovers [1, 2]. As the influence of spillovers in city street networks is significant [3], monitoring the traffic state of the roads and estimating the conditions under which a spillover will occur are important.

It has long been recognized that the queue length can represent the road traffic state intuitively, and it is the most common index to identify spillovers. Over the years, many

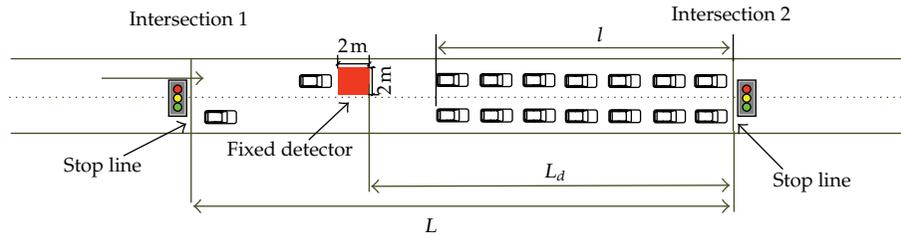


Figure 1: Location of queue detector.

researchers have dedicated themselves to this topic and three types of queue estimation models have been developed. The first one, which is based on the behavior of traffic shockwaves, was first proposed by Lighthill and Whitham for uninterrupted flow [4] and later improved and expanded by a number of researchers to signalized intersections [5–8]. The second one demonstrated by Webster [9] is based on the analysis of cumulative traffic input-output to a signal link. Later, it was improved by some researchers including Daganzo [10], Akcelik [11], and Vigos et al. [12]. As high-resolution traffic signal data, such as second-by-second detector data, vehicle-detector actuation events, and signal phase change events, is becoming increasingly available in recent years; a new model for estimating the maximum queue length during every cycle was developed by Liu et al. [13, 14].

Known probability distribution of traffic flow arrival rule is the premise for the first kind methods to estimate the queue length, and they seemly become no more applicable if the change of traffic state is very complicated or the traffic flow arrival rule does not suit to any known probability distribution. The second class methods can be used to estimate queue length both for macroscopic and microscopic levels; however, constructing the cumulative curve and solving the queue length in these methods are very arduous. In the third type models, the three break points which indicate a change in traffic state are difficult to identify if the residual queue length at the beginning of the red phase is greater than the distance between the stop line and the detector location. Moreover, traffic-wave theory, with its premise of continuous flow, is not suitable for practical application, as the discontinuity of traffic flow in urban street networks is significant under the influence of signal controls [15]; thus, methods taking traffic-wave theory as their basis are impracticable.

To close these gaps mentioned above, this paper analyzes the relationship between queue length and roll time occupancy, using the data collected by VISSIM simulations, and then develops a new method to estimate the queue length using the data of roll time occupancy. The finding of this paper may provide oversaturated arterials a basis to identify spillover conditions.

2. Computation of Roll Time Occupancy

With the aim of obtaining traffic data for predicting the road traffic state, fixed loop detectors are placed sufficiently upstream from the intersection stop line in SCOOT (Split, Cycle, and Offset Optimization Technique) and other systems [16]. These detectors can be called as queue detectors or advanced detectors, whose locations can be illustrated as in Figure 1.

The most common type of queue detector is a loop detector, which can provide traffic flow information in the form of pulse data, and is shown in Figure 2. The upper bars represent

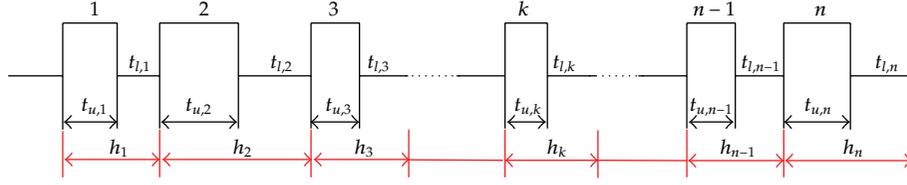


Figure 2: Pulse information from a loop detector.

the state where the detector is occupied by one vehicle, and the duration corresponds to occupancy time of this vehicle; the lower bars stand for the gaps between the rear of the forward vehicle and the start of the following vehicle. The sum of the durations of an upper and the following lower bars is the headway between two successive vehicles.

$t_{u,k}$, the occupancy time of the k th vehicle, can be obtained from the vehicle length and average speed data using:

$$t_{u,k} = \frac{(L_k + L_q)}{u_k}, \quad (2.1)$$

where u_k is the average speed of the k th vehicle when passing through the queue detector (m/s); L_k is the length of vehicle k (m); L_q is the length of queue detector (m), which is about 2.0 m in actual application.

As shown in (2.1), the lower the value of u_k , the higher the value of $t_{u,k}$; meanwhile, the expected speed of all drivers is the free flow speed, and a lower speed indicates a higher queue length (i.e., the state on the road is congested). As a consequence, it can be deduced that the occupancy time of one vehicle may characterize the traffic condition around the queue detector at the time when the vehicle is passing through and the queue length can be estimated by analyzing the data of occupancy.

Occupancy is the ratio of the sum of the durations of the upper bars during a given period (denoted as T) to the entire duration of that period. Aiming to identify the traffic condition and estimate the queue length at some roads instantaneously, the idea of roll time occupancy is proposed, which could indicate the traffic condition around a queue detector continuously within an interval of Δt . Figure 3 illustrates the statistical method of calculating occupancy time, taking $T = 5\Delta t$ as an example.

The roll time occupancy in period i , o_i , is given by

$$o_i = \frac{t_i}{T}. \quad (2.2)$$

Here, Δt is the time of rolling step (s), usually set to 1s in order to obtain the traffic state around the queue detector each second; T is the time length of each period for counting roll time occupancy (s), which is usually set to 5s in real application; t_i is the time for which queue detectors are occupied by vehicles in period i (s).

As mentioned above, the traffic condition will become congested around the queue detectors as the increasing of occupancy, and there is a positive correlation between roll time occupancy and queue length for known traffic intensity. Thus, the queue length on some roads can be estimated by analyzing the data of roll time occupancy.

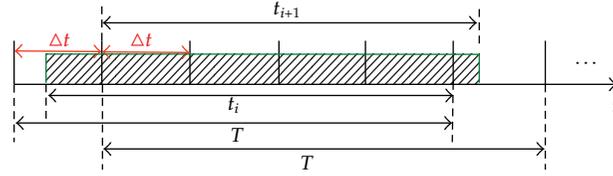


Figure 3: Calculation of roll time occupancy.

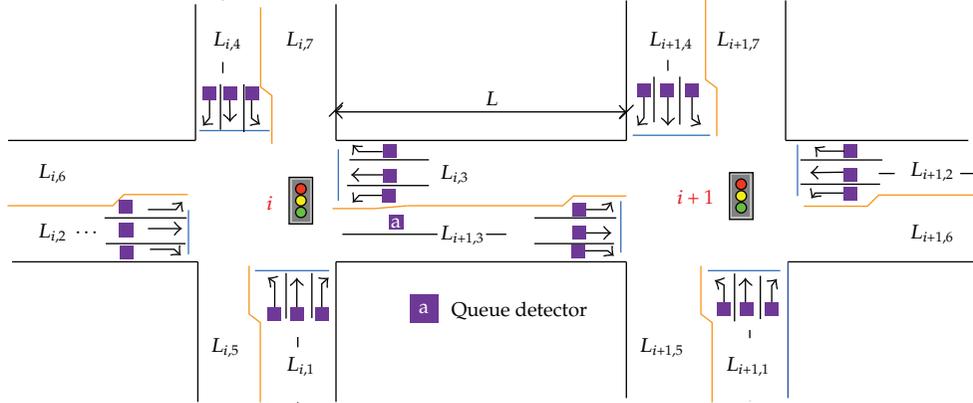


Figure 4: Experimental road grid for the simulation.

3. Relationship between Queue Length and Roll Time Occupancy

Generally, higher roll time occupancy corresponds to a longer queue length, and the relationship between the two can be obtained from traffic data.

3.1. Data Collection and Relevant Parameters

Traffic data collected by a queue detector over a given period of time can vary wildly under the influence of different traffic flow and road conditions. In this paper, traffic data was collected using a VISSIM simulation under different conditions. The experimental road grid is shown in Figure 4.

Here, $i + 1$ is the downstream intersection, and $L_{i+1,3}$ is the link of interest, where a queue detector is placed at the inside lane. The distance between the queue detector and the upstream intersection is 50 m, and the link lengths of the six approaches, other than $L_{i+1,3}$ and $L_{i,3}$, whose parameters vary and are given later, are listed in Table 1.

The signal timings of the two intersections are also presented in the Table 1, with phases 1, 2, 3, and 4 representing the East-West through phase, the East-West left turn phase, the South-North through phase, and the South-North left turn phase, respectively. The key parameters in VISSIM are set as follows: (1) the expected speeds of cars and buses are 15 m/s and 10 m/s, respectively; (2) the saturated flow is 1,800 veh/h for all lanes; (3) the traffic volume is set to be 1,200 veh/h for all approaches; (4) the length of channelization segment is 71 m for all approaches; (5) the ratio of right-turning to through to left-turning vehicles on all approaches is 1 : 2 : 1; (6) the random seed is 41; (7) when the speed is less than 0.28 m/s,

Table 1: Link lengths and phase sequences of the two intersections.

Intersection	Link length (m)			Green time (s)			
	Approach no. 1	Approach no. 2	Approach no. 4	Phase no. 1	Phase no. 2	Phase no. 3	Phase no. 4
i	550	550	550	32	20	36	20
$i + 1$	550	580	550	24	21	32	31

Table 2: Parameters relating to link no. 3 in the simulation environment.

Condition number	L (m)	n	w_d (m)	r (%)	Condition number	L (m)	n	w_d (m)	r (%)
1	300	2	3.5	5	7	300	2	3.5	5
2	350	2	3.5	5	8	300	2	3.5	10
3	400	2	3.5	5	9	300	2	3.5	15
4	350	1	3.5	5	10	300	2	3	5
5	300	2	3.5	5	11	300	2	3.5	5
6	300	3	3.5	5	12	300	2	3.75	5

we define the queue as building-up, and when the speed is greater than 1.39 m/s, we define the queue as dissipating; (8) the simulation time is 3,600 s.

With known traffic demand, the operational characteristics of the traffic flow on the stretch of road $L_{i+1,3}$ will differ depending on the parameters in VISSIM, including link length (L), lane width (w), number of lanes (n), and bus ratio (r). A total of twelve simulation conditions were designed for this paper; the parameters of which are listed in Table 2.

The following section will collect the traffic data under different conditions and analyze the relationship between the queue lengths on the interest link and roll time occupancy collected by queue detectors; Sections 3.3 and 3.4 will discuss the relationship between the parameters of the new model and these four simulation parameters mentioned above.

3.2. Queue Length and Roll Time Occupancy

The signal timing plans of the two intersections were set using the VAP module of VISSIM, and the traffic data was obtained regarding the roll time occupancy and maximum queue length with a rolling step of Δt by adjusting the data output interval. Figure 5 illustrates the relationship between roll time occupancy and queue length under simulation condition no. 1.

Similar relationship models between queue length and roll time occupancy can be obtained by changing the simulation conditions, and then an S-type logistic model for the relationship was developed based on the analysis of the points in Figure 5. The conceptual shape of the logistic model is shown in Figure 6 and is devised as follows:

$$o = o_{\min} + \frac{o_{\max} - o_{\min}}{1 + e^{[-b(l-l_w)]}}, \quad (3.1)$$

where o_{\min} is the minimum roll time occupancy; o_{\max} is the maximum roll time occupancy; l is the queue length of a road where spillovers appear regularly (m); l_w is the queue length at the deflection point of the curve and b determines the slope of the curve.

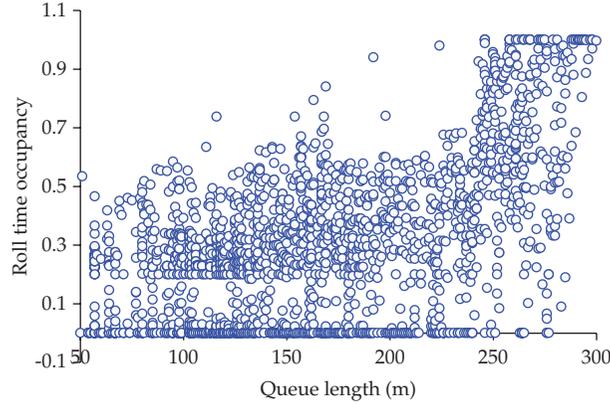


Figure 5: Relationship between l and o with $w_d = 3.5$ m.

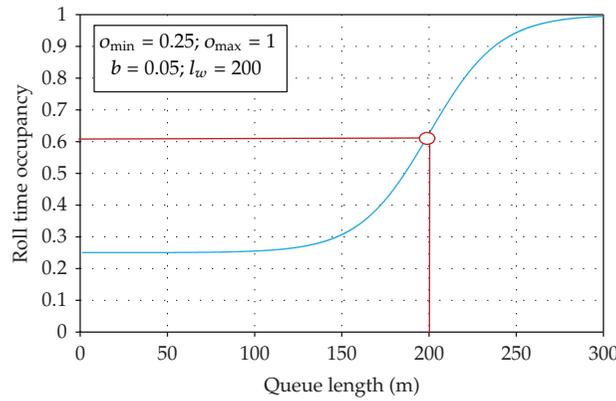


Figure 6: Conceptual shape of l versus o .

It is necessary to define the upper and lower values of the asymptotes, that is, the values of o_{\max} and o_{\min} . o_{\min} is the smallest roll time occupancy under a given simulation environment. The parameter, o_{\max} , which is calculated based on the simulation conditions, is the maximum roll time occupancy. It is equal to 1, which can be confirmed from the definition of roll time occupancy. Given the initial parameters and the assumed form of the model, it is possible to calibrate the model of each simulation environment by applying the best-fitting parameters, based on the data collected under each of the twelve simulation conditions. The result of this calibration for the simulation situation in Figure 5 is shown in Figure 7.

The S-shape of the curve is also superimposed over the entire range of queue length, up to 300 m. The data points for the simulation environment are presented in order to show the “goodness-of-fit” of the model to the data. In this study, the fitting relationship between o and l is

$$o = 0.2858 + \frac{0.7142}{1 + e^{[-0.0552(l-251.3985)]}}, \quad R = 0.7529. \quad (3.2)$$

Using the same method, equations were fitted for the twelve simulation situations, as given in Table 3.

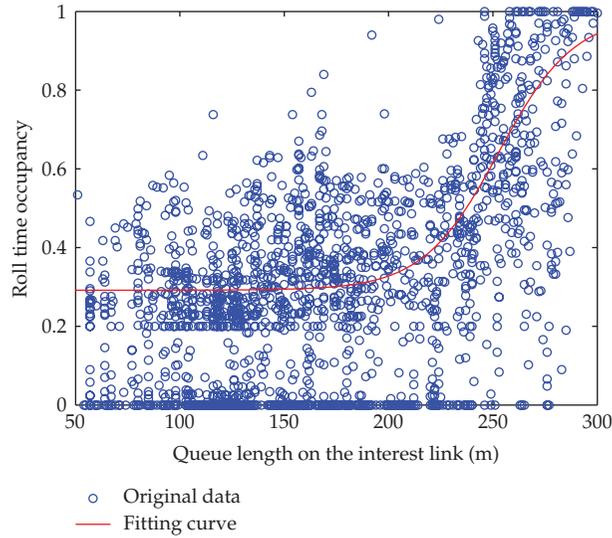


Figure 7: Calibration of S-curve between l and o .

Table 3: Fitting equations and correlation coefficients under different simulation conditions.

Number	Fitted equations	R	Number	Fitted equations	R
1	$o = 0.2858 + \frac{0.7142}{1 + e^{[-0.0552(l-251.3985)']}}$	0.7529	7	$o = 0.2858 + \frac{0.7142}{1 + e^{[-0.0552(l-251.3985)']}}$	0.7529
2	$o = 0.2801 + \frac{0.7199}{1 + e^{[-0.0516(l-299.76)']}}$	0.8716	8	$o = 0.3082 + \frac{0.6918}{1 + e^{[-0.0953(l-247.7518)']}}$	0.8000
3	$o = 0.2499 + \frac{0.7501}{1 + e^{[-0.0375(l-327.7079)']}}$	0.8319	9	$o = 0.3349 + \frac{0.6651}{1 + e^{[-0.0949(l-252.7541)']}}$	0.7671
4	$o = 0.3223 + \frac{0.6777}{1 + e^{[-0.0572(l-278.2494)']}}$	0.7652	10	$o = 0.2917 + \frac{0.6998}{1 + e^{[-0.0552(l-253.8729)']}}$	0.7409
5	$o = 0.303 + \frac{0.696}{1 + e^{[-0.086(l-249.0782)']}}$	0.7758	11	$o = 0.3330 + \frac{0.667}{1 + e^{[-0.0979(l-255.6569)']}}$	0.7416
6	$o = 0.2917 + \frac{0.7083}{1 + e^{[-0.0797(l-251.7239)']}}$	0.7847	12	$o = 0.2805 + \frac{0.7195}{1 + e^{[-0.0749(l-247.9341)']}}$	0.7953

Analyzing the fitting equations under different simulation conditions, the values of o_{\min} are similar in each case; however, the values of b and l_w vary greatly. Consequently, the four parameters, L , w_d , n , and r can be said to strongly influence the values of b and l_w , and it is thus necessary to obtain the relationship between o and l under normal conditions.

3.3. Regression Analysis for Parameter b

Parameter b represents the geometric complexity for constant traffic intensity. The complexity is calculated from the link length, lane width, number of lanes, and bus ratio. A macroscopic model was developed for parameter b through a multiple-stepwise linear regression of the four parameters under each simulation situation. Two models were produced by the statistical software, SPSS [17]; a summary of which is presented in Table 4.

Table 4: Model summaries for parameter b .

Model	R	R^2	Adjusted R^2	Std. error
1	0.663 ^a	0.440	0.384	0.0258196
2	0.813 ^b	0.661	0.586	0.0121482

^aPredictors: (constant), bus ratio; ^bPredictors: (constant), bus ratio, link length.

Table 5: ANOVA for parameter b .

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	0.002	1	0.002	7.855	0.019 ^a
	Residual	0.002	10	0.000		
	Total	0.004	11			
2	Regression	0.003	2	0.001	8.785	0.008 ^b
	Residual	0.001	9	0.000		
	Total	0.004	11			

^aPredictors: (constant), bus ratio; ^bPredictors: (constant), bus ratio, link length.

Table 6: Coefficients for parameter b .

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. Error	β		
1	Constant	0.041	0.010		4.151	.002
	Bus Ratio	0.403	0.144	0.663	2.803	.019
2	Constant	0.134	0.039		3.428	.008
	Bus Ratio	0.337	0.121	0.555	2.787	.021
	Link length	0.00028	0.000	-0.483	-2.425	.038

As the determination coefficient (R^2) of model no. 1 is less than 0.5, the goodness-of-fit is not acceptable. However, the determination coefficient of model no. 2 is 0.586, indicating an acceptable goodness-of-fit [18]. The variance analysis is presented in Table 5.

From the variance analysis of model no. 2 in Table 5, it can be seen that the model has very high precision ($F = 8.785, P = 0.008 < 0.01$). Moreover, all of the values of P for the three terms in t -test are less than 0.05 (see Table 6), which indicates that all three regression coefficients are significant with $\alpha = 0.05$. The regression coefficients are tabulated in Table 6.

In model no. 2, the regression coefficients of r , L , and the constant term are 0.337, 0.00028, and 0.134, respectively. Therefore, the regression equation can be expressed as, which indicates that the value of b is only affected by L and r the following:

$$b = -0.000228L + 0.337r + 0.134. \quad (3.3)$$

3.4. Regression Analysis for Parameter l_w

It is also necessary to replace the value of l_w (the queue length at the inflection point of the S-curve in Figure 6) with a regression model with the four aforementioned parameters under the twelve simulation situations. One model is produced by SPSS, as summarized in Table 7.

The determination coefficient (R^2) of this model is 0.959; the goodness-of-fit is acceptable. The variance analysis is listed in Table 8.

Table 7: Model summary for parameter l_w .

Model	R	R ²	Adjusted R ²	Std. error
1	0.979 ^a	0.959	0.955	5.3884779

^aPredictors: (constant), link length.

Table 8: ANOVA for parameter l_w .

Model	Sum of squares	df	Mean square	F	Sig.	
1	Regression	6737.63	1	6737.63	232.05	.000 ^a
	Residual	290.36	10	29.04		
	Total	7027.99	11			

^aPredictors: (constant), bus ratio; ^bPredictors: (constant), bus ratio, link length.

Table 9: Coefficients for parameter l_w .

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	β		
1	(Constant)	23.378	15.874		1.473	.172
	Link length	0.760	0.050	0.979	15.233	.000

The values of F and P in this model are 232.047 and 0.000, respectively, which indicates that the precision of this regression model is very high. Moreover, the constant term does not affect l_w significantly ($P = 0.172 > 0.05$) with $\alpha = 0.05$ (as shown in Table 9), so it could be deduced that the constant term should be rejected from the regression model. The coefficients of parameter l_w are listed in Table 9.

The coefficients of the constant term and link length are 23.378 and 0.0760, respectively, and the regression equation is given as follows:

$$l_w = 0.706L, \quad R^2 = 0.9793. \quad (3.4)$$

Equation (3.4) suggests that l_w is only affected by L .

When there is no real-time data on queue length, it is possible to utilize an average model to determine the relationship between roll time occupancy and queue length, which may represent the average conditions of a variety of links and provide a method to estimate the queue length in general conditions. In this study, the mean of o_{\min} was used in the twelve cases as the general value, and substitute (3.3) and (3.4) into (3.1), then, the average model for the roll time occupancy can be obtained:

$$o = 0.3056 + \frac{1 - 0.3056}{1 + e^{[-(0.000228L + 0.337r + 0.134)(l - 0.760L)]}}. \quad (3.5)$$

When the link length is 300 m, the number of lanes is 1, and the bus ratio is 5%, the average inflection roll time occupancy (i.e., the corresponding roll time occupancy when l is equal to l_w) is found to be 0.6662.

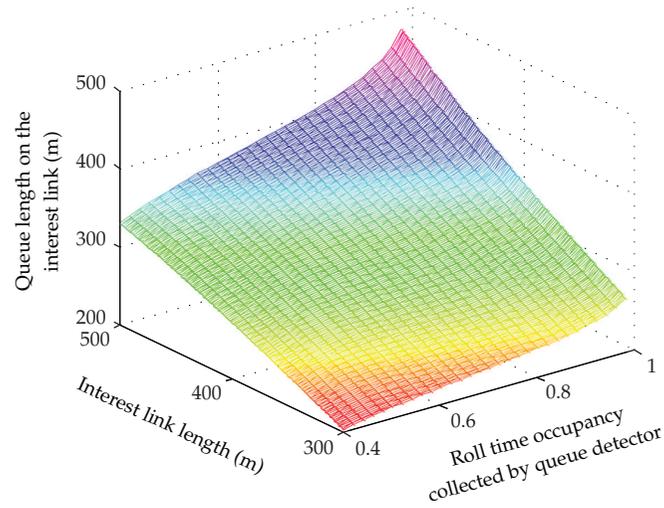


Figure 8: Influence of link length.

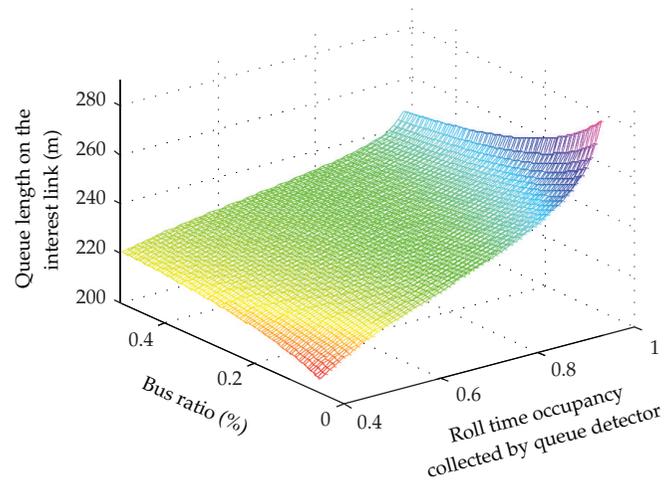


Figure 9: Influence of bus ratio.

4. Queue Length Model with Roll Time Occupancy

An estimation model for queue length using the data of roll time occupancy can be obtained by transforming (3.5), which gives

$$l_i = 0.706L + \frac{1}{0.000228L - 0.337r - 0.134} \ln\left(\frac{1 - o_i}{o_i - 0.3056}\right). \quad (4.1)$$

With known traffic conditions of one road, the influences of link length and bus ratio on the relationship between queue length and roll time occupancy are shown in Figures 8 and 9. (The bus ratio in Figure 8 is 5%; the link length in Figure 9 is 350 m).

The analysis of Figures 8 and 9 shows that

- (1) the relation forms between queue length and roll time occupancy keep invariant under the influence of link length and bus ratio;
- (2) the queue length is directly proportional to the link length when other parameters are constant, and the influence degree will become higher with roll time occupancy becoming greater;
- (3) with lower roll time occupancy, the influence of bus ratio on the relationship is insignificant. The reason may be that lower roll time occupancy represents lower traffic flow, and the difference of occupancy time between buses and cars when they passing through the queue detector is not obvious, which indicates that the occupancy time in every interval is almost uniform;
- (4) high roll time occupancy may be caused by two conditions: (a) the queue length approaches to the link length, and all vehicles pass through the detector with congested speed; (b) if the farthest point of queues is close to the queue detector, the speeds of vehicles passing through the detector should approach to the congested speed, and the average occupancy time per vehicle will become larger under the influence of bus ratio; based on the latter condition, it is concluded that the queue length will decrease with the increasing of bus ratio with an identically higher roll time occupancy (seen as in Figure 9).

Furthermore, in order to validate the queue length estimation model proposed in this paper, another simulation was finished with the same experimental road grid and signal timing scheme except that the bus ratio and interest length link are 8% and 375 m, respectively. The estimated value of queue length can be calculated based on (4.1) and the roll time occupancy collected by VISSIM. The average error of the 3,600 data series is about 23.14%, which indicates the precision of this new model is accepted.

5. Conclusions

In this paper, a method of estimating queue length based on fixed detector outputs was proposed, by focusing on the relationship between the roll time occupancy at a particular position and the queue length on one road. Twelve simulation conditions were designed using different combinations of these four parameters, including link length, lane width, number of lanes, and bus ratio. Based on the analysis of the simulation data, an S-type logistic model was chosen to represent the relationship between queue lengths and roll time occupancy, and equations were fitted for the twelve simulation situations. Then, the average model for roll time occupancy was achieved using successive multiple-stepwise linear regressions with the parameters of the fitting equations and simulation environments. Finally, a new estimation model of queue length based on roll time occupancy was obtained by transforming the average model, which can provide oversaturated arterials a basis to identify spillover conditions.

Considering the effect of traffic signal control, vehicles departed from the upstream intersection will pass through the detector with strong discontinuity and the queue detector would be idled for a long interval, which may bring the result that some lower roll time occupancy corresponds to higher queue length. Furthermore, there may be lane changing near the queue detector. The detector may not be occupied even if the queue length is larger

than the distance between the stop line and the queue detector, which also can lead to lower roll time occupancy corresponding to higher queue length. The estimation accuracy of the method for estimating queue length proposed in this paper can be improved, if the invalid data caused by the two factors mentioned above can be eliminated accurately.

Acknowledgment

This research was supported in part by the National High Technology Research and Development Program of China (no. 2011AA110304).

References

- [1] N. Geroliminis and A. Skabardonis, "Identification and analysis of queue spillovers in city street networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1107–1115, 2011.
- [2] W. W. Guo, W. H. Wang, G. Wets et al., "Influence of stretching-segment storage length on urban traffic flow in signalized intersection," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1401–1406, 2011.
- [3] D. C. Gazis, "Optimal control of a system of oversaturated intersections," *Operations Research*, vol. 12, no. 6, pp. 815–831, 1964.
- [4] M. J. Lighthill and G. B. Whitham, "On kinematic waves. I: flood movement in long rivers II: a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society. London A*, vol. 229, pp. 281–316, 1955.
- [5] G. Stephanopoulos and P. G. Michalopoulos, "Modelling and analysis of traffic queue dynamics at signalized intersections," *Transportation Research A*, vol. 13, no. 5, pp. 295–307, 1979.
- [6] P. G. Michalopoulos and G. Stephanopoulos, "Oversaturated signal systems with queue length constraints-II. Systems of intersections," *Transportation Research*, vol. 11, no. 6, pp. 423–428, 1977.
- [7] R. Akcelik and N. M. Roupail, "Overflow queues and delays with random and platooned arrivals at signalized intersections," *Journal of Advanced Transportation*, vol. 28, no. 3, pp. 227–251, 1994.
- [8] A. Sharma, D. M. Bullock, and J. A. Bonneson, "Input-output and hybrid techniques for real-time prediction of delay and maximum queue length at signalized intersections," *Transportation Research Record*, no. 2035, pp. 69–80, 2007.
- [9] F. V. Webster, *Traffic Signal Settings*, Road Research Technical Paper No. 39, Research Laboratory, London, UK, 1958.
- [10] C. F. Daganzo, "Queue spillovers in transportation networks with a route choice," *Transportation Science*, vol. 32, no. 1, pp. 3–11, 1998.
- [11] R. Akcelik, *A Queue Model for HCM 2000*, ARRB, Melbourne, Australia, 1999.
- [12] G. Vigos, M. Papageorgiou, and Y. Wang, "Real-time estimation of vehicle-count within signalized links," *Transportation Research C*, vol. 16, no. 1, pp. 18–35, 2008.
- [13] H. X. Liu, X. Wu, W. Ma, and H. Hu, "Real-time queue length estimation for congested signalized intersections," *Transportation Research C*, vol. 17, no. 4, pp. 412–427, 2009.
- [14] X. Wu, H. X. Liu, and D. Gettman, "Identification of oversaturated intersections using high-resolution traffic signal data," *Transportation Research C*, vol. 18, no. 4, pp. 626–638, 2010.
- [15] M. Papageorgiou, J. M. Blouffeville, and H. Haj-Salem, "Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part II: Coordinated on-ramp metering," *Transportation Research A*, vol. 24, no. 5, pp. 361–370, 1990.
- [16] J. Walmsley, "The practical implementation of SCOOT traffic control systems," *Traffic Engineering & Control*, vol. 23, no. 4, pp. 196–199, 1982.
- [17] M. J. Norusis, *The SPSS Guide to Data Analysis*, SPSS, Chicago, Ill, USA, 1990.
- [18] J. Pallent, *SPSS Survival Manual*, SPSS, Chicago, Ill, USA, 2001.

Research Article

A Node Model Capturing Turning Lane Capacity and Physical Queuing for the Dynamic Network Loading Problem

Mingxia Gao

School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

Correspondence should be addressed to Mingxia Gao, mxgao@mail.lzjtu.cn

Received 8 May 2012; Revised 30 August 2012; Accepted 30 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Mingxia Gao. 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.

An analytical dynamic node-based model is proposed to represent flows on a traffic network and to be utilized as an integral part of a dynamic network loading (DNL) process by solving a continuous DNL problem. The proposed model formulation has an integrate base to be structured with a link load computing component, where physical queuing and its influence were explicitly taken into account by dividing a link into two parts: running part and queuing part. The solution to the model is obtained by a hybridization algorithm of simulation and analytical approach, where an iteration process is conducted to update time-dependent network flow conditions after a reasonable discretization of the problem. The performance of the proposed model, as a DNL model, is tested on a sample network. It is seen that the proposed model provides consistent approximations to link flow dynamics. The dynamic node model proposed in this paper is unique in that it explicitly models directional queue in each turning lane and the First-In-First-Out (FIFO) rule at lane level rather than link level is pursued.

1. Introduction

The dynamic network loading (DNL) problem is the reproduction of variable link performances and network flow conditions by considering nodal and exiting flow characteristics [1–4] or by explicitly considering the flow propagation and the provision of real time path-link information [5–7], given path flows and link-performance functions. In the past two decades, DNL has been extensively studied owing to the needs of simulating urban traffic and solving dynamic traffic assignment (DTA) problems. The variation in model structure is heavily dependent on both assumptions made to obtain a solution for the problem, that is, the discretization dimension, queuing, and the criteria that affect the computation of link loads and path-link traveling times. One way to categorize the different approaches is discrete-flow

models and continuous-flow models. Discrete-flow models, also referred to as simulations, are usually classified into microscopic simulations and macroscopic simulations. In this paper we concentrate on continuous-flow models which are also referred as “analytic” or “macroscopic” models, and we do so for the usual reasons: low number of parameters to be calibrated, good computational performance, mathematical tractability. Microsimulations are time-based, meaning that individual vehicles are moved in short-time intervals (0.1–1 s) according to car following and other traffic behavior models [8]. The main assumption that is made during the construction of a model to solve DNL problem is on queuing and can be divided into point queuing and physical queuing. Considering the queuing assumption, the models of DNL can be further classified as link based, which vary with the adopted performance function such as link exit function and travel-time-function and node based. Although the performance-function-adopted-link-based modeling framework has been widely used for the analytical formulation of the DNL problem, little research has been conducted to explore their capabilities of reproducing realistic flow dynamics [9–11].

The term node based is generally used for models that explicitly consider the flow splitting rates which are the proportions of traffic leaving a node, assigned to each exiting link, and has been originated from the pioneering proposition of the cell-based traffic flow model in [12]: the cell transmission model (CTM). Node models can be classified as follows according to how link interactions are modeled: (1) competition-free nodes: only the flow conservation law is obeyed at such nodes. The competition-free node model is often seen in the analytical DTA research [13]. (2) Uncontrolled competition nodes: traffic from different incoming links and/or heading to different outgoing links would have to compete against each other for the limit capacities [14]. A typical example is freeway junctions (on- and off-ramps) without metering facilities. (3) Controlled competition nodes: the competition among different traffic streams is managed by a controlled logic, such as signalized intersections. No matter if controlled or uncontrolled competition nodes, the local demand and supply flow concept are utilized to put up an unifying framework for the modeling of intersections simultaneously with the imposition of boundary conditions to various network flow models such as first-order wave models in [12, 15], second-order wave models in [16], models that explicitly incorporate the spatial queuing effect [9], and link performance models [17]. Most of the aforementioned node-based models are proposed to overcome the deficiencies of link-based models and to be utilized in network traffic control and management applications, where merges and diverges excessively break down the stability of traffic flow.

However, some details that are really important and strongly influence loading results have not got satisfactory representation and enough attention. For example, most node models deal with flow propagation through nodes by distributing local demand of upstream links to downstream links constrained by prevailing supply, where the prevailing demand of a link depends on flows waiting to exit and the capacity of the link rather than the capacity of turning lanes. Queuing on an upstream link is treated as a whole rather than capturing turning directions of vehicles or directional lanes they belonging to. To satisfy FIFO at link level, such models are liable to cause unrealistic description of flow dynamics. For example, right-turning vehicles have to wait even if its downstream link has enough supply just because left-turning vehicles cannot cross the intersection for lacking of downstream supply, whether these vehicles (right turning and left turning) are in the same directional turning lane. In real traffic conditions, as pointed in HCM, flows propagating through an intersection are limited by the capacity of directional entry lanes, and drivers generally choose a turning lane according to their turning directions when approaching an intersection, which leads to several directional queues with different flow dynamics even on the same link.

Being cognizant of the insufficiency of link-based models in reproducing actual traffic dynamics and the motivation for node-based models, we have proposed an analytical dynamic node-based model that explicitly considers the influence of directional turning lane and physical queuing in different lanes. In our model, queues in each turning lane are explicitly modeled and FIFO at lane level is pursued. We have chosen to solve the continuous DNL problem analytically with our proposed methodology rather than formulating a theoretical high-order wave model for a node. The proposed node model has an integrate base to be structured with a link load computing component. The constraints of link dynamics, flow conservation, flow propagation, and boundary conditions are considered both in nodal rules and the link model. We obtained the solution of the model by coding a simulation-based hybridization algorithm after designing a discrete version of the problem.

This paper is organized as follows. In Section 2, model description, including the link model and the node model formulation is given. The solution procedure is explained in Section 3. Section 4 holds the numerical results of the solution method that is employed to solve the DNL problem on a sample network. Findings and discussions based on the results obtained conclude this paper as the final section.

2. Model Description

Given path flow and link performance functions, the CDNLP consists of determining time-dependant network flow conditions such as link travel times, link inflows, and link outflows. In the loading procedure, the modeling of traffic flow has two major facets: the representation of traffic dynamics on a link (homogeneous road segment) and on a node (boundary of several links, intersection). We propose an approach that evolves a link model with a node model, where a link is divided into two parts: the running part and the queuing part. The partial length of the running part is dependent on the prevailing traffic condition. Flow dynamics on the running part are described by a Travel Time Function (TTF-) based link model, and flow propagation through intersections is represented with a node model that describes directional queuing in different turning lanes explicitly. Without loss of generality, all links are assumed to have exclusive turning lanes for each turning direction at intersections. Time-dependent exit flow of a directional lane rather than link exit flow is modeled and calculated as the minimum value of prevailing demand and supply of the lane. For links holding mixed turning lanes, exit flow from mixed lanes involves distributing lane demand to downstream links, which can be dealt with using methods adopted in previous node models [18, 19] by taking a mixed lane as a link.

Some basic notations and variables are given first. The physical traffic network can be represented by a directed network $G = (N, A)$, where N is the set of nodes, and A is the set of links. In the following, the index r denotes an origin node, the index s denotes a destination node, the index a denotes a link, the index p denotes a path between the origin (O) and the destination (D), and the index aa^+ denotes a directional turning lane which connects link a and its downstream link a^+ . An O - D pair (r, s) is designated, where $O \subset S, D \subset S, r \in O, s \in D, rs \in R$, and $R \subset (O * D)$. The c of paths between an O - D pair (r, s) is denoted by p_{rs} .

2.1. Link Model

Link models for the DNL problems enable the specification of flow dynamics on a link in three ways: bottleneck models, whole link function models, and hydrodynamic models. In

bottleneck-type models, vehicles always move along a link at free-flow speed before they arrive at the exit node, where they form a FIFO queue if the outflow rate they induce exceeds the maximum discharge rate (bottleneck capacity) of the link [19]. In whole-link function models, a performance function (Travel-Time-based Function or Link Exit Function) is adopted to describe flow dynamics taking the whole link as homogeneous [20]. While hydrodynamic models view traffic as a continuous fluid represented by density, speed, and flow-rate; they are also known as kinematic wave (KW) models because their solutions can be categorized by combinations of kinematic waves in any of the three quantities [21]. We follow the approach of bottleneck models because flow dynamics of vehicles running on homogeneous segments and of those waiting in a queue for exiting should be treated differently. Also, some improvements are made to model congestions on a link segment and to capture the effect of physical queuing. In our approach, a link is divided into two parts: the running part and the queuing part. The prevailing partial length of the running part of link a (denoted with $L'_a(t)$) depends on the number of vehicles on the link and is calculated as below:

$$L'_a(t) = L_a \times \left(1 - \frac{xq_{aa^+}^*(t)}{xr_a(t) + xq_a(t)} \right), \quad (2.1)$$

where $xq_{aa^+}^*(t)$ is the number of vehicles in a turning lane of link a with longest queue at time t ; $xr_a(t)$ and $xq_a(t)$ denote the number of vehicles on the running part and the queuing part of link a at time t .

The traversal time on the running part of link a experienced by vehicles entering at time t (denoted with $dr_a(t)$) is a function of link volume as given in (2.2):

$$dr_a(t) = \frac{L'_a(t)}{v_a(t)}, \quad (2.2)$$

where $v_a(t)$ denotes the travel speed for vehicles entering link a at time t and is calculated with a modified Green-Shields equation [22] as given by (2.3):

$$v_a(t) = v_{a,\min} + (v_{a,f} - v_{a,\min}) \left(1 - \frac{k_a(t)}{k_{a,j}} \right), \quad (2.3)$$

where $v_{a,\min}$, $v_{a,f}$, $k_{a,j}$ denote the minimum speed, free-flow speed, and jam density of link a , respectively. The prevailing density of the running part of link a at time t (denoted with $k_a(t)$) is determined by the following equation:

$$k_a(t) = \frac{xr_a(t)}{L'_a(t) \cdot \text{lan}_a}, \quad (2.4)$$

where lan_a denotes the lane number of link a .

The path-specific exit flow of the running part of link a is determined by (2.5):

$$vr_{rsp}^a(t + dr_a(t)) = \frac{ur_{rsp}^a(t)}{1 + dr_a(t) - dr_a(t-1)}, \quad \forall r, s, t, a, \forall p \in p_{rs}, \quad (2.5)$$

where $ur_{rsp}^a(t)$ and $vr_{rsp}^a(t)$ denote entering flow and exit flow of the running part of link a of $rsO-D$ pair on path p at time t . $ur_{rsp}^a(t)$ can be determined beforehand as below:

$$ur_{rsp}^a(t) = \begin{cases} f_{rsp}(t), & a \text{ is the first link on path } p, \\ vq_{rsp}^{(a^-)a}(t), & \text{otherwise,} \end{cases} \quad a, a^- \in p, \quad (2.6)$$

where a_p^- is the link that precedes link a on path p ; $vq_{rsp}^{a_p^- a}(t)$ is the exit flow from the queuing part of link a^- in turning lane a^-a of $rsO-D$ pair on path p at time t ; $f_{rsp}(t)$ is the departure flow rate on path p between an $rsO-D$ pair departing at time t . Equation (2.6) ensures the FIFO behavior on the running part by forcing vehicles that enter the link at t to be pushed out at $t + dr_a(t)$.

The number of vehicles existing on the running part of link a is updated accordingly as given by (2.7) and (2.8):

$$xr_{rsp}^a(t+1) = xr_{rsp}^a(t) + ur_{rsp}^a(t) - vr_{rsp}^a(t), \quad \forall r, s, a, t, \quad \forall p \in p_{rs}, \quad (2.7)$$

$$xr_a(t) = \sum_{r \in R} \sum_{s \in S} \sum_{\substack{p: a \in p \\ p \in p_{rs}}} xr_{rsp}^a(t), \quad \forall a, t, \quad (2.8)$$

where $xr_{rsp}^a(t)$ denotes the number of vehicles on the running part of link a of $rsO-D$ pair on path p at time t .

2.2. Node Model

We only consider nodes with competition. Most node models presented in previous research deal with flow propagation through nodes by distributing demand of upstream links to downstream links. The demand of a link, D , is the maximum possible exit flow rate, that is,

$$D = \min\{C, Q\}, \quad (2.9)$$

the supply of a link, S , is the maximum possible receiving flow rate, that is,

$$S = \min\{C, R\}, \quad (2.10)$$

where C is the flow (exit) capacity; Q is the rate of flow ready to exit; R is the maximum entry flow rate to the downstream link permitted by the current traffic condition. The partial flow exit from an upstream link entering a downstream link is calculated according to the turning proportion and constrained by prevailing supply. Taking Figure 1 as an example, the queue on link a (or link b) is modeled as a whole, time-dependent partial flow exiting from link a to link c (or d) is calculated as follows: exit flow from link a is determined first with the formula mentioned above, and the partial flow is calculated with the total exit flow from link a multiplying by a turning proportion. To satisfy FIFO rule at link level, such models

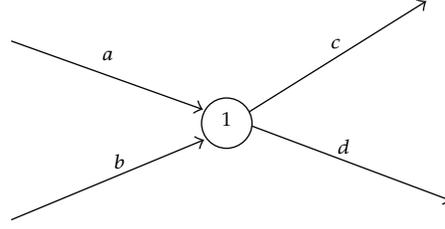


Figure 1: A sample network.

may lead to unnecessary queuing of vehicles. For example, vehicles on link a ready to enter link c will have to wait even if link c has enough supply, just for the reason that link d got a traffic jam and cannot allow any entering flow at current time. In other words, vehicles with different turning directions will have to travel synchronously to satisfy FIFO at link level.

In real traffic conditions, as pointed in HCM, there exist directional turning lanes at intersections especially those with signal controlled, and the capacity of turning lanes rather than that of the whole link actually plays role in limiting flow exiting a link. Drivers generally change or choose turning lanes according to their turning directions when approaching an intersection, which leads to different queues with different flow dynamics in turning lanes. To overcome this shortcoming in describing real dynamics of traffic flow, queuing in different turning lanes should be specifically described, and FIFO at lane level rather than link level should be pursued.

In such situation, the input to the node model is time-dependent entering flow to the queuing part of a link in each turning lane, which can be determined by the exit flow of the running part of the same link according to path flows as given below:

$$uq_{rsp}^{aa^+}(t) = vr_{rsp}^{a,a^+}(t), \quad \forall r, s, t, \forall p \in p_{rs}, a, a^+ \in p, \quad (2.11)$$

$$uq_{aa^+}(t) = \sum_{r \in R} \sum_{s \in S} \sum_{\substack{p: p \in p_{rs} \\ a, a^+ \in p}} uq_{rsp}^{aa^+}(t), \quad \forall r, s, t, \forall a, a^+ : a_h = a_t^+, \quad (2.12)$$

where $uq_{rsp}^{aa^+}(t)$ is the partial entering flow to the queuing part of link a in turning lane aa^+ of $rsO-D$ pair on path p at time t , and $uq_{aa^+}(t)$ is the total entering flow to the queuing part of link a in turning lane aa^+ at time t .

Based on the entering flow, the exit flow of the directional queuing part can be calculated with prevailing demand and supply as following:

$$vq_{aa^+}(t) = \min[S_{aa^+}(t), R_{aa^+}(t)], \quad \forall a, a^+ : a_h = a_t^+, \quad (2.13)$$

where $vq_{aa^+}(t)$ is the exit flow from the queuing part of link a in turning lane aa^+ at time t ; $S_{aa^+}(t)$ is the partial demand that is present from link a to link a^+ at time t and is determined with (2.14):

$$S_{aa^+}(t) = \begin{cases} uq_{aa^+}(t), & \text{if } xq_{aa^+}(t) = 0, uq_{aa^+}(t) < c_{aa^+}, \\ c_{aa^+}, & \text{otherwise,} \end{cases} \quad \forall a, a^+ : a_h = a_t^+, \quad (2.14)$$

where $xq_{aa^+}(t)$ denotes the number of vehicles on the queuing part of link a in turning lane aa^+ at time t ; c_{aa^+} denotes the capacity of turning lane aa^+ on link a .

The partial supply of link a^+ allocated to link a at time t (denoted with $R_{aa^+}(t)$) can be calculated by (2.15), as shown below:

$$R_{aa^+}(t) = \beta_{aa^+} \cdot R_{a^+}(t), \quad \forall a, a^+ : a_h = a_t^+,$$

$$R_{a^+}(t) = \begin{cases} s_{a^+}, & \text{if } x_{a^+}(t) < h_{a^+} \\ vq_{a^+}(t), & \text{otherwise} \end{cases} \quad \forall a, a^+ : a_h = a_t^+, \quad (2.15)$$

where β_{aa^+} is the proportionality coefficient which depends on the lane number and control mode of an intersection and treated as a constant; h_{a^+} is the maximum number of vehicles that link a^+ can accommodate; s_{a^+} is the entering capacity of link a^+ ; a_h is the head node of link a ; a_t^+ is the tail node of link a^+ .

With partial exit flow of each turning lane known, the total exit flow of the queuing part of link a (denoted with $vq_a(t)$) can be determined by (2.16) as given below:

$$vq_a(t) = \sum_{a^+ : a_h = a_t^+} vq_{aa^+}(t), \quad \forall a. \quad (2.16)$$

The relationships between disaggregated and aggregate variables are calculated as follows:

$$vq_{rsp}^{aa^+}(t) = \begin{cases} \frac{xq_{rsp}^{aa^+}(t)}{xq_{aa^+}(t)} \cdot vq_{aa^+}(t), & \text{if } xq_{aa^+}(t) > 0, \\ \frac{uq_{rsp}^{aa^+}(t)}{uq_{aa^+}(t)} \cdot vq_{aa^+}(t), & \text{otherwise.} \end{cases} \quad \forall r, s, t, \forall p \in p_{rs}, \quad (2.17)$$

The splitting rate $\lambda_{aa^+}^p$ for the flow that is exiting link a in turning lane aa^+ along path p is calculated as given by (2.18). The constraint associated with $\lambda_{aa^+}^p$ is given by (2.19), which expresses that at turning lane level, the FIFO rule holds.

$$\lambda_{aa^+}^p = \begin{cases} \frac{xq_{rsp}^{aa^+}(t)}{xq_{aa^+}(t)}, & \text{if } xq_{aa^+}(t) > 0, \\ \frac{uq_{rsp}^{aa^+}(t)}{uq_{aa^+}(t)}, & \text{otherwise,} \end{cases} \quad p \in p_{rs}, a, a^+ \in p, \quad (2.18)$$

$$\sum_{p \in p_{rs}} \lambda_{aa^+}^p = 1. \quad (2.19)$$

The number of existing vehicles at time t , specified to path, turning lane(s), and queuing part, is updated as follows:

$$xq_{rsp}^{aa^+}(t+1) = xq_{rsp}^{aa^+}(t) + uq_{rsp}^{aa^+}(t) - vq_{rsp}^{aa^+}(t), \quad \forall r, s, t, \forall p \in p_{rs}, a, a^+ \in p, \quad (2.20)$$

$$xq_{aa^+}(t) = \sum_{r \in R} \sum_{s \in S} \sum_{\substack{p: p \in p_{rs} \\ a, a^+ \in p}} xq_{rsp}^{aa^+}(t), \quad \forall a, a^+ : a_h = a_t^+, \quad (2.21)$$

$$xq_a(t) = \sum_{a^+ : a_h = a_t^+} xq_{aa^+}(t), \quad \forall a. \quad (2.22)$$

3. Solution Procedures

The solution procedures sought for the DNL problem can be clustered into three approaches: (1) an analytical approach, (2) a simulation-based approach, and (3) a hybridization of these two approaches. The DNL problem solved with an analytical approach possesses the desired properties of both the solution and the link performance function [23]. The simulation-based approach, that is, INTEGRATION, DYSMART, and other propositions, provides more flexibility and consequently enables the reproduction of detailed flow dynamics both in dealing with more complex traffic patterns and in obtaining more realistic loading results lacking the desired solution properties [24]. The hybrid approach set out by the incorporation of an analytical formulation with simulation and utilized in several studies, including [12, 22, 25] and this paper, has become an efficient alternative in network performance modeling. The feature that it simultaneously possesses both ability of analytical models in obtaining more accurate solutions and the ability of simulation models in obtaining computationally efficient solutions in network traffic modeling context has motivated us to employ a hybrid approach in this paper.

The analysis period is divided into small time intervals with same length σ . The interval σ is treated as unit of time, and time period that is not an exact multiple of σ involved in loading procedure is processed with linear approximation. Given path flows during *O-D* traffic period, we can get entering flow to the running part of each link by (2.6).

The following algorithm summarizes the determination of flows and the number of vehicles on all links by the proposed node-based modeling approach when a network is considered.

- (1) Set $t = 0$ (first time interval), clear the network, that is,

$$xr_{rsp}^a(t) = 0, \quad xq_{rsp}^{aa^+}(t) = 0, \quad \forall r, s, t, a, p, \forall a^+ : a_t^+ = a_h. \quad (3.1)$$

- (2) Determine the number of vehicles on the running part and the directional queuing part on each link with (2.8) and (2.22), respectively.
- (3) Determine the disaggregate inflow $ur_{rsp}^a(t)$ to running parts of links with (2.6).
- (4) Calculate the travel time $dr_a(t)$ of each link with (2.1)–(2.4) and obtain outflow $vr_{rsp}^a(t + dr_a(t))$ by (2.5).
- (5) Calculate the disaggregate number of vehicles $xr_{rsp}^a(t)$ with (2.7).
- (6) Obtain the disaggregate inflow $uq_{rsp}^{aa^+}(t)$ with (2.11).
- (7) Calculate the aggregate inflow $uq_{aa^+}(t)$ with (2.12) and calculate the number of vehicles in each turning queue $xq_{aa^+}(t)$ with (2.21).
- (8) Obtain $vq_{aa^+}(t)$ and $vq_a(t)$ by (2.13)–(2.16).

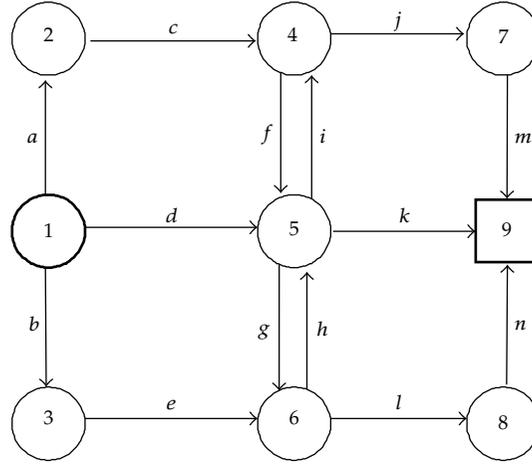


Figure 2: Sample network used for testing trials.

- (9) Determine the disaggregate variable $vq_{rsp}^{aa^+}(t)$ and $xq_{rsp}^{aa^+}(t)$ with (2.17) and (2.20), respectively.
- (10) If the demand is discharged from all the paths during the analysis period, stop; otherwise, set $t =$ (following time interval) and go to step (2).

4. Numerical Implementations

4.1. Sample Network

The performance of the proposed model is tested on a sample network shown in Figure 2 with several paths between the given $O-D$ pair. The link characteristics assigned are given in Table 1, and the characteristics of turning lane(s) at the entrance of each intersection are given in Table 2. The traffic flow interval, σ , is 1 min, and the departure period, T , is 30 min and is divided into 5 intervals with same length. There is only one $O-D$ pair and seven routes used in this network. Route departure flows, $f_{rsp}(t)$, have a constant value during each departure interval, given in Table 3 and $f_{rsp}(t) = 0$ at other times.

4.2. Test Results

The performance of the proposed model is evaluated with a number of critical terms from the simulation. The outflow diagram of flow propagation at node 5 is given in Figure 3. The dynamic outflow and queue length diagrams in each turning lane on link d are given in turn in Figures 4 and 5.

In Figure 3, it is seen that the loadings to link k required longer time, due to a longer flow profile of paths involving this link. It can be seen from Figure 4 that outflows of different turning lanes on link d appear to vary in similar manners, which reach saturation in a few minutes and keep the state for some time (41 minutes in the right turn lane, 52 minutes in the left turn lane, and 69 minutes in the through lane) before a sharp decline. From Figure 5,

Table 1: Assigned link characteristics of sample network.

Link number	Free-flow speed v_f (km/h)	Minimum speed v_{\min} (km/h)	Link length l (m)	Lane number lan	Capacity rc_a (pcu/h)	Jam density k_j (pcu/km/lan)
<i>a</i>	30	12	980	2	1900	140
<i>b</i>	32	10	950	2	1800	140
<i>c</i>	30	5	800	2	1700	140
<i>d</i>	36	15	1750	3	2600	140
<i>e</i>	30	12	900	2	1850	140
<i>f</i>	30	10	860	2	1600	140
<i>g</i>	28	10	750	2	1850	130
<i>h</i>	30	10	950	2	1900	130
<i>i</i>	28	13	800	2	1850	135
<i>j</i>	30	15	830	2	2000	150
<i>k</i>	35	15	1250	2	2050	153
<i>l</i>	30	15	850	2	1700	160
<i>m</i>	32	14	900	1	1100	150
<i>n</i>	30	12	880	2	2200	145

Table 2: Capacities of turning lanes at intersection entrance.

Link number	Turn direction and downstream link	Capacity (pcu/h)
<i>a</i>	Right turn $\rightarrow c$	1500
<i>b</i>	Left turn $\rightarrow e$	1600
<i>c</i>	Left turn $\rightarrow f$	900
	Through $\rightarrow j$	950
	Right turn $\rightarrow g$	900
<i>d</i>	Left turn $\rightarrow i$	305
	Through $\rightarrow k$	315
<i>e</i>	Left turn $\rightarrow h$	329
	Through $\rightarrow l$	880
<i>f</i>	Left turn $\rightarrow k$	265
	Through $\rightarrow g$	507
<i>g</i>	Left turn $\rightarrow l$	907
<i>h</i>	Right turn $\rightarrow k$	917
	Through $\rightarrow i$	362
<i>i</i>	Right turn $\rightarrow j$	1700
<i>j</i>	Right turn $\rightarrow m$	1800
<i>l</i>	Left turn $\rightarrow n$	1750

Table 3: Route departure flow rates in 30 minutes.

Route	Departure interval				
	[0, 5]	[6, 11]	[12, 17]	[18, 23]	[24, 30]
(1) <i>d-k</i>	600 (pcu/h)	720	730	660	0
(2) <i>d-i-j-m</i>	450	540	520	450	0
(3) <i>b-e-h-k</i>	600	720	720	710	550
(4) <i>b-e-l-n</i>	900	1080	1080	1080	870
(5) <i>d-g-l-n</i>	1090	1320	1320	1130	0
(6) <i>a-c-f-k</i>	410	480	510	510	50
(7) <i>a-c-j-m</i>	1100	1320	1320	1320	140

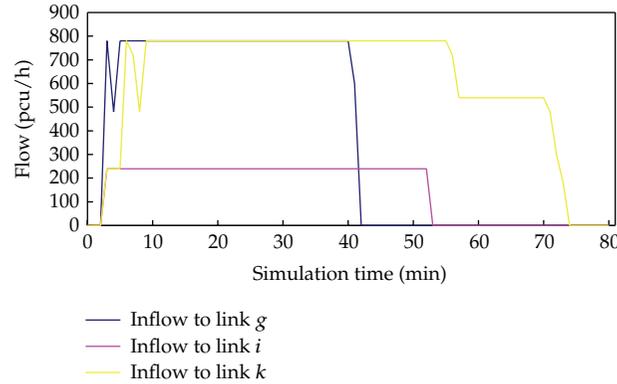


Figure 3: Outflows at node 5 of sample network.

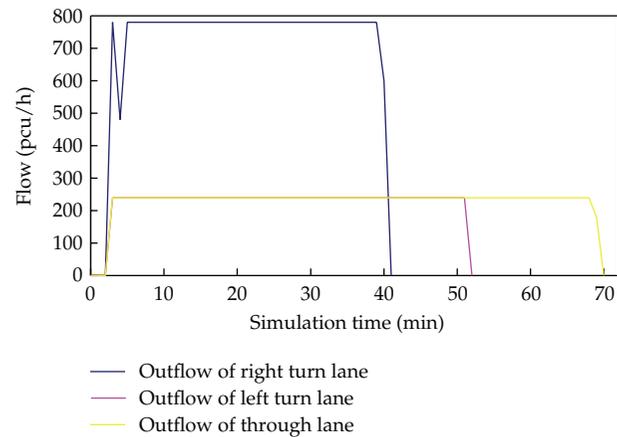


Figure 4: Outflows from turning lane(s) of link *d*.

we can see that the number of queuing vehicles in each turning lane increases to a maximum value after the entrance lanes got saturation.

4.3. Comparison of Different Models

In this part, link *d* was implemented using other models for comparison. Travel-Time-Function-based link models (TTF) that has been used in [22], node model pursuing FIFO at link level (for simplicity named NM-1) adopted in [18], and node model presented in this paper (for simplicity named NM-2) are tested. In TTF-based link model, a link was treated as a whole, and the travel speed for a vehicle entering link *a* at time *t* was calculated with the modified Green-Shields model shown in (2.6). The difference is that the traversal time was calculated with the total length of link *a*, not the partial length of the running part, divided by the speed at time *t*. NM-1 has a similar framework with NM-2, in which a link is divided into two parts and the flow dynamics on running part were described with a travel time function. It is the description of flow dynamics on queuing parts that makes NM-2 different from NM-1. In NM-1, time-dependent exit flow of a whole link but not a turning lane is calculated with

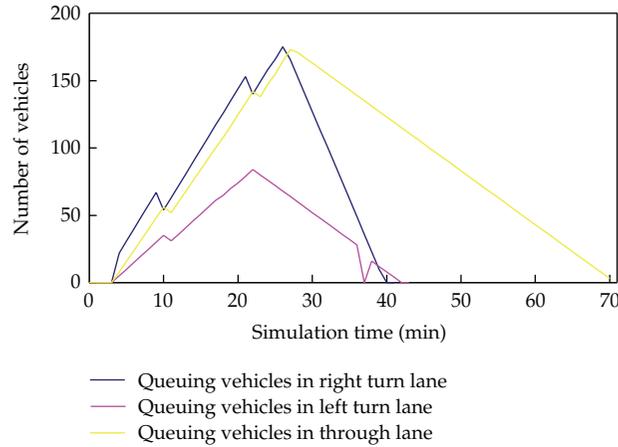


Figure 5: Queue length in turning lane(s) of link d .

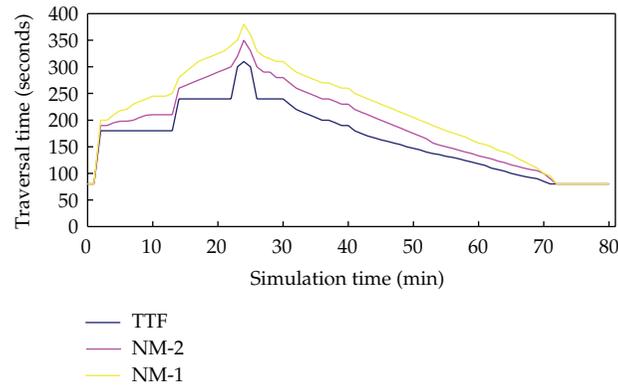


Figure 6: Time-dependent traversal times on link d in different scenarios.

its prevailing demand and supply; queuing vehicles with different directions are described as mixed together; the prevailing demand waiting to exit is limited by the capacity of the link, rather than specified to turning lanes.

Figure 6 compares the time-dependent link traversal times on link d produced by the three models mentioned above. As shown, NM-1 overestimated the congestion on link d , while TTF led to a slight underestimation. The reason lies in that the TTF-based link model ignores queuing delay before intersections, and NM-1, pursuing FIFO at link level, may lead to unnecessary queuing of some vehicles.

5. Conclusion

In this paper, an analytical node-based model has been proposed for the continuous dynamic network loading problem (CDNLP). Taking account of a set of analytical rules, an algorithm has been derived. The new dynamic node model proposed in this paper is unique in that it explicitly captures the turning lane capacity and directional physical queuing in the description of traffic propagation through nodes.

In the proposed node model, directional queue in each turning lane is explicitly considered and the FIFO rule at lane level rather than link level is pursued. A directional exit flow function is presented to update the exit flow from each turning lane with the constraints of lane capacity, downstream link capacity, and flow conservation, and so forth. A travel time function (TTF-) based link model has been evolved to a node model in the modeling structure. To capture queue spillback caused by the capacity of turning lanes as well as the downstream link, the concept of prevailing partial length is adopted, which tells the length of partial link that can be used by running vehicles and is determined by the number of running vehicles as well as queuing vehicles. The time to travel the running part of a link equals to the partial length divided by the prevailing speed calculated with a modified Green-shields formulation and is common to vehicles entering the link at the same time. An algorithm is presented for the CDNL problem with our node-based modeling approach, which is based on simulation and updates the performance of running parts as well as queuing parts in order of time step.

The drawback is that the node model integrates a priori mean effects on flows of a traffic signal without explicitly representing the alternation of green and red stages, and traffic wave can not be tracked, for the running part of a link is treated as a homogeneous road segment.

The proposed node-based model can be easily integrated as a flow modeling component of a dynamic traffic-assignment process, enabling its utilization in a wide range of intelligent transportation system applications. For example, the realistic representation of traffic flow dynamics enables the proposed model to be easily utilized in advanced traveler information systems (ATISs). The prediction on link performances can be obtained in terms of real-time flow volume data inputs. These predictions can be basic inputs to ATIS applications, such as variable message signs for route guidance. Dynamic signal optimization and ramp metering are other possible topics that the proposed model's extensions can study for capacity management and speed regulation.

Acknowledgments

The author would like to thank the supports by the New Teacher Fund for Doctor Station of the Ministry of Education with no. 20116204120005, by Program of Humanities and Social Science of Education Ministry of China for Western and Frontier under Grant no. 12XJCZH002 and 11XJC630009.

References

- [1] J. H. Wu, Y. Chen, and M. Florian, "The continuous dynamic network loading problem: a mathematical formulation and solution method," *Transportation Research Part B*, vol. 32, no. 3, pp. 173–187, 1998.
- [2] Y. W. Xu, J. H. Wu, M. Florian, P. Marcotte, and D. L. Zhu, "Advances in the continuous dynamic network loading problem," *Transportation Science*, vol. 33, no. 4, pp. 341–353, 1999.
- [3] M. Blumberg and H. Bar-Gera, "Consistent node arrival order in dynamic network loading models," *Transportation Research Part B*, vol. 43, no. 3, pp. 285–300, 2009.
- [4] J. M. Rubio-Ardanaz, J. H. Wu, and M. Florian, "Two improved numerical algorithms for the continuous dynamic network loading problem," *Transportation Research Part B*, vol. 37, no. 2, pp. 171–190, 2003.

- [5] V. Astarita, "A continuous time link model for dynamic network loading based on travel time function," in *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pp. 79–102, Lyon, France, 1996.
- [6] H. Bar-Gera, "Continuous and discrete trajectory models for dynamic traffic assignment," *Networks and Spatial Economics*, vol. 5, no. 1, pp. 41–70, 2005.
- [7] R. Jayakrishnan, H. S. Mahmassani, and T. Y. Hu, "An evaluation tool for advanced traffic information and management systems in urban networks," *Transportation Research Part C*, vol. 2, no. 3, pp. 129–147, 1994.
- [8] Q. Yang and H. N. Koutsopoulos, "A microscopic traffic simulator for evaluation of dynamic traffic management systems," *Transportation Research Part C*, vol. 4, no. 3, pp. 113–129, 1996.
- [9] V. Astarita, M. Florian, M. Mahut, and J. H. Wu, "Modeling the spill-back of congestion in link based dynamic network loading models: a simulation model with application," in *Proceedings of the 14th International Symposium on Transportation and Traffic Theory*, pp. 55–573, Jerusalem, 1999.
- [10] M. Carey and Y. E. Ge, "Convergence of a discretised travel-time model," *Transportation Science*, vol. 39, no. 1, pp. 25–38, 2005.
- [11] X. Nie and H. M. Zhang, "A comparative study of some macroscopic link models used in dynamic traffic assignment," *Networks and Spatial Economics*, vol. 5, no. 1, pp. 89–115, 2005.
- [12] C. F. Daganzo, "The cell transmission model, part II: network traffic," *Transportation Research Part B*, vol. 29, no. 2, pp. 79–93, 1995.
- [13] D. K. Merchant and G. L. Nemhauser, "Model and an algorithm for the dynamic traffic assignment problems," *Transportation Science*, vol. 12, no. 3, pp. 183–199, 1978.
- [14] C. F. Daganzo, "Properties of link travel time functions under dynamic loads," *Transportation Research Part B*, vol. 29, no. 2, pp. 95–98, 1995.
- [15] J. P. Lebacque, "The Godunov scheme and what it means for first order traffic flow models," in *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pp. 647–677, Lyon, France, 1996.
- [16] S. Mammar, J. P. Lebacque, and H. Haj Salem, "Intersection modeling based on the second order traffic flow model (ARZ)," in *Proceedings of the 11th EWGT Meeting*, pp. 435–440, 2006.
- [17] H. B. Celikoglu, "A dynamic node model with a quadratic travel time function based link model," in *Proceedings of the 11th EWGT Meeting*, pp. 658–666, 2006.
- [18] H. B. Celikoglu, E. Gedizlioglu, and M. Dell'Orco, "A node-based modeling approach for the continuous dynamic network loading problem," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 165–174, 2009.
- [19] H. M. Zhang and Y. Nie, "Modeling network flow with and without link interaction: properties and implications," in *Proceedings of the 84th annual meeting of the Transportation Research Board*, Washington, DC, USA, 2006.
- [20] M. Carey and M. McCartney, "Behaviour of a whole-link travel time model used in dynamic traffic assignment," *Transportation Research Part B*, vol. 36, no. 1, pp. 83–95, 2002.
- [21] C. F. Daganzo, "A variational formulation of kinematic waves: solution methods," *Transportation Research Part B*, vol. 39, no. 10, pp. 934–950, 2005.
- [22] R. Jayakrishnan, W. K. Tsai, and A. Chen, "A dynamic traffic assignment model with traffic-flow relationships," *Transportation Research Part C*, vol. 3, no. 1, pp. 51–72, 1995.
- [23] B. W. Wie, R. L. Tobin, and M. Carey, "The existence, uniqueness and computation of an arc-based dynamic network user equilibrium formulation," *Transportation Research Part B*, vol. 36, no. 10, pp. 897–918, 2002.
- [24] H. B. Celikoglu and M. Dell'Orco, "Mesoscopic simulation of a dynamic link loading process," *Transportation Research Part C*, vol. 15, no. 5, pp. 329–344, 2007.
- [25] A. K. Ziliaskopoulos, "Linear programming model for the single destination System Optimum Dynamic Traffic Assignment problem," *Transportation Science*, vol. 34, no. 1, pp. 37–49, 2000.

Research Article

Modeling of Signal Plans for Transit Signal Priority at Isolated Intersections under Stochastic Condition

Lv Bin

School of Traffic and Transportation, Lanzhou Jiaotong University, Gansu, Lanzhou 730070, China

Correspondence should be addressed to Lv Bin, tylbxx@163.com

Received 10 May 2012; Revised 30 August 2012; Accepted 30 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Lv Bin. 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.

Transit signal priority (TSP) is recognized as having the potential to improve transit service reliability at small cost to general traffic. The popular preference for TSP encounters the challenges of various and challenging test scenarios. According to the stochastic characteristics of traffic flow, the signal timing model was established for TSP at an isolated signal intersection, where the passenger average delay was used as the optimization objective, and the weights of all phases were considered. The priority logic that is considered in the study provides cycle length and green time within a fixed-time traffic signal control environment. Using the Gauss elimination, the quantitative relationships were determined between phase clearance reliability (PCR), cycle length, and green time. Simulation experiments conducted by the particle swarm optimization (PSO) algorithm indicated that (1) the random variation of arrival rate has an obvious effect on traffic signal settings; (2) the proposed TSP model can reduce passenger delays, especially under stochastic traffic flow.

1. Introduction

Transit signal priority (TSP) strategies will have the potential improvements in transit schedule reliability, which have been recognized to speed up transit service in common. The TSP strategies can also be categorized into two groups: optimization based and logic rule based [1]. The existed TSP strategies, which are used at isolated signalized intersections, focus on intersections primarily without exclusive bus lanes and operate with an acyclic manner often [2, 3].

Yagar and Han [4, 5] analyzed the influences that passengers' getting on and off transits at transit stations have on the traffic stream of intersections. In order to optimize transit and street cars delay, a signal timing optimization model was built by taking

the influences of passengers' getting on and off transits into consideration. According to the changes of such parameters as passenger amount, on time performance of transits, transit departure frequency, and so forth, Mirchandani [6] established a transit priority weight model in order to optimize signal timing parameters.

Aimed at passenger average delay, Yang et al. [7] founded an optimal liner planning model of transit signal priority control system at isolated intersections, thus realizing preferential control passive transit signals at isolated intersections. Sunkari et al. [8] proposed another technique that used the delay equation from the 1985 highway capacity manual (HCM). The HCM delay equation was applied to the no-TSP signal timings and to signal timing plans associated with four TSP cases, namely, maximum and minimum green extension and maximum and minimum early green.

With the purpose of optimizing phase sequence and time of green light, Ma et al. [9] have designed priority control point number model of transit signals with different departure frequencies and consider to optimize signal timing with the aim of minimizing average delay of transits when the departure frequency is of integral multiple of the length of the cycle. Bie et al. [10] analyzed the relations of the length of the cycle, average passenger, and transit delay indicators at intersections, determined the method to calculate the cycle length, and taking the minimization of average passenger delay as a goal for optimization, brought forward green signal ratio optimization method according to surplus green light time allocated to passenger amount of each phase under the precondition of no traffic jam. Abdy and Hellenga [11] proposed TSP model that is developed on the basis of deterministic queuing theory. In developing the proposed model, analytical expressions are developed assuming that the intersection modeled as a D/D/1 queuing system.

In Polus' study [12], the observed travel time of bus on links was found to follow Beta distribution, whose parameters were estimated and found linearly related with each other and with the link length in the off peak hours. Because of the influence of bus dwell time, number of stops, and passenger activities (i.e., arrival, boarding, and alighting), the dispatching headway of bus was assumed gamma distributed [13]. The fact that a design of timing plan of test intersections, typically regardless of the stochastic characteristics of traffic flow and the demands of passenger, necessarily makes test scenarios less challenging may be the reason why timing parameters was found having an insignificant impact on the performance of TSP strategies [14]. Moreover, the traditional measures of effectiveness, such as delay, stops, travel time, transit schedule reliability, and fuel assumption [14], are fully vehicle oriented and the interests of passenger are neglected to a great extent.

This paper attempts to analyze the impact of timing parameters on the development and performance of an optimization-based transit signal priority, which is for use at isolated signalized intersections with two approaches busway. The study differs from previous studies in two ways. First, the attribute for passenger average delay is introduced to analyze the principles of timing parameters. Second, the stochastic characteristics of traffic flow were analyzed, and a model was established for TSP at an isolated signal intersection.

2. Model Building

The essence of transit priority is to demonstrate care for people. In optimizing the signal timing parameters of transits, we should consider the changing tendency of both delays of street cars and average passenger delays at intersections as well [15]. Therefore, we assume that the total delays at the approach of intersection are equal to the product of average

vehicle delay and the number of passengers in all vehicles at this approach, and that the total delays of pedestrians at the intersection are equal to the sum of delays of pedestrians at all approaches.

Suppose that the average number of passengers carried by transits is P_b and that of street cars is P_s , then the total delays of people in one signal cycle D_p will be

$$D_p = \sum_{i=1}^n \sum_{j=1}^{m_i} D_{ij}^p = C \sum_{i=1}^n \sum_{j=1}^{m_i} (d_{ij}^s Q_{ij}^s P_s + d_{ij}^b Q_{ij}^b P_b), \quad (2.1)$$

where D_{ij}^p is the total passengers delay of j approach of i phase in one cycle (s); C is cycle length at a signal intersection (s); d_{ij}^s , and d_{ij}^b are the delays of street cars and transits at the j approach of i phase (s). Q_{ij}^s is the arrival rate of street cars at the j approach of i phase (pcu/s); Q_{ij}^b means the arrival rate of transits at the j approach of i phase (veh/s); m_i signifies the amount at the approach of i phase.

Define d_p^i as the passenger average delay at the i phase. This value is related to the arrival rate of street cars Q_{ij}^s , that of transits Q_{ij}^b at each approach related to the i phase and their respective passenger carrying numbers P_b and P_s :

$$d_p^i = \frac{D_p}{C \sum_{j=1}^{m_i} (Q_{ij}^s P_s + Q_{ij}^b P_b)}. \quad (2.2)$$

According to the definition above, a signal timing optimization model can be built with regard to the optimization of transit priority signal timing at isolated intersection under stochastic conditions:

$$\min \sum_{i=1}^n u_p^i d_p^i, \quad (2.3)$$

$$\text{s.t. } P\{v_i C \leq g_i s_i\} = \alpha_i, \quad \forall i, \quad (2.4)$$

$$\sum_{i=1}^n g_i + L = C, \quad (2.5)$$

$$L = \sum_{i=1}^n (L_s + I_i - A), \quad (2.6)$$

$$0 \leq g_{\min}^i \leq g_i, \quad (2.7)$$

where n is the number of phase at signal intersections; u_p^i is the weight factor of passenger average delay of i phase and can be determined according to the grade of road or the arrival rate of transits; v_i , s_i , and g_i are the arrival rate (pcu/s), the saturated flow rate (pcu/s) of critical lane, and the green time (s) of i phase, respectively. C is the cycle length (s); α_i is the phase clearance reliability (PCR) at i phase; PCR is the probability that traffic could be entirely discharged within the available green time of the i phase; L is the total lost time of each cycle (s); A and L_s are the yellow time (s), the startup lost time (s) of each phase. I_i is

the intergreen interval time (s) of i phase. g_{\min}^i is the minimum green time of i phase which should meet pedestrians crossing the street and traffic safety (s).

In this model, the goal of the objective function (2.3) is to minimize the summation of passenger average delay of each phase, and the value of passenger average delay of each phase is correlative to PCR (α_i).

The constraint condition (2.4) is a PCR constraint that should be satisfied by the i phase [16]. The PCR could reflect the actual traffic running of each phase, and it could also be used to evaluate whether signalized timing parameters of each phase could meet traffic demand. Based on this, PCR would be used to calculate and analyse the reliability of signalized intersection comprehensively. PCR associates with v_i and s_i in a given timing plan at an intersection. It is stated that the arrival has no relation with saturated flow rate in previous researches, and the arrival rate and the capacity are independent from each other. Therefore, the saturated flow rate is assumed constant. If the arrival rate is constant, then PCR is constant. When the arrival rate is random variable, the PCR varies with arrival rate within the specified period. To make the state more stable, or get the higher value of the PCR at an intersection, it is necessary to consider the optimization of signal timing parameters, and all of them are affected by the arrival rate.

Equation (2.5) means that the sum of green time of each phase and time loss of each cycle equals the cycle length; (2.6) stands for the total time loss of cycle signals constituted by the time loss of each phase; (2.7) refers to the minimum green light time constraint that should be satisfied by phase green time for the safe intersection of pedestrians and motor vehicles.

The research team conducted a traffic investigation at a signal intersection in Lanzhou City. The results and data analysis indicated that the traffic flow arrival rate was assumed to be normal distribution (normal distribution can achieve good effects when fitting the traffic data of the various directions, the various vehicle models, and the various approaches of heavy traffic congested roads and intersections. Therefore, if the valid model testing results cannot be affected, let the random variable v_i (road vehicle arrival rate of critical approach) of i phase be subject to normal distribution, and its distribution parameter is (μ_i, σ_i^2) . Using the Gauss elimination, the model can be transformed into

$$\begin{aligned} & \min \sum_{i=1}^n \mu_p^i d_p^i \\ & \text{s.t.} \begin{cases} g_1 = \frac{L \prod_{i=2}^n B_i}{\prod_{i=1}^n B_i - \prod_{i=2}^n B_i - \sum_{k=2}^n \left(\prod_{i=1, i \neq k}^n B_i \right)}, & g_i = \frac{g_1 B_1}{B_i}, \\ C = \frac{g_i s_i}{\mu_i + \sigma_i \cdot \Phi^{-1}(\alpha_i)}, \\ L = \sum_{i=1}^n (L_s + I_i - A), \\ 0 \leq g_{\min}^i \leq g_i, \end{cases} \end{aligned} \quad (2.8)$$

where $B_i = s_i / (\mu_i + \sigma_i \cdot \Phi^{-1}(\alpha_i))$.

According to different specific control targets, for example, different grades of roads at each direction at the intersections and the amount of transit flows, we should consider using

road directions with relatively higher grade and maximize the traffic capacity at phases with large transit flow. We can reflect the above requirement by reasonably setting the value of u_p^i in the model.

3. Solution Algorithm

Particle swarm optimization (PSO) algorithm is an evolutionary computation technique. It optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. PSO optimizes a problem by having a population of candidate solutions and moving these particles around in the search space according to simple mathematical formulae over the particle's position and velocity. Each particle's movement is influenced by its local best known position and is also guided toward the best known positions in the search space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions.

PSO is originally attributed to Kennedy [17] and was first intended for simulating social behavior, as a stylized representation of the movement of organisms in a bird flock or fish school. The algorithm was simplified and it was observed to be performing optimization. PSO is a metaheuristic as it makes few or no assumptions about the problem being optimized and can search very large spaces of candidate solutions. More specifically, PSO does not use the gradient of the problem being optimized, which means PSO does not require that the optimization problem be differentiable as is required by classic optimization methods, such as gradient descent and quasi-Newton methods. PSO can therefore also be used on optimization problems that are partially irregular, noisy, change over time, and so forth. PSO algorithm displays apparent advantages in solving stochastic expected value model. Hence, this paper will throughout use it in model solution.

3.1. PSO Algorithm

Every particle in the particle swarm will fly in the searched space at certain speed $V_i = (V_{i1}, V_{i2}, \dots, V_{id})$ in the n dimension space. First, those particles will randomly generate a group of initial solutions x_1, x_2, \dots, x_N , where N represents the number of particles. Adjust the dynamic location and speed according to the flying experience of particles in solution space and the group, and use the adaptive value to evaluate the merits of solutions. Choose current individual extremum P_{best} , global extremum G_{best} , and record their locations. Update the speed and location of the next generation of particles according to formula (3.1) to (3.3) and continue the next iteration. The equation of updating is

$$V_{id} = \beta \times V_{id} + a_1 \times \text{rand}() \times (P_{id} - X_{id}) + a_2 \times \text{rand}() \times (P_{gd} - X_{id}), \quad (3.1)$$

$$V_{id} = V_{\max}, \quad \text{if } V_{id} > V_{\max}, \quad (3.2)$$

$$V_{id} = -V_{\max}, \quad \text{if } V_{id} < -V_{\max},$$

$$X_{id} = X_{id} + V_{id}, \quad (3.3)$$

where β is inertia weight which can maintain the kinetic inertia of particles and enable them to explore new areas. a_1 and a_2 stand for positive acceleration constant which is usually

made 2, and they can make every particle move towards P_{best} and G_{best} in accelerated motion. $\text{rand}()$ is the random number distributed equally on $[0, 1]$ and can be used to imitate random disturbances of group behaviors in nature. P_{id} and P_{gd} are the d dimension component of individual and overall extremum value, respectively.

Equation (3.2) constraints the largest speed of particles: if the accelerated speed of the particle at present will make one of its dimension speed component V_{id} exceed this dimension's largest speed V_{max} , then the speed of this dimension will be limited as V_{max} which determines the searching precision of a particle in the solution space. If its V_{max} is too large, a particle is prone to fly past the optimal solution; otherwise, it will probably be trapped in partial searching space, unable to realize overall searching. If the searching space of the issue is limited within $[-X_{\text{max}}, X_{\text{max}}]$, then we can assume that $V_{\text{max}} = kX_{\text{max}}$, $0 \leq k \leq 1$.

3.2. Procedures of Algorithm

By combining stochastic simulation and PSO algorithm, we can arrive at the solving algorithm.

- (1) Initialize the particle group in n dimension problem space: suppose the size of the group is popsize , a random number will appear in the feasible region of green light time g_i at the i phase; use the expected value estimation method of stochastic stimulation to compute and test the feasibility of this random number, namely, judging whether g_i can satisfy $B_1 - B_1B_n + B_n \neq 0$ and $0 \leq g_{\text{min}}^i \leq g_i$; repeat this process for popsize times and popsize initial feasible particles will be obtained: $\mathbf{g}_j = (g_{j1}, g_{j2}, \dots, g_{jm})$, $j = 1, 2, \dots, \text{popsize}$; initialize the speed parameter.
- (2) Compute the adaptive value, namely, $u_p^i d_p^i$, of every particle with the expected value estimation method of stochastic stimulation.
- (3) Compare the adaptive value of each particle with that at the best position experienced by it; if the former is better, then take it as the best position at present.
- (4) Compare the adaptive value of each particle with the best adaptive value experienced by the particle throughout the whole situation; if the former is better, then take it as the best position of the whole situation at present.
- (5) Evolve according to the updating (3.1), (3.2), and (3.3).
- (6) Compute $u_p^i d_p^i$ of particles updated with the expected value estimation method of random stimulation and test the feasibility of particles; if the feasibility is acceptable, then keep their original positions unchanged.
- (7) Repeat (2) to (6) until the largest iteration presupposed or a good enough adaptive value is obtained.
- (8) Export the best particles and corresponding adaptive values as optimal solutions.

4. Simulation Calculation

To verify the model and the formulas, let the intersection be four approaches in main direction of road (including one busway) and two approaches in minor direction of road, as shown in Figure 1. Three-phase signal control plan is adopted at the intersection. Saturated flow rate of each approach equals to $s_0 \times f_b$, where saturated flow rate of straight approach equals to

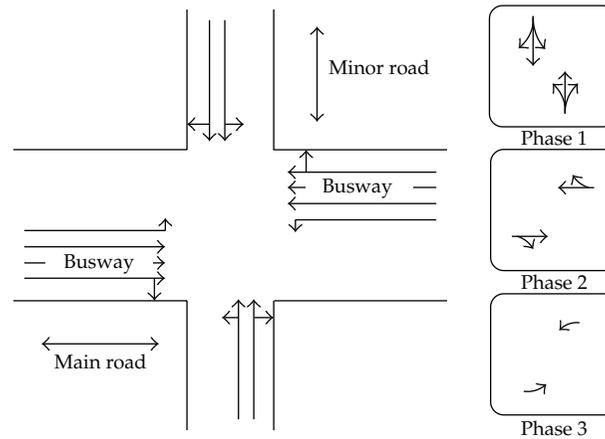


Figure 1: The graph of intersection channelization and phase scheme.

1800 pcu/h, saturated flow rate of left-turn approach equals to 1600 pcu/h, $g_{\min}^1 = g_{\min}^2 = 20$ s, $g_{\min}^3 = 10$ s, and $L = 9$ s. Each of phase clearance reliability is equal, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$. The number of passenger street cars is 2.5, and the number of passenger buses is 40. Simulation results are presented in the following sections.

Let the vehicle arrival rates of all phases be subject to the normal distribution, where arrival rate expectation of street cars is 1100 pcu/h, arrival rate expectation of buses is 240 veh/h, and the buses are running in the main road. The traffic flow ratio of left-turn, straight, and right-turn is 5%, 85%, and 10%, respectively. If street cars and buses arrival rate is constant, that is, the arrival rate of street cars is 1100 pcu/h, while that of the buses is 240 veh/h, then use the solving model to acquire the followings: cycle length = 76 s, $g_1 = 26$ s, $g_2 = 31$ s, and $g_3 = 10$ s. Using the PSO algorithm, the results are shown from Tables 1, 2 and Figure 2 when arrival rate of social vehicles and buses is random variable.

As shown in Tables 1, 2 and Figure 2:

- (1) to ensure the phase clearance reliability without reducing and to meet the random features of the arrival rate, the cycle length and the green time need to be set larger at signalized intersection;
- (2) to get the larger phase clearance reliability, a larger cycle length is needed. And if the vehicle arrival rate is increasing, this trend would be more obvious. As shown in Table 1, if the variance of street cars arrival rate is 220 pcu/h, and the buses arrival rate is constant, the phase clearance reliability increased from 0.75 to 0.95 and the cycle length from 89 s to 126 s, and the phase which has the larger variance of arrival rate needs much green time;
- (3) if all phases have the equal phase clearance reliability, then the phase with higher arrival rate should be allocated with much green time;
- (4) to obtain the higher phase clearance reliability with the larger variance of arrival rate, the cycle length of an intersection would exceed the maximum value. On this circumstance, using fixed-time control mode would not meet the control requirements. Therefore, actuated control or adaptive control mode may be considered.

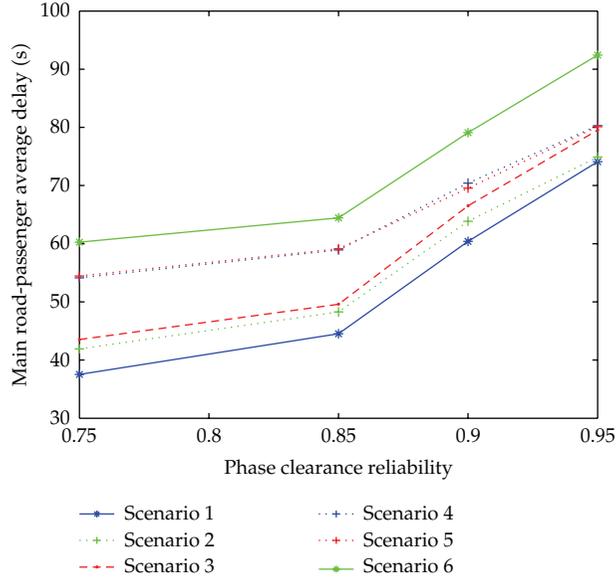


Figure 2: Average delay per passenger of main road.

Table 1: Calculation results when arrival rate of social vehicles is random variable.

α	Scenario 1 variance of street cars = 110 pcu/h number of bus is constant				Scenario 2 variance of street cars = 220 pcu/h number of bus is constant				Scenario 3 variance of street cars = 330 pcu/h number of bus is constant			
	C (s)	g_1 (s)	g_2 (s)	g_3 (s)	C (s)	g_1 (s)	g_2 (s)	g_3 (s)	C (s)	g_1 (s)	g_2 (s)	g_3 (s)
0.75	82	28	32	13	89	31	33	16	95	34	36	16
0.85	92	34	36	13	101	37	38	17	116	44	45	18
0.90	99	37	39	14	112	43	43	17	126	51	48	18
0.95	108	42	43	14	126	48	50	19	135	53	53	20

5. Conclusion

This paper analyzes the limitations of existing transit signal priority control method and considers the stochastic characteristics of arrival rate of street cars and transits; aimed at signal plans for transit signal priority at isolated intersections under the condition of TSP, it establishes a signal timing model with the target of optimizing passenger average delay, designs a corresponding PSO algorithm, and carries out stimulation.

The signal plans for transit signal timing priority model at isolated intersections can reflect the control strategy of transit priority, with the indicator of passenger average delay being a key factor. The optimizing method of phase sequence is an important research content of prior transit signal control method. This research assumes that transit stream travels on the main roads, and there is no turning transit which does not conform to the actual conditions; therefore, the issue of signal plans for transit signal priority at joint phase should be further considered.

Table 2: Calculation results when arrival rate of social vehicles and buses are random variables.

α	Scenario 4				Scenario 5				Scenario 6			
	variance of street cars = 110 pcu/h variance of buses = 24 veh/h				variance of street cars = 220 pcu/h variance of buses = 48 veh/h				variance of street cars = 330 pcu/h variance of buses = 72 veh/h			
	C (s)	g_1 (s)	g_2 (s)	g_3 (s)	C (s)	g_1 (s)	g_2 (s)	g_3 (s)	C (s)	g_1 (s)	g_2 (s)	g_3 (s)
0.75	85	29	33	14	98	35	38	16	103	38	40	16
0.85	99	37	39	14	109	39	42	19	123	48	45	21
0.90	107	41	43	14	119	45	46	19	135	49	54	23
0.95	118	45	49	15	132	49	54	20	144	54	57	24

Acknowledgments

The work described in the paper was supported by Natural Science Foundation of China (no. 50968009) and the Research Fund for the Doctoral Program of Higher Education under Grant (no. 20096204110003).

References

- [1] H. Li and P. D. Prevedouros, "Traffic adaptive control for oversaturated isolated intersections: model development and simulation testing," *Journal of Transportation Engineering*, vol. 130, no. 5, pp. 594–601, 2004.
- [2] F. Dion and B. Hellenga, "A rule-based real-time traffic responsive signal control system with transit priority: application to an isolated intersection," *Transportation Research Part B*, vol. 36, no. 4, pp. 325–343, 2002.
- [3] H. F. Xu, K. P. Li, and M. M. Zheng, "Isolated transit signal priority control strategy based on logic rule," *China Journal of Highway and Transport*, vol. 21, no. 5, pp. 96–102, 2008 (Chinese).
- [4] S. Yagar, "Efficient transit priority at intersections," *Transportation Research Record*, vol. 1390, pp. 10–15, 1993.
- [5] S. Yagar and B. Han, "A procedure for real-time signal control that considers transit interference and priority," *Transportation Research Part B*, vol. 28, no. 4, pp. 315–331, 1994.
- [6] P. Mirchandani, A. Knyazyan, and L. Head, "An approach towards the integration of bus priority, traffic adaptive signal control, and bus information/scheduling systems," *Computer-Aided Scheduling of Public Transport*, vol. 505, no. 4, pp. 319–334, 2001.
- [7] X. G. Yang, Y. Lin, and M. S. Hang, "Study of solution for transit priority signal," *China Journal of Highway and Transport*, vol. 14, pp. 101–105, 2001 (Chinese).
- [8] S. R. Sunkari, P. S. Beasley, T. Urbanik, and D. B. Fambro, "Model to evaluate the impacts of bus priority on signalized intersections," *Transportation Research Record*, no. 1494, pp. 117–123, 1995.
- [9] W. Ma, X. Yang, and M. Yun, "Transit signal priority strategies based on the consideration of bus frequency," *Journal of Tongji University*, vol. 35, no. 11, pp. 1470–1475, 2007 (Chinese).
- [10] Y. M. Bie, H. Zhu, D. H. Wang, D. F. Ma, and X. M. Song, "Development and evaluation of a passive bus signal priority algorithm at isolated intersection," *Journal of Beijing University of Technology*, vol. 37, no. 4, pp. 522–528, 2011 (Chinese).
- [11] Z. R. Abdy and B. R. Hellenga, "Analytical method for estimating the impact of transit signal priority on vehicle delay," *Journal of Transportation Engineering*, vol. 137, no. 8, pp. 589–600, 2011.
- [12] A. Polus, "Modeling and measurements of bus service reliability," *Transportation Research*, vol. 12, no. 4, pp. 253–256, 1978.
- [13] J. Lin and M. Ruan, "Probability-based bus headway regularity measure," *IET Intelligent Transport Systems*, vol. 3, no. 4, pp. 400–408, 2009.

- [14] F. Dion, H. Rakha, and Y. Zhang, "Evaluation of potential transit signal priority benefits along a fixed-time signalized arterial," *Journal of Transportation Engineering*, vol. 130, no. 3, pp. 294–303, 2004.
- [15] H. M. Niu, "Determination of the skip-stop scheduling for a congested transit line by bilevel genetic algorithm," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1158–1167, 2011.
- [16] B. Lv and H. M. Niu, "Optimization for signal plans at isolated intersections on random condition," *Journal of Traffic and Transportation Engineering*, vol. 10, no. 6, pp. 116–120, 2010 (Chinese).
- [17] J. Kennedy, "Particle swarm: social adaptation of knowledge," in *Proceedings of the IEEE International Conference on Evolutionary Computation (ICEC '97)*, pp. 303–308, April 1997.

Research Article

A Simulation-Based Dynamic Stochastic Route Choice Model for Evacuation

Xing Zhao, Gang Ren, Chao Fan, and Chen-Zi Ding

School of Transportation, Southeast University, No. 2, Sipailou, Nanjing 210096, China

Correspondence should be addressed to Xing Zhao, bright-0701@163.com

Received 7 May 2012; Accepted 25 July 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Xing Zhao 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 establishes a dynamic stochastic route choice model for evacuation to simulate the propagation process of traffic flow and estimate the stochastic route choice under evacuation situations. The model contains a lane-group-based cell transmission model (CTM) which sets different traffic capacities for links with different turning movements to flow out in an evacuation situation, an actual impedance model which is to obtain the impedance of each route in time units at each time interval and a stochastic route choice model according to the probit-based stochastic user equilibrium. In this model, vehicles loading at each origin at each time interval are assumed to choose an evacuation route under determinate road network, signal design, and OD demand. As a case study, the proposed model is validated on the network nearby Nanjing Olympic Center after the opening ceremony of the 10th National Games of the People's Republic of China. The traffic volumes and clearing time at five exit points of the evacuation zone are calculated by the model to compare with survey data. The results show that this model can appropriately simulate the dynamic route choice and evolution process of the traffic flow on the network in an evacuation situation.

1. Introduction

Evacuation is one of the most important measures adopted in emergency response to protect masses and to avoid both physical and property damages. Reflecting dynamic propagation characteristics of evacuation traffic flow appropriately is the core theory problem in estimating the evacuation time and evaluating evacuation plans reasonably.

In the studies throughout the world that focus on the evacuation route choice problem, modeling methods include static traffic assignment and dynamic network traffic flow theory. The dynamic models can reflect the propagation process of evacuation traffic flow more effectively than the static method. The core problem of the dynamic method is describing the dynamic propagation and stochastic characteristics of evacuation traffic flow.

The urban dynamic network traffic flow theory under normal conditions has been developed for nearly 20 years with the evolution of intelligent transportation systems, and the macrosimulation model in static traffic flow has been imported into the dynamic flow theory. The cell transmission model (CTM) is one of the dynamic traffic performance models. CTM was first proposed by Daganzo to simulate the traffic flow on highways [1] and was then expanded to network traffic [2], which compromised the accuracy in simulation and optimized the mathematical resolution reasonably. This model can describe dynamic traffic propagation characteristics and capture the phenomena of shockwaves, queue formation, and queue dissipation effectively. For simulating the dynamic propagation process of network traffic flow more accurately, based on the CTM, Lo and Szeto [3] established an optimal dynamic route choice model via the variational inequality. Two years later, Szeto and Lo [4] contributed a DTA variational inequality optimum model considering both choices of route and departing time under the condition of elastic traffic demand. Then, they proposed a disequilibrium DTA model and obtained the route impedance function through CTM simulation in 2005 [5].

Based on previous studies on the DTA problem, the dynamic propagation characteristics of traffic flow in an evacuation situation were researched further. Tuydes and Ziliaskopoulos [6] simulated the escape behaviors in fire via a modified CTM model and linear programming. Dixit and Radwan [7] studied the evacuation problem before typhoon approaches and discussed relevant questions such as evacuation scheduling, route planning, and evacuation destination arrangement based on CTM.

However, during evacuation, road users are not bound to choosing their routes according to the optimum system. To address this problem, some existing models assumed that drivers' route choices were based on current road and traffic conditions. The NETVAC1 model allowed drivers to choose a turning movement at each intersection at each time interval based on the anterior traffic conditions, while the CEMPS model followed the shortest path-based mechanism [8] to make decisions. These models were essentially myopia evacuation route choice models. In fact, road users on the network may consider about the whole evacuation route but cannot obtain all of the traffic information exactly, so stochastic route selection is unavoidable. Therefore, the perceived impedance of a route is an estimate of the actual impedance in a practical application. There is a stochastic variable between the actual impedance and the perceptive impedance, which leads to the problem of stochastic user equilibrium (SUE).

Different assumptions of the estimated term generate different SUE models under normal conditions, in which the multinomial logit (MNL) model and Multinomial Probit (MNP) model are the most widely applied. With further study of the SUE problem, the dynamic stochastic user optimal problem (DSUO), which is an extension of the static stochastic user equilibrium problem, was proposed by Daganzo and Sheffi [9]. As implied by Szeto and Lo [10], it is important to develop and adopt the route choice principle in the DTA model that is consistent with the actual travel behavior. Subsequently, scholars from various countries have paid more attention to the dynamic stochastic user equilibrium theory. Vythoulkas [11], Cascetta and Cantarella [12], Cascetta et al. [13], Lim and Heydecker [14], Sun et al. [15], and Han [16] carried out studies on the DSUO traffic assignment problem, most of which assumed the logit-based model for the route choice behavior of travelers. However, the irrelevant alternatives deficiency of the logit-based model in the modeling is known. Therefore, Zhang et al. [17] presented a time-dependent stochastic user equilibrium (TDSUE) traffic assignment model within a probit-based path choice decision framework. Meng and Khoo [18] did a comparison study to investigate the efficiency and accuracy of the Ishikawa

algorithm with the method of successive averages for the probit-based dynamic stochastic user optimal (P-DSUO) traffic assignment problem.

Generally speaking, on one hand, previous models of the dynamic evacuation route choice problem did not consider the characteristic of the evacuation traffic flow. In an evacuation situation, the large density of traffic flow makes it difficult for drivers to exchange lanes, while the minimum headway diminishes compared with normal conditions [19, 20]. On the other hand, Cova and Johnson [21] pointed out that the main delay in evacuation appears with crossing; however, the traditional CTM model could not clearly simulate the transmission process of the traffic flow at an intersection.

Therefore, the objective of this paper is to simulate the evolution process of the traffic flow on the network and the stochastic route choice in an evacuation situation under determinate road network, signal design, and OD demand. The model contains three parts: the lane-group-based cell transmission model (CTM) which sets different traffic capacities for links with different turning movements to flow out, the actual impedance model which is to obtain the impedance of each route in time units at each time interval, and the stochastic route choice model according to the probit-based stochastic user equilibrium. It can be applied to estimate the real-time evacuation traffic condition and provide the basis for evaluating the performance of evacuation plans developed in response to the possibility of an event or a disaster.

2. Model Formulation

The dynamic stochastic route choice model for evacuation contains three parts, the CTM model, the actual impedance model, and the stochastic route choice model, which is applied to simulate a dynamic propagation process, estimate actual impedance in time units of routes, and simulate route choice of vehicles, respectively. The logistic relationships among the three parts and the structure of the synthesis model are shown in Figure 1.

2.1. Lane-Group-Based CTM Model

The traditional CTM model divides one direction of each street on the network into small homogeneous segments, called cells, while the lane-group-based CTM model divides every link into cells to simulate the evolution process of traffic flow. Drivers cannot change lanes easily in evacuation situations given the large traffic density and car-following phenomenon. Hence, it is assumed that each vehicle considers the desired link fully when the vehicle enters into the roadway.

Based on the transmission mechanism of the traditional CTM model, the traffic propagation rule at intersections can be reflected through different constraints of the first and last cells of the link: the first cell may be an ordinary cell or a merging cell [22], thereinto, the traffic capacity of the merging cell decreases; the end cell may be an ordinary cell or a diverging cell and has a fixed signal phase [23]. However, compared with the traditional CTM model, the set of cells in the proposed model is divided into subsets more specifically. On the one hand, the proposed model sets different traffic capacities for links with different turning movements to flow out; therefore, the last cells of the link are classified precisely considering flow-out directions; on the other hand, the source cells and sink cells are also divided into subsets in considering if the links connected with them are single or not. The lane-group-based

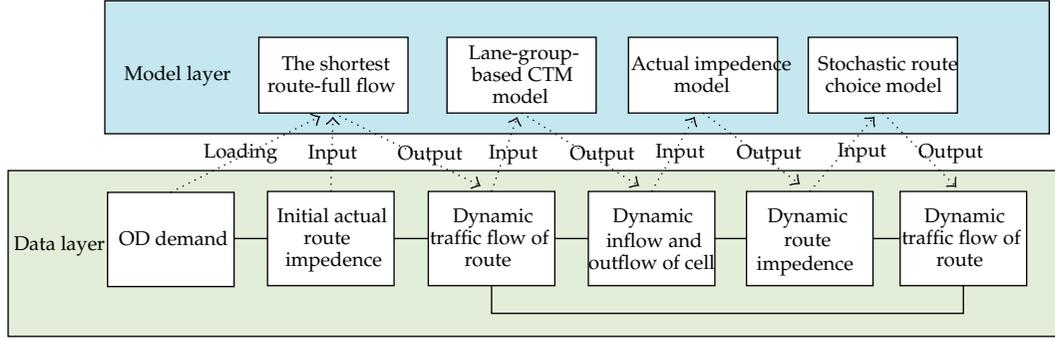


Figure 1: Flow chart of the model.

CTM model in evacuation situation that describes the evolution of traffic flow is expressed as follows.

(1) *Single Link*

The equation of traffic flow propagation can be expressed as

$$\begin{aligned}
 x_{a_n}^t &= x_{a_n}^{t-1} + y_{a_{n-1}a_n}^{t-1} - y_{ab}^{t-1}, \\
 y_{ab}^t &= \min\{x_{a_n}^t, p_a^t N_a Q_{a_n(s/l/r)}^t, N_b Q_{b_1}^t, \delta_{b_1}^t (N_b X_{b_1}^t - x_{b_1}^t)\}, \\
 a &= \Gamma^{-1}(b), \quad a_1, a_2, \dots, a_n \in a, \quad a, b \in L, \quad t \in T.
 \end{aligned} \tag{2.1}$$

In particular, the number of vehicles in the source cell can be acquired by loading values and outflows at each time interval. Thus, when the first cell of a link is connected with the source cell R only:

$$\begin{aligned}
 x_R^t &= x_R^{t-1} + f_R^t - y_{Ra}^{t-1}, \\
 y_{Ra}^t &= \min\{x_{Ra}^t, N_a Q_{a_1}^t, \delta_{a_1}^t (N_a X_{a_1}^t - x_{a_1}^t)\}, \\
 x_R^0 &= f_R^0.
 \end{aligned} \tag{2.2}$$

The sink cell can be considered as a storeroom with infinite capacity. When the last cell of a link is connected with the sink cell S only:

$$\begin{aligned}
 x_{a_n}^t &= x_{a_n}^{t-1} + y_{a_{n-1}a_n}^{t-1} - y_{aS}^{t-1}, \\
 y_{aS}^t &= \min\{x_{a_n}^t, N_a Q_{a_n}^t\}.
 \end{aligned} \tag{2.3}$$

(2) *Merging Link*

The equation of traffic flow propagation of a merging link is

$$\begin{aligned}
 x_{b_1}^t &= x_{b_1}^{t-1} + \sum_{a \in \Gamma^{-1}(b)} y_{ab}^{t-1} - y_{b_1 b_2}^{t-1}, \\
 y_{ab}^t &= \min \left\{ x_{a_n}^t, N_a p_a^t Q_{a_n(s/l/r)}^t, p_a^t N_b Q_{b_c}^t, p_a^t \delta_{b_1}^t (N_b X_{b_1}^t - x_{b_1}^t) \right\}, \\
 a &\in \Gamma^{-1}(b), \quad a_1, a_2, \dots, a_n \in a, \quad a, b \in L, \quad t \in T.
 \end{aligned} \tag{2.4}$$

In particular, for a merging link that not only connected with links but also with a source cell, it is supposed that the vehicles in the links have priority over the source cell.

$$\begin{aligned}
 x_{b_1}^t &= x_{b_1}^{t-1} + \sum_{a \in \Gamma^{-1}(b)} y_{ab}^{t-1} - y_{b_1 b_2}^{t-1}, \\
 y_{ab}^t &= \min \left\{ x_{a_n}^t, N_a Q_{a_n(s/l/r)}^t, N_b Q_{b_c}^t, \delta_{b_1}^t (N_b X_{b_1}^t - x_{b_1}^t) \right\}, \\
 y_{Rb}^t &= \min \left\{ x_R^t, \min \left[N_b Q_{b_c}^t, \delta_{b_1}^t (N_b X_{b_1}^t - x_{b_1}^t) \right] - y_{ab}^t \right\}, \\
 a &\in \Gamma^{-1}(b), \quad R \in \Gamma^{-1}(b), \quad t \in T.
 \end{aligned} \tag{2.5}$$

(3) *Diverging Link*

Each vehicle chooses a route between the OD pair when loaded at the origin and then propagate to the diverging link of the route after some time. Therefore, the proportion of vehicles moving from the diverging link to each downstream link in time interval t can be reckoned by the route choice results of vehicles existing in the end cell of this diverging link currently. The equation of traffic flow propagation of a diverging link is

$$\begin{aligned}
 x_{a_n}^t &= x_{a_n}^{t-1} + y_{a_{n-1} a_n}^{t-1} - \sum_{b \in \Gamma(a)} y_{ab}^{t-1}, \\
 y_{ab}^t &\leq \min \left\{ pr_b^t x_{a_n}^t, \delta_{b_1}^t (N_b X_{b_1}^t - x_{b_1}^t), N_b Q_{b_1}^t, pr_b^t N_a p_a^t Q_{a_n(s,l,r)}^t \right\}, \\
 pr_{a-b}^{t+n} &= \frac{\sum_{rs} \sum_{\forall K, b \in K} f_K^{rs}(t_i)}{\sum_{rs} \sum_{\forall K, b \in K} f_K^{rs}(t_i) + \sum_{rs} \sum_{\forall K, c \in K} f_K^{rs}(t_j) + \dots}, \\
 b, c, \dots &\in \Gamma(a), \quad a_1, a_2, \dots, a_n \in a, \quad a \in L, \quad t \in T.
 \end{aligned} \tag{2.6}$$

In particular, it is assumed that when the source cell is diverged, vehicles in the source cell follow a uniform distribution to flow into the downstream links of the source cell. The equation can be expressed as

$$\begin{aligned} x_{Ra}^t &= x_{Ra}^{t-1} + f_a^t - \sum_{a \in \Gamma(R)} y_{Ra}^{t-1}, \\ y_{Ra}^t &= \min\{pr_a^t x_{Ra}^t, N_a Q_{a_1}^t, \delta_{a_1}^t (N_a X_{a_1}^t - x_{a_1}^t)\}, \\ \sum_{a \in \Gamma(R)} x_{Ra}^0 &= f_R^0. \end{aligned} \quad (2.7)$$

(4) Ordinary Cells within a Link

The equation of traffic flow propagation of ordinary cells within the link is almost the same as the traditional CTM model.

$$\begin{aligned} x_{a_i}^t &= x_{a_i}^{t-1} + y_{a_k a_i}^{t-1} - y_{a_i a_j}^{t-1}, \\ y_{a_k a_i}^t &= \min\{x_{a_k}^t, N_a Q_{a_k}^t, N_a Q_{a_i}^t, \delta_{a_i}^t (N_a X_{a_i}^t - x_{a_i}^t)\}, \\ k \in \Gamma^{-1}(i), \quad j \in \Gamma(i), \quad \forall(k, i) \in a, \quad a \in L, \quad t \in T. \end{aligned} \quad (2.8)$$

2.2. Actual Impedance Model

Due to the restrictions of the traffic capacity of the first and last cells of a link, the inflow of a link during time interval $[t, t+1]$ will be divided into N suboutflows of the link during time intervals $[t, t+1], [t+1, t+2], \dots, [t+N-1, t+N]$. Denote the suboutflow during time interval $[t+n, t+n+1]$ as $\mu_{a\bullet}^{t+n}$ and ensure that the value of $\mu_{a\bullet}^{t+n}$ satisfies the FIFO rule [24]:

$$\mu_{a\bullet}^{t+n} = \begin{cases} 0 & o_a(t+n+1) < x_a^t, \\ o_a(t+n+1) - x_a^t & o_a(t+n) < x_a^t \leq o_a(t+n+1) \leq x_a^t + y_{\bullet a}^t, \\ y_{a\bullet}^{t+n} & \text{when } x_a^t \leq o_a(t+n) \leq o_a(t+n+1) \leq x_a^t + y_{\bullet a}^t, \\ x_a^t + y_{\bullet a}^t - o_a(t+n) & x_a^t \leq o_a(t+n) \leq x_a^t + y_{\bullet a}^t < o_a(t+n+1), \\ 0 & x_a^t + y_{\bullet a}^t < o_a(t+n), \end{cases} \quad (2.9)$$

where

$$\sum_{n=0}^N y_{a\bullet}^{t+n} = x_a^t + y_{\bullet a}^t, \quad (2.10a)$$

$$y_{a\bullet}^t = \sum_{b \in \Gamma(a)} y_{ab}^t, \quad a, b \in L, \quad t \in T, \quad (2.10b)$$

$$y_{\bullet a}^t = \sum_{l \in \Gamma^{-1}(a)} y_{la}^t, \quad l, a \in L, t \in T, \quad (2.10c)$$

$$o_a(t+N) = \sum_{n=0}^{N-1} \sum_{b \in \Gamma(a)} y_{ab}^{t+n}, \quad a, b \in L, t \in T. \quad (2.10d)$$

Then, the average impedance in time units of a link can be acquired:

$$\bar{\tau}_a(t) = \frac{\mu_{a\bullet}^{t+1} + \dots + (N-1)\mu_{a\bullet}^{t+N-1}}{\mu_{a\bullet}^t + \mu_{a\bullet}^{t+1} + \dots + \mu_{a\bullet}^{t+N-1}}. \quad (2.11)$$

The routes between OD pairs are composed of links. Therefore, based on the impedance of links, the actual impedance in time units of route K between OD pair rs at time interval t can be calculated as follows:

$$c_K^{rs}(t) = \bar{\tau}_r(t) + \bar{\tau}_a(\bar{\tau}_r(t) + t - 1) + \bar{\tau}_b(t + \bar{\tau}_a(\bar{\tau}_r(t) + t - 1)) + \dots \quad (2.12)$$

$$r \in R, \quad a, b \dots \in L, K, \quad K \in K_{rs}, \quad b \in \Gamma(a).$$

2.3. Stochastic Route Choice Model

(1) Perceived Impedance in Time Units

As mentioned previously, drivers perceive the route impedance differently. To treat this problem, let the perceived route impedance consist of two parts: the actual impedance and an error term. Assume that

$$\hat{c}_K^{rs}(t) = c_K^{rs}(t) + \varepsilon_K^{rs}(t), \quad t \in T. \quad (2.13)$$

Based on the study of Daganzo [2], the principle of probit-based stochastic user equilibrium, which takes the assumption of a normal distribution, is adopted in this research. Therefore, the distribution of the perceived impedance in time units of path K with nonoverlapping links can be obtained, that is:

$$\text{VAR}(\hat{c}_K^{rs}(t)) = \beta \tilde{t}_K^0, \quad \hat{c}_K(t) \sim N(c_K(t), \beta \tilde{t}_K^0), \quad (2.14)$$

where \tilde{t}_K^0 is the free-flow impedance in time units of route K ; β could be interpreted as the variance of the perceived impedance over a route K at time interval t .

The covariance between the perceived impedance of two routes with overlapping links is expressed as follows:

$$\text{cov}(\hat{c}_{K_i}^{rs}(t, f), \hat{c}_{K_j}^{rs}(t, f)) = \text{cov}(\varepsilon_{K_i}^{rs}(t), \varepsilon_{K_j}^{rs}(t)) = \beta \times \sum_{a \in A} \tilde{t}_a^0 \delta_{a, K_i}^{rs} \delta_{a, K_j}^{rs}, \quad (2.15)$$

$$K \in K_{rs}, \quad r \in R, \quad s \in S, \quad t \in T,$$

where $\delta_{a, K}^{rs}$ is a 0-1 parameter, which takes a value of 1 if link a is on the route K of OD pair rs ; otherwise, it takes a value of 0. \tilde{t}_a^0 is the free-flow impedance in time units of link a [9].

In summary, the perceived route impedance in time units of path K between OD pair rs at time interval t follows a multivariate normal distribution, that is:

$$\vec{c}_K^{rs}(t) \sim \text{MVN}(\vec{c}_K^{rs}(t), \Sigma^{rs}), \quad (2.16)$$

where the diagonal terms of Σ^{rs} are the variances given in (2.14) and the off-diagonal terms are the covariance described in (2.15).

(2) Probability $P_{K_i}^{rs}(t, f)$ and Traffic Volume $f_K^{rs}(t)$

Each evacuated driver estimates impedance of all routes between the OD pair at the time interval loading at the origin and chooses the route for which the impedance is perceived to be the least of all the optional routes. Hence, the probability $P_{K_i}^{rs}(t, f)$ that a driver selects K_i between K_i and K_j at time interval t can be expressed as

$$P_{K_i}^{rs}(t, f) = \Pr\left[\hat{c}_{K_i}^{rs}(t, f) \leq \hat{c}_{K_j}^{rs}(t, f), \forall K_j \in K_{rs}, K_i \neq K_j\right]. \quad (2.17)$$

Furthermore, on the basis of normal distribution properties, it can be calculated as follows:

$$\Pr\left[\hat{c}_{K_i}^{rs}(t, f) \leq \hat{c}_{K_j}^{rs}(t, f)\right] = \Phi\left[\frac{-c_{K_i}^{rs}(t, f) + c_{K_j}^{rs}(t, f)}{\sqrt{\beta \tilde{t}_{K_i}^0 + \beta \tilde{t}_{K_j}^0 - 2 \text{cov}(c_{K_i}^{rs}(t, f), c_{K_j}^{rs}(t, f))}}\right]. \quad (2.18)$$

If the routes between OD pair rs are greater than two, the probability $P_{K_i}^{rs}(t, f)$ of drivers choosing route K_i at time interval t is

$$P_{K_i}^{rs}(t, f) = \Pr\left[\hat{c}_{K_i}^{rs}(t, f) \leq \min\{\hat{c}_K^{rs}(t, f)\}\right]. \quad (2.19)$$

A Monte Carlo simulation can be applied to estimate the probability of each route between each OD pair chosen by drivers.

Based on the probability, the traffic volume of route $K \in K_{rs}$ loading at the source cell R during time interval $[t, t + 1]$: $\vec{f}_K^{rs}(t)$ can be acquired:

$$\vec{f}_K^{rs}(t) = \vec{P}_K^{rs}(t, f) q_{rs}(t), \quad \sum_{K \in K_{rs}} \vec{f}_K^{rs}(t) = q_{rs}(t), \quad f_K^{rs}(t) \geq 0, \quad K \in K_{rs}, \quad r \in R, \quad s \in S, \quad t \in T, \quad (2.20)$$

where $q_{rs}(t)$ is the traffic demand between OD pair rs loading at the source cell R during time interval $[t, t + 1]$.

(3) Objective Function

The stochastic route choice problem is equivalent to finding vectors $\vec{f}_K^{*rs}(t)$ satisfying the following equation:

$$\begin{aligned} \vec{f}_K^{*rs}(t) > 0, \quad \tilde{c}_K^{rs}(t, \vec{f}^*) &= \tilde{c}_{\min}^{rs}(t), \\ \vec{f}_K^{*rs}(t) = 0, \quad \tilde{c}_K^{rs}(t, \vec{f}^*) &> \tilde{c}_{\min}^{rs}(t), \end{aligned} \quad (2.21)$$

where

$$\vec{f}_K^{*rs}(t) \geq 0, \quad K \in K_{rs}, \quad r \in R, \quad s \in S, \quad t \in T, \quad (2.22a)$$

$$\sum_{K \in K_{rs}} \vec{f}_K^{*rs}(t) = q_{rs}(t), \quad \vec{f}_K^{*rs}(t) \geq 0, \quad K \in K_{rs}, \quad r \in R, \quad s \in S, \quad t \in T, \quad (2.22b)$$

$$\vec{P}_K^{rs}(t, \vec{f}^*) q_{rs}(t) = \vec{f}_K^{*rs}(t), \quad K \in K_{rs}, \quad r \in R, \quad s \in S, \quad t \in T. \quad (2.22c)$$

2.4. Solution Algorithm

Step 1 (Initialization). Calculate the free-flow impedance in time units of each route to find the shortest one of each OD pair and assign all of the traffic demands of the corresponding origins on them in each time interval. Record the initial traffic volume of route K between OD pair rs loading at the source cell R during time interval $[t, t + 1]$: $\vec{f}_K^{(1)} = (\vec{f}_K^{(1)}(t), K \in K_{rs}, t \in T)$. Set $n = 1$.

Step 2. The number of iterations is n . Update the traffic volume of route K between OD pair rs loading at the source cell R at time interval t : $\vec{f}_K^{(n)}(t) = \vec{f}_K^{(n-1)}(t) + (1/(n-1))(\vec{y}_K^{(n-1)}(t) - \vec{f}_K^{(n-1)}(t))$.

Step 3. Update the actual impedance in time units of route K at time interval t : $c_K^{rs}(t)$ using the proposed lane-group-based CTM model and the actual impedance model.

Step 4. Calculate the probability of route K between OD pair rs at time interval t chosen by drivers $\vec{P}_K^{rs}(t, \vec{f}^{(n)})$ and the auxiliary traffic volume of route K between OD pair rs loading at the source cell R at time interval t : $\vec{y}_K^{(n)}(t) = q_{rs}(t) \vec{P}_K^{rs}(t, \vec{f}^{(n)})$.

Step 5. If convergence is attained, stop, and $\vec{f}_K^{(n)}(t) = \vec{f}_K^{*(n)}(t)$. If not, set $n = n + 1$ and go to Step 2.

Convergence criterion: $\sqrt{\sum_{n=i-2}^i \sum_{rs} \sum_{K \in K_{rs}} \sum_{t=0}^{T-1} [\vec{f}_K^{rs(n)}(t) - \vec{y}_K^{rs(n)}(t)]^2} \leq \sum_{rs} \xi(K_{rs} - 1), \xi \leq 0.2$.

3. Model Verification

Based on the distribution of parking lots and the road network data, we calculated the evaluation traffic volumes and clearing time of each exit point of the evacuation zone after the

opening ceremony of the 10th National Games of China to compare with survey data to verify the model's effectiveness.

3.1. Building of Evacuation Network

The 10th National Games of China, which were held in Nanjing Olympic Sports Center, led to many traffic needs. According to the usage data supplied by Traffic Administration Bureau, streets on the northern side of the Olympic Center were used for inside driveway parking lots, which parked 1100 vehicles; it also provided two inside driveway parking lots on the eastern side, which parked 465, 385 vehicles separately, and there was an underground parking on the southern side that was not only for the audience but also for players and servicers, which had been used in 439 parking spaces.

The managers conducted some traffic management such as contraflow in the evacuation zone nearby the Olympic Sports Center to handle the large-scale demand of short-term traffic evacuation. All roads within the region applied one-way access during evacuation to compose a closed area allowing traffic to flow out only. Thus, the parking lots are set as origins for evacuation, and the exit links of the northern side and eastern side are assumed to connect to a virtual sink cell separately, which means the whole evacuation network has two destinations. The evacuation network with origins and destinations is shown in Figure 2(a). According to survey data, the evacuation traffic demand between originals and destinations is known, and the OD demand table and the dynamic loading conditions of each origin are shown in Figure 2(b). The specific lane-group-based cell structure is shown in Figure 3.

Specific information of each link is listed in Table 1.

Based on the exiting studies [25], we set the values of the previously mentioned parameters as in Table 2.

3.2. Evacuation Route Choice

The solution algorithm presented previously is coded in Microsoft visual C++ and run on a desktop personal computer with CPU of Intel Core(TM)2 2.2GHz and RAM of 2GB. The computing time to converge is about 16.9 minutes. The long computing time can be reduced by diminishing sampling size for Monte Carlo simulation. When the traffic demand gets larger, the computing time will not elongate obviously unless the loading time intervals become more.

Figure 4 depicts the convergent trend of the algorithm for solving the network of Figure 3 with the aforementioned data. According to Figure 4 it can be seen that convergence at iteration 13th with an average absolute error of 1.4 satisfies the stop criterion.

According to the proposed model, the evacuation route choice result and other important calculations are as follows.

(1) Clearing Time of the Evacuation Network

This is defined as the evacuating time from when the opening ceremony finishes to when the last evacuees arrive at a virtual sink cell. This indicator is the most important one to reflect the performance of evacuation and to evaluate the evacuation plans. Figures 2(a) and 3 show that the links lc , nc , pc are connected with exit D_1 , while ea , kc , qc are connected with exit D_2 . In this case, the clearing time of the evacuation network is the maximum value of all route evacua-

Table 1: Basic information of the links.

Link	Amount of cells	Amount of lanes
<i>aa</i>	4	5
<i>ea</i>	6	10
<i>ia</i>	14	4
<i>ma</i>	4	4
<i>ta</i>	3	8
<i>ya</i>	12	4
<i>fc</i>	4	4
<i>jc</i>	4	4
<i>nc</i>	20	8
<i>ba</i>	4	5
<i>fa</i>	14	6
<i>ja</i>	14	3
<i>na</i>	4	8
<i>ua</i>	3	8
<i>cc</i>	12	4
<i>gc</i>	4	4
<i>kc</i>	6	8
<i>pc</i>	20	6
<i>ca</i>	8	5
<i>ga</i>	14	5
<i>ka</i>	14	3
<i>pa</i>	4	4
<i>wa</i>	12	6
<i>dc</i>	12	3
<i>hc</i>	4	8
<i>lc</i>	10	10
<i>qc</i>	12	6
<i>da</i>	8	5
<i>ha</i>	14	4
<i>la</i>	4	4
<i>qa</i>	4	4
<i>xa</i>	12	5
<i>ec</i>	12	3
<i>ic</i>	4	4
<i>mc</i>	11	10

tion times, 2940 s. The model result of the detailed clearing time of each exit link is shown in Table 3.

(2) Evacuation Route Choice Result

The evacuation route choice result at each time interval can deduce the total traffic volume of each route during the whole evacuation period which is shown in Table 4.

Table 2:

Length of time interval	5 s
Jam density	0.2 veh/m
Free-flow speed	54 km/h (i.e., 15 m/s)
Backward propagation speed	6 m/s
Straight-through capacity of a cell	2160 veh/h/lane (i.e., 3 veh/interval/lane)
Left-turn capacity of a cell	2.4 veh/interval/lane
Right-turn capacity of a cell	2.7 veh/interval/lane
Merging capacity of a cell	2.7 veh/interval/lane
Length of a cell	75 m
Carrying capacity of cell	15 veh/lane
β	0.5

Table 3: Clearing time of each exit.

Exit link	Exit	Evacuation time (5 s)
<i>lc</i>	D ₁	543
<i>pc</i>	D ₂	576
<i>kc</i>	D ₂	553
<i>nc</i>	D ₂	548
<i>ea</i>	D ₁	512
<i>qc</i>	D ₁	588

(3) The Dynamic Traffic Volume

We can obtain the number of vehicles in each cell and outflow of each cell at each time interval by solving x_a^t and $y_{a_i a_j}^t$. Figure 5 shows that the traffic volume of point 1 is the sum of the outflows of links *qc* and *kc*, while the traffic volume of point 5 is the outflow of link *pc*. The calculation of the time-sharing traffic volumes of points 1 and 5 during each time interval are shown in Figure 5.

3.3. Results Comparison

We compare the Previous computed results with the field survey data to verify the validity of the dynamic stochastic route choice model.

The survey collected traffic volumes of the exit points (1~5) of the evacuation zone from 22:00–23:30. The distribution of these points is shown in Figures 2(a) and 3. Among them, traffic volumes of points 1 and 5 are recorded at intervals of three minutes. The comparison result between the model calculations and the survey data of traffic volume of five exit points during the whole evacuating process is shown in Table 5, and the comparison results of the time-sharing traffic volumes of points 1 and 5 are shown in Figure 5.

In Table 5, the values of traffic volumes of the proposed model of each point during the whole evacuation process are calculated by the corresponding route's traffic volume in Table 4. The comparison result shows that model results of the total traffic volumes of exit points are similar to the survey data. Among the five points, the exit points of the routes between the OD pairs O₂D₁, O₃D₁, O₂D₂, and O₄D₂ are certain. Between the OD pair O₁D₁, compared to the chosen route *aa-ca-ea*, the impedance of other routes increases by 40% or

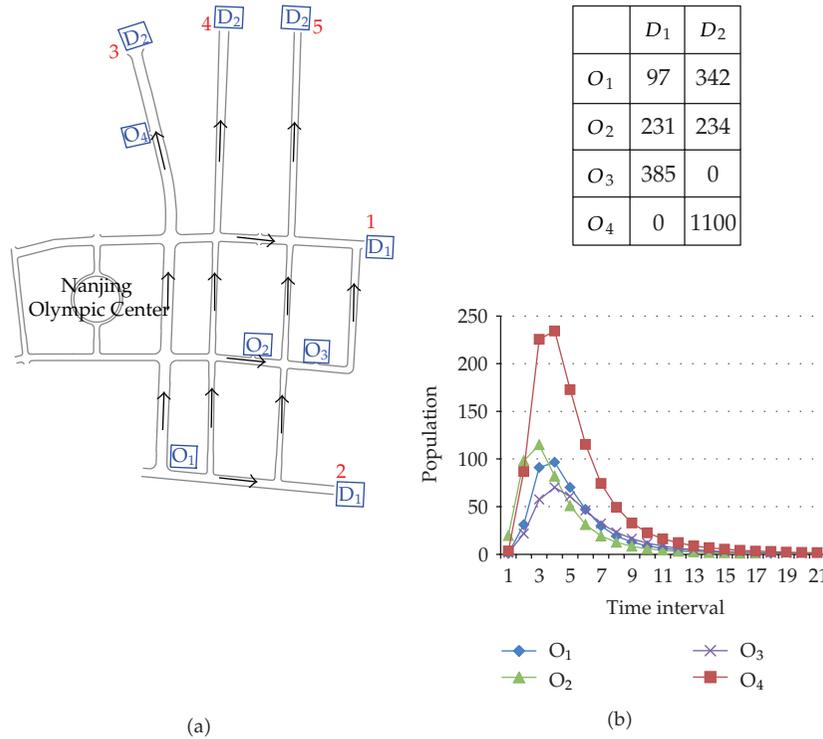


Figure 2: (a) Evacuation network and (b) OD demand table and dynamic loading condition.

Table 4: Traffic volume of each route.

Route	Point	Traffic volume (veh)
<i>pa-ta-ua-qc</i>	1	223
<i>aa-ca-ea</i>	2	97
<i>ga-wa-gc-nc</i>	4	38
<i>qa-ec-pc</i>	5	234
<i>qa-dc-kc</i>	1	8
<i>ga-xa-mc-lc</i>	3	118
<i>ba-ia-cc-nc</i>	4	88
<i>ba-ia-ya-hc-jc-pc</i>	5	1
<i>ua-qc</i>	1	385
<i>lc</i>	3	1100
<i>fa-ma-cc-nc</i>	4	94
<i>fa-ma-ya-hc-jc-pc</i>	5	3

more, so the traffic demand between this OD pair is all evacuated from this route. Therefore, the values of the total traffic volumes of the model at points 1, 2, and 5 are fully consistent with the survey data. However, the values of the total traffic volume at point 3 and point 4 have some errors because the value of traffic volume of the route *ga-xa-mc-lc* tends to be smaller, and the sum of the traffic volumes of routes *ga-wa-gc-nc*, *ba-ia-cc-nc*, *fa-ma-cc-nc* tends to be larger compared to the survey data. Figure 5 shows that the average values of absolute values of time-sharing traffic volume error at points 1 and point 5 are 3.83 veh/3 min and

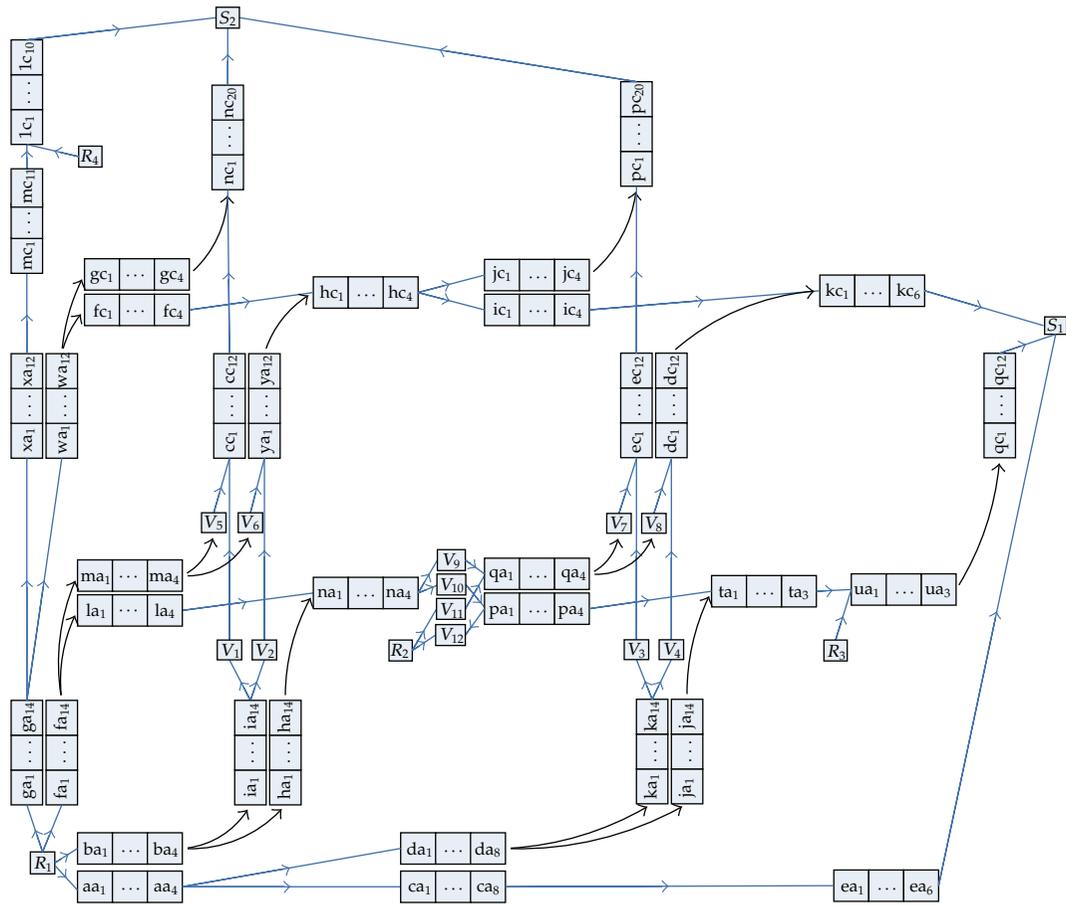


Figure 3: Cell representation of the evacuation network.

Table 5: Comparison of model calculation and survey data of traffic volume.

Point	Destination	Traffic volume	Model calculation of traffic volume	Survey data of traffic volume	Error
1	D ₁	713	616	616	0
2			97	97	
3	D ₂	1676	1218	1180	2.27%
4			220	258	
5			238	238	
Total			2389		2.27%

2.82 veh/3 min, respectively, which fit the distribution of the evacuation process of traffic flow in reality properly.

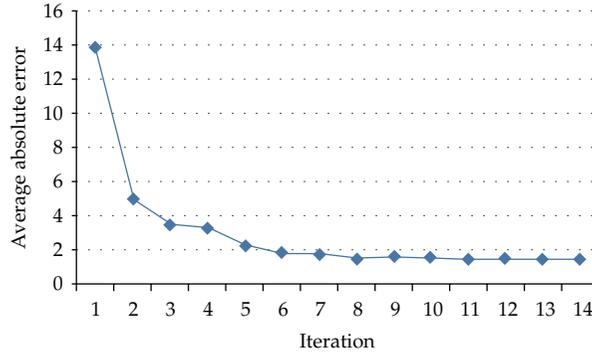


Figure 4: Convergent trend for the evacuation network.

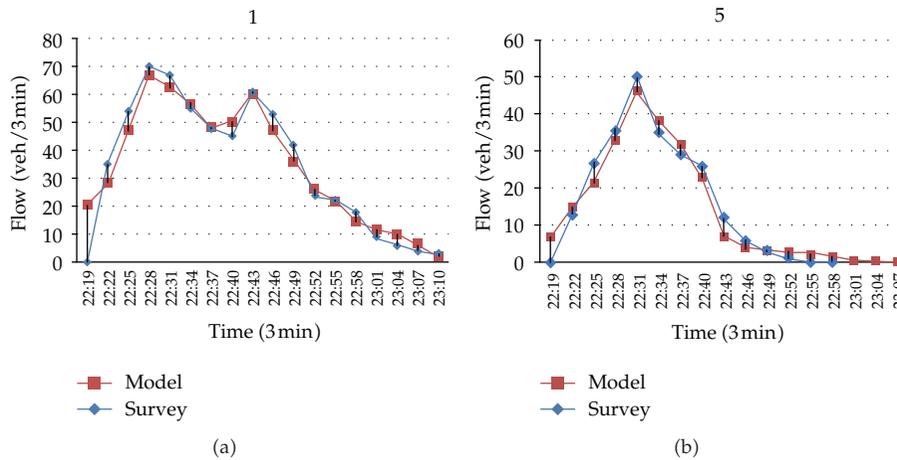


Figure 5: Comparison of model calculation and survey data of time-sharing traffic volume.

4. Conclusions

The dynamic stochastic evacuation route choice model is established to simulate the evolution process of the traffic flow on the network and the stochastic route choice in an evacuation situation under determinate road network, signal design, and OD demand. It contains three parts: the lane-group-based CTM model, the actual impedance model, and the stochastic route choice model.

Considering the large traffic density which makes it difficult for vehicles to exchange lanes in an evacuation situation, this paper established a lane-group-based CTM model, which detailed the propagation process of the traffic flow with different flowing-out turning movements on the basis of the car-following phenomenon in an evacuation situation. This part obtains the inflow and outflow of each cell. Because evacuation in an instant time is of the essence, a realistic model of traffic network performance under a dynamic load is necessary.

Based on the lane-group-based CTM model for evacuation, the piecewise function was established to obtain the actual impedance of each link at each time interval and the dynamic

route impedance; then, combined with the principle of stochastic user equilibrium, we confirmed the error term of the route impedance and acquired the perceived impedance which is taken to be the main criterion for the decision of evacuation route choice.

To verify the effectiveness of this model, this paper applies the proposed model to calculate the evacuation traffic volumes and clearing time of each exit point of the evacuation zone after the opening ceremony of the 10th National Games of China based on the distribution of parking lots and traffic data of the road network. The comparison between the computed results of the proposed model and field survey data proves that this model can reflect the dynamic propagation characteristics of evacuation traffic flow appropriately.

In an emergency evacuation, the OD demand table is not known a priori. Traffic route choice model needs to reflect the emergency circumstances; therefore the estimation of OD demand should be constructed as part of the modeling effort—a subject for further research.

Further studies in the calibration for each parameter in the proposed model under different familiarity of drivers to evacuation network and different levels of emergency evacuation situation are necessary. Design and development of the user interface of this model could simplify the cellular process of the traffic networks and enhance the practicality and operability of the model.

Notations

t :	Set of discrete time intervals
R :	Set of source cell (origin)
S :	Set of sink cell (destination)
L :	Set of links, a link equivalent to a link with a unique turning movement at intersection
$Q_{a_i}^t$:	Maximum number of vehicles that can flow out (traffic capacity) of cell a_i at time interval t
$Q_{a_n(s/l/r)}^t$:	Traffic capacity of the end cell of a through link/left-turn link/right-turn link at time interval t
$Q_{a_c}^t$:	Traffic capacity of merging cell
$\delta_{a_i}^t$:	Ratio of the free-flow speed and backward speed of cell a_i at time interval t
$\Gamma(i)/\Gamma(a)$:	Set of successor cells i or link a
$\Gamma^{-1}(i)/\Gamma^{-1}(a)$:	Set of predecessor cells to cell i or link a
$X_{a_i}^t$:	Maximum number of vehicles in cell a_i at time interval t
p_a^t :	Signal and priority control parameter of link a
f_R^t :	Evacuation demand generated from source cell R at time interval t
$f_K^{rs}(t)$:	Evacuation demand of path K between OD pair rs loading at the source cell R at time interval t
N_a :	Number of lanes of link a
$x_{a_i}^t$:	Number of vehicles in cell a_i at time interval t
x_a^t :	Number of vehicles in link a at time interval t
$y_{a_i a_j}^t$:	Number of vehicles moving from cell a_i to cell a_j at time interval t
y_{ab}^t :	Number of vehicles moving from link a to link b at time interval t
$y_{a\bullet}^t$:	Number of vehicles moving out of link a at time interval t
$\mu_{a\bullet}^{t+n}$:	Number of vehicles moving into at time interval t and moving out of link a at time interval $(t+n)$

- $O_a(t)$: Accumulative number of vehicles moving out of link a during time interval $[0, t]$
- pr_{a-b}^t : Proportion of vehicles moving from link a to its downstream link b at time interval t
- $\bar{\tau}_a(t)$: The average impedance in time units of a link a at time interval t
- $c_K^{rs}(t)$: The actual impedance in time units of route K between OD pair rs at time interval t
- $\hat{c}_K^{rs}(t)$: The perceived impedance in time units of route K between OD pair rs at time interval t
- $\varepsilon_K^{rs}(t)$: A random error term of impedance of route K between OD pair rs at time interval t .

Acknowledgments

This research is supported by the National Natural Science Foundation of China (no. 51078086) and (no. 51278101). The authors appreciate the Jiangsu Provincial Key Laboratory of Transportation Planning and Management of Southeast University.

References

- [1] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research B*, vol. 28, no. 4, pp. 269–287, 1994.
- [2] C. F. Daganzo, "The cell transmission model, part II: network traffic," *Transportation Research B*, vol. 29, no. 2, pp. 79–93, 1995.
- [3] H. K. Lo and W. Y. Szeto, "A cell-based variational inequality formulation of the dynamic user optimal assignment problem," *Transportation Research B*, vol. 36, no. 5, pp. 421–443, 2002.
- [4] W. Y. Szeto and H. K. Lo, "A cell-based simultaneous route and departure time choice model with elastic demand," *Transportation Research B*, vol. 38, no. 7, pp. 593–612, 2004.
- [5] W. Y. Szeto and H. K. Lo, "Non-equilibrium dynamic traffic assignment," in *Proceedings of the 16th International Symposium on Transportation and Traffic Theory (ISTTT '05)*, pp. 427–445, Maryland, USA, 2005.
- [6] H. Tuides and A. Ziliaskopoulos, "The network evacuation problem and solution algorithms," in *INFORMS Annual Meeting*, San Francisco, Calif, USA, 2005.
- [7] V. V. Dixit and E. Radwan, "Hurricane evacuation: origin, route, and destination," *Journal of Transportation Safety and Security*, vol. 1, no. 1, pp. 74–84, 2009.
- [8] S. An, J. X. Cui, and J. Wang, "An international review of road emergency transport and evacuation," *Journal of Transportation Systems Engineering and Information Technology*, vol. 8, no. 16, pp. 38–45, 2008.
- [9] C. F. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transportation Science*, vol. 11, no. 3, pp. 253–274, 1977.
- [10] W. Y. Szeto and H. K. Lo, "Dynamic traffic assignment: properties and extensions," *Transportmetrica*, vol. 2, no. 1, pp. 31–52, 2006.
- [11] P. C. Vythoulkas, "A dynamic stochastic assignment model for the analysis of general networks," *Transportation Research B*, vol. 24, no. 6, pp. 453–469, 1990.
- [12] E. Cascetta and G. E. Cantarella, "A day-to-day and within-day dynamic stochastic assignment model," *Transportation Research A*, vol. 25, no. 5, pp. 277–291, 1991.
- [13] E. Cascetta, F. Russo, and A. Vitetta, "Stochastic user equilibrium assignment with explicit route enumeration: comparison of models and algorithms," in *Proceedings of International Federation of Automatic Control, Transportation Systems*, pp. 1078–1084, Chania, Greece, 1997.
- [14] Y. Lim and B. Heydecker, "Dynamic departure time and stochastic user equilibrium assignment," *Transportation Research B*, vol. 39, no. 2, pp. 97–118, 2005.
- [15] D. Sun, R. F. Benekohal, and S. T. Waller, "Bi-level programming formulation and heuristic solution approach for dynamic traffic signal optimization," *Computer-Aided Civil and Infrastructure Engineering*, vol. 21, no. 5, pp. 321–333, 2006.
- [16] S. Han, "Dynamic traffic modelling and dynamic stochastic user equilibrium assignment for general road networks," *Transportation Research B*, vol. 37, no. 3, pp. 225–249, 2003.

- [17] K. Zhang, H. S. Mahmassani, and C. C. Lu, "Probit-based time-dependent stochastic user equilibrium traffic assignment model," *Transportation Research Record*, no. 2085, pp. 86–94, 2008.
- [18] Q. Meng and H. L. Khoo, "A computational model for the probit-based dynamic stochastic user optimal traffic assignment problem," *Journal of Advanced Transportation*, vol. 46, no. 1, pp. 80–94, 2012.
- [19] H. Hu, X. M. Liu, and X. K. Yang, "Car-following model of emergency evacuation based on minimum safety distance," *Journal of Beijing University of Technology*, vol. 33, no. 10, pp. 1070–1074, 2007.
- [20] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research C*, vol. 19, pp. 1202–1214, 2011.
- [21] T. J. Cova and J. P. Johnson, "A network flow model for lane-based evacuation routing," *Transportation Research A*, vol. 37, no. 7, pp. 579–604, 2003.
- [22] H. K. Lo, "A dynamic traffic assignment formulation that encapsulates the cell transmission model," in *Proceedings of the 14th International Symposium on Transportation and Traffic Theory*, pp. 327–350, Jerusalem, Israel, 1999.
- [23] C. Wright and P. Roberg, "The conceptual structure of traffic jams," *Transport Policy*, vol. 5, no. 1, pp. 23–35, 1998.
- [24] A. P. Lian, Z. Y. Gao, and J. C. Long, "Dynamic user optimal assignment problem of link variables based on the cell transmission model," *Acta Automatica Sinica*, vol. 33, no. 8, pp. 852–859, 2007.
- [25] A. Z. Li, *Research on Lane Distribution Conditions and Capacity of Urban Road Intersections [Ph.D. thesis]*, Southeast University, Nanjing, China, 2008.

Research Article

Study on the Stochastic Chance-Constrained Fuzzy Programming Model and Algorithm for Wagon Flow Scheduling in Railway Bureau

Bin Liu

Traffic and Transportation School, Lanzhou Jiaotong University, Lanzhou 730070, China

Correspondence should be addressed to Bin Liu, liubin0909@mail.lzjtu.cn

Received 15 May 2012; Revised 13 August 2012; Accepted 15 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Bin Liu. 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 wagon flow scheduling plays a very important role in transportation activities in railway bureau. However, it is difficult to implement in the actual decision-making process of wagon flow scheduling that compiled under certain environment, because of the interferences of uncertain information, such as train arrival time, train classify time, train assemble time, and flexible train-size limitation. Based on existing research results, considering the stochasticity of all kinds of train operation time and fuzziness of train-size limitation of the departure train, aimed at maximizing the satisfaction of departure train-size limitation and minimizing the wagon residence time at railway station, a stochastic chance-constrained fuzzy multiobjective model for flexible wagon flow scheduling problem is established in this paper. Moreover, a hybrid intelligent algorithm based on ant colony optimization (ACO) and genetic algorithm (GA) is also provided to solve this model. Finally, the rationality and effectiveness of the model and algorithm are verified through a numerical example, and the results prove that the accuracy of the train work plan could be improved by the model and algorithm; consequently, it has a good robustness and operability.

1. Introduction

The train work plan is the core of the daily work plan and the data hub of types of scheduling work in railway bureau. It plays an important role as the whole link between Railway Ministry of China and railway stations and depots. The main purpose of train work plan is to allocate wagons to departure trains. And the wagon flow has to match the time limitation of wagon operations on marshalling stations and the wagon flow direction and train-size limitation. Besides, the train work plan has to meet the demand of wagon loading plan and

empty wagon reposition plan. All these train operations could be classified as *wagon flow scheduling problem*. The optimization objective of wagon flow scheduling is to accelerate the rolling stock turnover, to reduce the wagon residence time at railway stations, and to reduce the empty wagon running distance.

For railway wagon flow scheduling problem, many scholars have done a lot of beneficial researches, especially for wagon flow scheduling in railway marshalling stations. Gulbroden studied the railway scheduling in marshalling station by using operations research [1]. Yager and his partner developed an efficient sequencing model for humping in railway marshalling stations [2]. Carey and Carville developed scheduling heuristics analogous to those successfully adopted by train planners using “manual” methods [3]. Lentink et al. discussed how to use network flow method to establish mathematical model to solve train scheduling problem [4]. The robust optimization in railway transportation is discussed by Marton et al. in 2007 [5].

In China, many railway transportation organization methods are fundamentally different from other countries. Wang presented a concept of “*price*” and used some techniques to transform the wagon flow allocating problem into a transportation problem model in operations research; the objective of the model is to minimize the total price so that the satisfactory solution can be attained by using the calculating method on table [6]. It is worthy of mentioning that the literature [7] is one of the most important literatures in this research field, and many successive studies derive from this. He et al. developed a fuzzy dispatching model for wagon flow scheduling in railway marshalling station and designed a genetic algorithm to obtain the satisfactory solution [8]. And He et al. developed an integrated dispatching model for railway station operations and a computer-aided decision support system [9]. Liu et al. developed a chance-constrained programming model which aimed to reduce the residence time of wagons in the marshalling station and the average delay time of departure trains and designed an improved genetic algorithm to solve the problem [10]. Li et al. addressed the problem of optimizing the marshalling station stage plan with the random train arrival time and developed a dependent-chance programming model and designed a hybrid intelligent algorithm based on stochastic simulation and tabu search [11]. And Li et al. put forward a brief survey of stage plans under certain and uncertain environments and with computer-aided dispatching methods and systems. He pointed out the existing and unresolved problems in application of the current theories and methods. What is more, he investigated the direction of future research of railway marshalling stations stage plan [12].

In recent years, with the wide use of Train Dispatch and Management System (TDMS) and Synthetically Automatic Marshalling (SAM) station system in China, the wagon flow information between railway bureau and railway stations are shared completely, so we put forward a new transportation organization concept that is “*integralization of railway bureau and railway stations.*” Thus the wagon flow scheduling in railway bureau can replace the railway station wagon flow scheduling in great extent, and the accuracy of the wagon flow scheduling plan and the overall transportation organization efficiency can be improved dramatically.

Based on above literatures and practical situation in China, considering all kinds of uncertain factors in wagon flow scheduling, such as the stochastic train arrival time and fuzzy train-size limitation of the departure trains, a stochastic chance-constrained fuzzy multiobjective model for the flexible wagon flow scheduling problem is set up. And a hybrid intelligent algorithm based on ant colony optimization (ACO) and genetic algorithm (GA) is also given in this paper.

2. Stochastic Chance-Constrained Fuzzy Programming Model

In this section, we aim at maximizing the satisfaction of departure train-size limitation and minimizing the wagon residence time at railway station to establish the multiobjective optimization model for wagon flow scheduling problem in railway bureau. First, we analyze and formulate the constraints of wagon flow scheduling by considering the some uncertain factor. Then we will summarize and formulate the objective function.

2.1. Start Time Constraints to Classify a Train

The train arrival time is stochastic. The lag between train actual and planed arrival time is a random variable with normal distribution, denoted by ε_i . Let T_i'' be the planed arrival time of train i , so the actual arrival time of train i is $T_i' = T_i'' + \varepsilon_i$, T_i the earliest time after the inspection of train i , $T_i = T_i' + a_i$, where a_i is the inspection time of train i . Assume that a train classifying process is in a time segment k . Let t_k be the start time to classify a train in the k th time segment, J_k^i a boolean variable whose value is 1 if train i is classified in the k th time segment, otherwise the value is 0; let n be the total number of arriving trains in the stage [8]. So the start time constraints to classify a train are as follows:

$$\Pr\left(t_k - \sum_{i=1}^n T_i J_k^i \geq 0\right) \geq \alpha_1, \quad k = 1, 2, \dots, n. \quad (2.1)$$

Expression (2.1) denotes that the start time to classify train i in the k th time segment must be after the end time of inspection train i . Because of the stochasticity of train arrival time, the end time of train inspection is also stochastic. The expression (2.1) is chance constrained, and the probability of expression (2.1)'s holding is more than or equal to α_1 , where α_1 is the given confidence level.

2.2. Start Time Constraints to Assemble a Train

Assume that a train assemble process is in a time segment k' . Let b_j be the process time of assemble train j ; it is a random variable with normal distribution. Let n' be the total number of departure trains in the stage.

Let t'_k be the start time to assemble a departure train in the k th time segment, $P_{ik}^{jk'}$ be a boolean variable whose value is 1 if train i which is classified in the k th time segment delivers wagons to departure train j which is assembled in the k' th time segment, otherwise the value is 0 [7]. So the start time constraints to assemble a departure train are as follows:

$$\Pr\left(t_k + \sum_{i=1}^n j_i J_k^i - t'_{k'} \leq M \left(1 - \sum_{i=1}^n \sum_{j=1}^{n'} P_{ik}^{jk'}\right)\right) \geq \alpha_2, \quad k = 1, 2, \dots, n; \quad k' = 1, 2, \dots, n', \quad (2.2)$$

where j_i is the random classifying time of train i with normal distribution. And M is an extremely big positive number. Expression (2.2) denotes that the start time to assemble train j in the k' th time segment must be after the end time to classify train i if inbound train i

delivers wagons to departure train j . And the probability of expression (2.2)'s holding is more than or equal to α_2 , where α_2 is the given confidence level.

2.3. Wagon Flow Delivers Relationship Constraints

Let m_{ir} be the arrival wagon number in classified train i whose destination is direction r . And v_j is the train-size limitation of departure train j , q is total number of destination direction in arrival train, x_{ir}^j is decision variable which means the departure train j wagon number whose destination direction is r and from arrival train i , and $\Omega(j)$ is the total destination direction number of departure train j . Q_j is a boolean variable, its value is 1 if the j th scheduled departure line is occupied by the train, otherwise the value is 0 [7]. So wagon flow delivers relationship constraints that are as follows:

$$\sum_{j=1}^{n'} x_{ir}^j = m_{ir}, \quad i = 1, 2, \dots, n; \quad r = 1, 2, \dots, q, \quad (2.3)$$

$$\sum_{i=1}^n \sum_{r=1}^q x_{ir}^j = v_j \cdot Q_j, \quad j = 1, 2, \dots, n'. \quad (2.4)$$

Expression (2.3) denotes that the wagon to direction r from arrival train i can be delivered to different departure train with the same direction. Expression (2.4) denotes the departure train-size limitation which will be further discussed in Section 2.6.

2.4. Train Departure Time Constraints

Let d'_j be the scheduled departure time of train j in train timetable, h_j the inspection time of departure train j , z the convoy time needed of the departure train from classification yard to departure yard, d_j the latest time of the assembling of train j that should be completed, $d_j = d'_j - h_j - z$, $B_{k'}^j$ a boolean variable whose value is 1 if train j is assembled in the k th time segment, otherwise the value is 0, and d_j^* the lag between train j scheduled departure time and actual departure time [9]. So train departure time constraints are as follows:

$$\Pr\left(t_{k'}^j + b_j B_{k'}^j - d_j \leq M(1 - B_{k'}^j)\right) \geq \alpha_3, \quad k' = 1, 2, \dots, n'; \quad j = 1, 2, \dots, n'. \quad (2.5)$$

Expression (2.5) denotes that the completed time to assemble train j must not exceed the latest time that is determined by the train timetable so that train j will depart on time. As the classifying time of a train is stochastic, the probability of expression (2.5)'s holding is more than or equal to α_3 , where α_3 is the given confidence level. And M is an extremely big positive number.

2.5. Logic Constraints

In order to guarantee the logical relationships among variables in the model, the following logic constraints have to conform [8]:

$$\begin{aligned} \sum_{i=1}^n J_k^i &= 1, \quad k = 1, 2, \dots, n, \\ \sum_{k=1}^n J_k^i &= 1, \quad i = 1, 2, \dots, n. \end{aligned} \quad (2.6)$$

Expression (2.6) denotes only one train can be classified in each time segment; one train can be classified only once:

$$\begin{aligned} \sum_{j=1}^{n'} B_{k'}^j &= 1, \quad k' = 1, 2, \dots, n', \\ \sum_{k'=1}^{n'} B_{k'}^j &= 1, \quad j = 1, 2, \dots, n'. \end{aligned} \quad (2.7)$$

Expression (2.7) denotes only one train can be assembled in each time segment; one train can be assembled only once:

$$\sum_{j=1}^{n'} \sum_{k'=1}^{n'} P_{ik}^{jk'} \leq M \cdot J_k^j, \quad i = 1, 2, \dots, n; \quad k = 1, 2, \dots, n. \quad (2.8)$$

Expression (2.8) denotes if the arrival train delivers wagons to departure train j ; train i must be classified, where M is an extremely big positive number:

$$\sum_{i=1}^n \sum_{k=1}^n P_{ik}^{jk'} \leq M \cdot B_{k'}^j, \quad j = 1, 2, \dots, n'; \quad k' = 1, 2, \dots, n'. \quad (2.9)$$

Expression (2.9) denotes if the departure train j will be assembled; it must have arrival trains deliver their wagons to the departure train j , where M is an extremely big positive number:

$$\begin{aligned} J_k^i &\in \{0, 1\}, \quad P_{ik}^{jk'} \in \{0, 1\}, \\ B_{k'}^j &\in \{0, 1\}, \quad Q_j \in \{0, 1\}, \\ i &= 1, 2, \dots, n, \quad j = 1, 2, \dots, n', \\ k &= 1, 2, \dots, n, \quad j = 1, 2, \dots, n'. \end{aligned} \quad (2.10)$$

2.6. Objective Function

Let T_e be the end time of the wagon flow scheduling stage, then the objective function is as follows [8]:

$$\max Z_1 = \sum_{j=1}^{n'} \left(\sum_{i=1}^n x_i^j (T_e - d_i' - d_j^*) \right). \quad (2.11)$$

Let $\lambda_j(x_j)$ be the satisfactory function of the actual train size x_j compared to the expected train-size v_j ; it is a trapezoidal form fuzzy number as follows [8]:

$$\lambda_j(x_j) = \begin{cases} 0, & x_j < \underline{v}_j^1; \\ \frac{x_j - \underline{v}_j^1}{\underline{v}_j^1 - \underline{v}_j^1}, & \underline{v}_j^1 \leq x_j < \underline{v}_j^1; \\ 1, & \underline{v}_j^1 \leq x_j \leq \underline{v}_j^2; \\ \frac{\underline{v}_j^2 - x_j}{\underline{v}_j^2 - \underline{v}_j^2}, & \underline{v}_j^2 < x_j \leq \underline{v}_j^2; \\ 0, & x_j > \underline{v}_j^2, \end{cases} \quad (2.12)$$

where \underline{v}_j^1 is the minimum number, \underline{v}_j^2 is the maximum number, and $[\underline{v}_j^1, \underline{v}_j^2]$ is the expected interval of numbers.

Considering the fuzzy train-size limitation of departure train, we formulated the second objective function:

$$\max Z_2 = \sum_{j=1}^{n'} \lambda_j \left(\sum_{i=1}^n x_i^j \right). \quad (2.13)$$

3. Chance Constraint Conversion

For a chance constraint $\Pr\{g(x, \xi) \leq 0\} \geq \alpha$, where ξ is a random variable with distribution function Φ , if function $g(x, \xi)$ has the form $g(x, \xi) = h(x) - \xi$, then $\Pr\{g(x, \xi) \leq 0\} \geq \alpha$ if and only if $h(x) \leq K_\alpha$, where $K_\alpha = \sup\{K \mid K = \Phi^{-1}(1 - \alpha)\}$. So the deterministic equivalent form of $\Pr\{g(x, \xi) \leq 0\} \geq \alpha$ is as follows [13]:

$$h(x) \leq K_\alpha, \quad K_\alpha = \sup\{K \mid K = \Phi^{-1}(1 - \alpha)\}. \quad (3.1)$$

The same goes for $\Pr\{g(x, \xi) \geq 0\} \geq \alpha$; the deterministic equivalent form is as follows [13]:

$$h(x) \geq K_\alpha, \quad K_\alpha = \inf\{K \mid K = \Phi^{-1}(\alpha)\}. \quad (3.2)$$

We can convert all the chance constraints in the model above into deterministic equivalent form by expressions (3.1) and (3.2). Let Φ_J be the probability distribution function of j_i , Φ_D the probability distribution function of ε_i , and Φ_B the probability distribution function of b_j , and then we have the following deterministic equivalent form.

Equivalence formula for expression (2.1) is

$$t_k - \left(T_i'' + a_i + \Phi_D^{-1}(\alpha_1) \right) \cdot \sum_{i=1}^n J_k^i \geq 0, \quad k = 1, 2, \dots, n. \quad (3.3)$$

Equivalence formula for expression (2.2) is

$$M \left(1 - \sum_{i=1}^n \sum_{j=1}^{n'} P_{ik}^{jk'} \right) + t'_k - t_k - \Phi_J^{-1}(\alpha_2) \cdot \sum_{i=1}^n J_k^i \geq 0, \quad k = 1, 2, \dots, n; \quad k' = 1, 2, \dots, n'. \quad (3.4)$$

Equivalence formula for expression (2.5) is

$$M \left(1 - B_{k'}^j \right) - t'_k - \Phi_B^{-1}(\alpha_3) \cdot B_{k'}^j + d_j \geq 0, \quad k' = 1, 2, \dots, n'; \quad j = 1, 2, \dots, n'. \quad (3.5)$$

4. A Hybrid Algorithm Based on ACO and GA

The wagon flow scheduling problem is an NP-complete problem proved by Dahlhaus et al. [14]. In this section, we focus on the hybrid algorithm design based on ant colony optimization (ACO) and genetic algorithm (GA). ACO algorithms are the most successful and widely recognized algorithmic techniques based on ant behaviors, initially proposed by Dorigo in 1992 in his Ph.D. thesis [15]. Genetic algorithms are developed by Holland in 1975. It is a powerful and broadly applicable stochastic search and optimization techniques, inspired by natural evolution, such as inheritance, crossover, mutation, and selection [16]. In this paper, the hybrid algorithm is mainly based on ACO; the crossover and mutation operator of GA is used to avoid the "premature" or "stagnation" of ACO.

Let the arrival train set be DD whose element is dd_1, dd_2, \dots, dd_m ordered by the train arrival time, and let the departure train set be CF whose element is cf_1, cf_2, \dots, cf_n ordered by the train departure time in this stage. These two sets are denoted by $DD = \{dd_1, dd_2, \dots, dd_m\}$ and $CF = \{cf_1, cf_2, \dots, cf_n\}$, respectively. The train makeup destination direction is a set denoted by $\Omega = \{r_1, r_2, \dots, r_q\}$ [6].

From Section 2.1, we know that $T_i' = T_i'' + \varepsilon_i$, $T_i = T_i' + a_i$ is the earliest start time to classify train i , so the actual classifying time cannot be earlier than T_i . Let t_k be actual start time to classify train i in the time segment k . Assume that the departure train sequence is $j_1, j_2, \dots, j_{m'}$ in which arrival train can deliver wagons to them and their departure time is $d'_{j_1}, d'_{j_2}, \dots, d'_{j_{m'}}$. From the train operating process, it is known that if the end time to classify the arrival train is later than $d'_{j_n} - h_j - z - b_j$ (where $1 \leq n \leq m'$), then the arrival train cannot deliver wagons to the departure train.

Define the classifying time window $[E_i, D_i]$ for train i , where E_i is the earliest start time and D_i the latest end time to classify train i . Thus the actual classifying time should be between E_i and D_i . Let ω_i be the penalty factor for the delay to classify train i , and in this paper ω_i is the wagon number of arrival train i .

4.1. Initialization

A classifying sequence of arrival trains can be regarded as an ant's travel path. For example, $(3, 1, 5, \dots, i)$ represents the trains classified by the order $3, 1, 5, \dots, i$, and there are n nodes on the path which represents the arrive trains, respectively. If the i th ant passing node is j , it means putting the arrival train j in the position i to classify. In the process of ants travel, the passed nodes make up the train collection and it is the taboo list *tabuk*, so the every completion of ant's travel makes a new solution.

4.2. Transition Probability

Let $L = \{(i_1, i_2) \mid i_1, i_2 \in DD\}$, and we set up a network $G = (DD, L)$, the purpose about this network is to search path that mostly satisfies the constraints of the departure train, such as train-size limitation, punctuality, and inviolate wagon flow direction. At first the pheromone on each edge is equal. And then every ant must make a choice to move to next node; it means that train will be classified in next step.

Suppose that, at time t , the probability of ants s to transfer from train i_1 to train i_2 is [15, 17]

$$P_{i_1, i_2}^s(t) = \begin{cases} \frac{[\tau_{i_1, i_2}]^\alpha \cdot [\eta_{i_1, i_2}]^\beta}{\sum_{z \notin \text{tabuk}} [\tau_{i_1, z}]^\alpha \cdot [\eta_{i_1, z}]^\beta}, & i_2 \notin \text{tabuk} \\ 0, & i_2 \in \text{tabuk}, \end{cases} \quad (4.1)$$

in which η_{i_1, i_2} is the heuristic information

$$\eta_{i_1, i_2} = \frac{1}{d_i + \xi_1 [\omega_i \max(0, C_i - D_i)] + \xi_2 ((D_i - E_i) / b_j)}. \quad (4.2)$$

The *tabuk* is the *tabu* list that stands for the set of arrival train which has already been classified; C_i is the end time for the actual classifying time of train i ; ξ_1 and ξ_2 are the weight coefficients; b_j is the time of classifying operation; α and β are the parameters used to control the relative importance of pheromone and heuristic information.

4.3. Selection and Local Search Strategy

Let q_0 be a constant, $q \in (0, 1)$ is a random number, if $q \leq q_0$, the next node the ants transfer to is a node that makes $[\tau_{i_1, i_2}]^\alpha \cdot [\eta_{i_1, i_2}]^\beta$ has the maximum value; otherwise, the node will be ensured upon the transition probability by taking the traditional roulette method.

If a local optimal solution is found in the early iteration of ACO, it is easier to appear "premature" or "stagnation" phenomenon, and there is also a need to apply a local search strategy in order to adjust the obviously inappropriate classifying order. So the crossover and mutation strategy of GA is adopted in this paper.

When an ant completes a tour, a train classifying sequence is obtained, then static wagon allocating method is adopted to calculate the "price" of the classifying sequence [6], and then two of the "minimum price" classifying sequences are selected to crossover.

The crossover strategy [18]: for chromosomes P_1 and P_2 , randomly generating two random numbers to determine the crossover position, exchange the classifying order between the crossover locations P_1 and P_2 ; if the gene is repeated between and outside of the crossover position, then delete the gene in this location, and then put the lacked gene to the chromosome by ascending order, and then the new chromosomes P_1'' and P_2'' are obtained. Then we calculate P_1'' and P_2'' classifying price and compare to the corresponding price of P_1 and P_2 . We select the minimum price chromosome P and execute mutation operation, exchange the genes of the two positions which are determined by two random numbers, and calculate the price. Thus the current optimal classifying sequence is represented by the path where the price is the less one of P and P' .

4.4. Pheromone Updating Strategy

The pheromone can be updated as follows:

$$\tau_{i1,i2}(t+1) = (1 - \rho)\tau_{i1,i2}(t) + \Delta\tau_{i1,i2}, \quad (4.3)$$

where ρ is the parameter to control the pheromone evaporation rate between time t and $t+1$; $1 - \rho$ is the retention of the pheromone in the current path. At the beginning, $\tau_{i1,i2} = c$ (c is a constant), and $\Delta\tau_{i1,i2}(t+1)$ is the residues pheromone on the passing edge. If the current path is the optimal one, then $\Delta\tau_{i1,i2}(t+1) = 1/P^*$, where P^* is the total price of the optimal sequence; otherwise, $\Delta\tau_{i1,i2}(t+1) = 0$.

4.5. The Steps of the Algorithm

Step 1. Initialization. According to the train arrival information to calculate train the classifying time window, and initialize wagon allocating price table. Set the same amount of pheromone on each edge.

Step 2. Sort trains by their arrival time, and update the wagon allocating price table and calculate the price.

Step 3. Place each ant to each node in G , and set tabu list with the corresponding node.

Step 4. Take an ant, calculate the transition probability of selecting the next node to update the tabu list, and then calculate the transition probability, select the node, and update the tabu list again until traverse through all the nodes.

Step 5. Calculate the pheromone that the ant left to each edge, then the ant die.

Step 6. Repeat Steps 3 and 4 until all the ants finish their tour.

Step 7. Calculate the prices of each path that ants choose.

Step 8. Choosing two of the smallest price paths (a path represents a chromosome) P_1 and P_2 to compare with path in Step 3, select the less one and make them crossover to obtain new paths P_1'' and P_2'' .

Table 1: Information of arrival trains.

Train code	Arrival time	Train makeup	Train code	Arrival time	Train makeup
dd0	0	A/15, B/22, C/10, D/12	dd6	112	A/24, D/22
dd1	10	A/30, B/20	dd7	136	A/25, C/23
dd2	35	A/15, C/35	dd8	152	B/20, C/10, D/18
dd3	41	A/35, C/15	dd9	172	A/30, C/18
dd4	58	B/15, C/15, D/15	dd10	208	B/25, C/20
dd5	92	B/20, d/25	dd11	225	A/20, C/23

Step 9. Calculate paths price of P_1'' and P_2'' , compare with the prices of P_1 and P_2 , and select one of the smallest P .

Step 10. Execute mutation operation for P , and calculate the path price after the mutation compared with the price of P , and then select the less one as the optimal path so far.

Step 11. Update the current optimal path, and empty the tabu list tabuk.

Step 12. Judge whether the iterations hit the predetermined number, or whether there is stagnation. If it does, we terminate the algorithm and the output current optimal path; otherwise, go to Step 3, execute the next loop of iteration.

5. An Illustrative Example

We take a certain wagon flow scheduling platform as example in one of the railway bureaus in China. Assume that some technological standard operation time is as follows: convoy time is 10 min, arrival inspection time is 35 min, and departure inspection time is 25 min. The lag between train actual and planed arrival time is a normal distribution variable $\mathcal{N}(0, 5)$, train classify time obeys $\mathcal{N}(15, 3)$, and train assemble time obeys $\mathcal{N}(15, 3)$, and departure train-size limitation is a fuzzy trapezoidal variable with parameters of (40, 45, 50, 52).

Since train classify and assemble time conforms $\mathcal{N}(\mu, \sigma^2)$, and the equivalence formula of them can convert to $\inf\{K \mid K = \Phi^{-1}(\alpha)\}$ by expression (3.4), so $\Phi^{-1}(0.95) = 1.6449$ when the confidence level $\alpha = 95\%$.

The arrival train information is shown as in Table 1. In the convenience of calculating, we set the start time stage is 0 and convert the train arrival time is an integer number which stands for the minutes that train arrival from the stage start time [19]. And assume that there are four train destination directions denoted by A, B, C, and D. The train 0 is a dummy train that represents the wagon flow in the beginning of this stage.

Suppose that all the arrival train can be classified immediately. We can calculate the initial wagon allocating price table according to static wagon allocating problem [6]. In this paper, the Java programming language is used to implement the algorithm above with parameters $\alpha = 1$, $\beta = 1$, $\xi_1 = 1$, $\xi_2 = 2$, $q_0 = 0.6$, $\rho = 0.7$. The satisfactory solution of wagon flow scheduling is shown in Table 2.

From Table 2, we know that all of the departure trains meet the train-size limitation from the point of the fuzzy constraint. And in this stage, there are 495 wagons scheduled to the departure trains. Since we consider the stochasticity and fuzziness in the model and algorithm, the robustness and operability of the work plan of railway bureau is prompted greatly.

Table 2: Information of departure trains.

Train code	Departure time	Train makeup: wagon source	Train code	Departure time	Train makeup: wagon source
cf1	112	B/50: dd0/22, dd1/28	cf6	235	A/50: dd3/15, dd6/24, dd7/11
cf2	142	A/50: dd0/15, dd2/15, dd3/20	cf7	266	D/42: dd5/2, dd6/22, dd8/18
cf3	150	C/50: dd0/10, dd1/20, dd2/20	cf8	285	B/20, C/32: dd8/20, dd2/2, dd3/15, dd4/15
cf4	198	D/50: dd0/12, dd4/15, dd5/23	cf9	328	C/51: dd7/23, dd8/10, dd9/18
cf5	216	B/37, C13: dd1/2, dd4/15, dd5/20, dd2/13	cf10	356	A/50: dd7/14, dd9/30, dd11/6

6. Conclusions

In this paper, considering the stochasticity of train arrival time, train classify time, and train assemble time and fuzzy train-size limitation, a stochastic chance-constrained fuzzy multiobjective model for wagon flow scheduling is set up based on the uncertain programming theory. By analyzing the model in detail, a hybrid intelligent algorithm based on ACO and GA is given. Furthermore, a numerical example is also offered to verify the rationality and effectiveness of the model and algorithm. As we know, the China railway informatization is very fast in recent years. But the TDMS, SAM, and other management information system are separate and not intelligent in some extent. So it needs to integrate the related systems by optimizing the transportation business models. The model and algorithm proposed in this paper provide the theoretical basis for integrating and optimizing the related systems. We hope the TDMS and SAM will be more practical and intelligent by using our model and algorithm in this paper.

In the future, we will study how to refine the basic wagon flow information, as well as the robust theory for wagon flow scheduling, and how to use synergetic theory in wagon flow and locomotive scheduling.

References

- [1] O. Gulbroden, "Optimal planning of marshalling yard by operation research," *ROCEEDINGS of the Symposium on the use of Cybernetics on the Railways*, pp. 226–233, 1963.
- [2] S. Yager, F. ESaccomanno, and Q. Shi, "An efficient sequencing model for humping in a railway," *Transportation Research A*, vol. 17, pp. 251–262, 1988.
- [3] M. Carey and S. Carville, "Scheduling and platforming trains at busy complex stations," *Transportation Research Part A*, vol. 37, no. 3, pp. 195–224, 2003.
- [4] R. M. Lentink, P. J. Fioole, L. G. Kroon, and C. Woudy, *Applying Operations Research Techniques to Planning Train Shunting*, John Wiley & Sons, Hoboken, NJ, USA, 2006.
- [5] P. Marton, J. Maue, and M. Nunkesser, "An improved train classification procedure for the hump Yard Lausanne Triage," in *Proceedings of the 9th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'09)*, September 2009.
- [6] C. G. Wang, "Study on wagon-flow allocating problem in a marshalling station by using calculating method on-table," *Journal of the China Railway Society*, vol. 24, no. 4, pp. 1–5, 2002.
- [7] S. He, *Optimization for railway hub work plan—study on model and algorithm of work plan for railway marshalling station and daily and shift plan for railway hub [Ph.D. thesis]*, Southwest jiaotong university, 1996.
- [8] S. He, R. Song, and S. S. Chaudhry, "Fuzzy dispatching model and genetic algorithms for railyards operations," *European Journal of Operational Research*, vol. 124, no. 2, pp. 307–331, 2000.
- [9] S. He, R. Song, and S. S. Chaudhry, "An integrated dispatching model for rail yards operations," *Computers and Operations Research*, vol. 30, no. 7, pp. 939–966, 2003.
- [10] T. Liu, S. W. He, B. H. Wang, and J. An, "Stochastic chance constrained programming model and solution of marshalling station dispatching plan," *Journal of the China Railway Society*, vol. 29, no. 4, pp. 12–17, 2007.
- [11] H. D. Li, S. W. He, R. Song, and L. Zheng, "Stochastic dependent-chance programming model and algorithm for stage plan of marshalling station," *Journal of Transportation Systems Engineering and Information Technology*, vol. 10, no. 1, pp. 128–133, 2010.
- [12] H. Li, S. He, B. H. Wang et al., "Survey of stage plan for railway marshalling station," *Journal of the China Railway Society*, vol. 33, no. 8, pp. 13–22, 2011.
- [13] B. Liu, *Theory and Practice of Uncertain Programming*, Physica, Heidelberg, Germany, 2002.
- [14] E. Dahlhaus, P. Horak, M. Miller, and J. F. Ryan, "The train marshalling problem," *Discrete Applied Mathematics*, vol. 103, no. 1–3, pp. 41–54, 2000.

- [15] M. Dorigo and T. Stutzle, *Ant Colony Optimization*, The MIT Press, Cambridge, Mass, USA, 2004.
- [16] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Optimization*, John Wiley & Sons, 2000.
- [17] W. J. Gutjahr, "A Converging ACO Algorithm for Stochastic Combinatorial Optimization," in *Proceedings of the Stochastic Algorithms: Foundations and Applications (SAGA '03)*, A. Al-Brecht and K. Steinhofel, Eds., vol. 9 of *Springer LNCS 2827*, pp. 10–25, 2003.
- [18] J. Wu, *Research on improved performance of ant colony algorithm by genetic algorithm [M.S. thesis]*, Taiyuan University of Technology, 2007.
- [19] Y. Jing and C. G. Wang, "Model and algorithm of dynamic wagon-flow allocating in a marshalling yard under uncertainty conditions," *Journal of the China Railway Society*, vol. 32, no. 4, pp. 8–12, 2010.

Research Article

Coordination Game Analysis through Penalty Scheme in Freight Intermodal Service

Jian Liu, Jianning Yu, and Yinzen Li

School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China

Correspondence should be addressed to Jian Liu, liujianlz@sohu.com

Received 10 May 2012; Revised 14 August 2012; Accepted 15 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Jian 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.

We study coordination mechanisms through penalty schemes to cooperate the behavior of two firms as successive segment carriers to make transport plan separate in freight intermodal market. Based on the different cost structure and service level constraint to two firms, we compare the decision making in two possible decision systems, that is, centralized system and decentralized system. In a centralized system—the first best case as a benchmark is contrasted with decentralized system. In the decentralized system, a Stackelberg game model is formulated between two firms. Some discordant decisions would be made by firm *I*'s overestimate motivation and firm *II*'s undersupply motivation. Our primary objective is to design penalty schemes to coordinate the interactions for two firms. The study shows in a decentralized system, setting suitable penalty schemes can coordinate the two firms' decision. We also study the feasible range of penalty parameters, and some important managerial insights are then deduced. In the end, a numerical example is provided to verify the validity of results, some concluding remarks are presented subsequently.

1. Introduction

Over the past decades, the increasing importance of international logistics has forced many firms to consider utilizing intermodalism to substantially decrease logistics costs. Intermodal freight transportation can be defined as the movement of goods from origins to destinations in one and the same loading unit or vehicle by successive transportation modes. Its goal is to provide an integration and effective seamless door-to-door service. During the whole operation process, there are always more than two separate firms involved in one intermodal freight transport service, therefore, the multiside participation is an outstanding feature, and the coordination or cooperation among the multiactor's is a core problem, which has been considered as a challenging issue by many practitioners and researchers.

Numerous studies have been done on the intermodal freight transportation industry. Macharis and Bontekoning [1] and Bontekoning et al. [2] provided an insightful review of the development and the related study in the intermodal freight transport. They argued that intermodal freight transportation research is emerging as a new transportation research application field, and it still is in a preparadigmatic phase. Specifically, because the physical network can be easily modeled as a network flow problem and there are especially efficient network flow algorithms, many researches use network to simulate actual intermodal operations [3–7]. However, most papers above focus on the operation management by single decision maker with the objective of expected profit maximization or expected cost minimization, and little account of interactions based on the multiactor's behavior is taken. Game theory is an appropriate tool for analyzing real situations where multiple agents are involved in decision and their actions are interrelated. Hurtely and Petersen [8] established a game-theoretic model to analyze the equilibrium behavior between carrier and shipper in freight transport market, by using a particular form of nonlinear tariff, they showed that the user equilibrium and system optimum can be simultaneously satisfied in an incomplete market. Xiao and Yang [9] subsequently developed a partially noncooperative game model among shippers, carriers, and infrastructure companies. Zhang et al. [10] examined the effect of multimodal integration between two different transport chains, they found an improvement in multimodal integration by a forwarder airline alliance, and it would not increase the alliance's output but improve both consumer surplus and total surplus.

Nowadays, with the change of production mode from centralized system to decentralized system, a new organization mode—virtual organization (VO)—is rising. The cooperation and coordination have become an important management issue with which more and more decentralized decision cases appeared in real business practice. It is a great challenge to traditional administration and management. There were extensive literatures which focus on the coordination among multiactors especially on supply chain management. Since the outcome in decentralized system is inefficient, cooperation among firms by means of coordination of actions may improve the individual profits. Nagarajan and Sobic [11] and Guardiola et al. [12] studied the cooperation in supply chain by cooperative game theory. Celikbas et al. [13] studied coordination mechanisms through penalty schemes between manufacturing and marketing departments which enable organizations to match demand forecasts with production quantities. Raju and Roy [14] studied a game model to understand how firm and industry characteristics moderate the effect of market information on cooperation. Cachon [15] reviewed the supply chain coordination with contracts, he discussed numerous supply chain models, and in each model the supply chain's actions are identified. Nevertheless, most studies on the coordination problem in a decentralized system are set in manufacturing industry and focus on how to coordinate interactions among supplier, manufacturer, and retailer in supply chains, seldom involve freight transport market, especially on how to coordinate participants' behavior (shipper, carrier, forwarder, etc.) in intermodal operation process. In above studies, the transport is not considered as an independent system but incorporated in the process of supply, manufacture, or retail.

In fact, based on the increase of global business, a kind of new organization structure, that is, VO with temporary, dynamic, and loose characteristics is being established in intermodal business process. It does require developing proper mechanisms to coordinate the behavior of all separate actors. In this paper, we develop incentive mechanisms for coordination actions to make transport plan between the two separate firms which offer complementary transportation service in an intermodal freight transport market. By comparing the performance of centralized and decentralized system under a stochastic

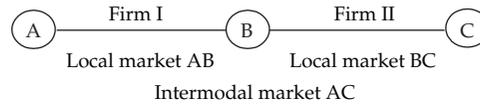


Figure 1: An intermodal market network.

demand, some theoretical analyses on game theory are deduced and some managerial insights are proposed subsequently.

The rest of the paper is organized as follows. Section 2 describes business background and sets up our basic model. Section 3 sets suitable penalty scheme to coordinate between the two separate firms. In Section 4, a case study is used to testify the propositions and results. In the end, some concluding remarks are presented in Section 5.

2. Model

2.1. Scenario and Notation

In order to describe the interactions among the different agents, we consider a simple intermodal network, as depicted in Figure 1, the network consists of three nodes, namely, A, B, and C. Consequently, there are three origin-destination markets, namely, AB, BC, and AC, of which AB involves ground transportation of cargo between A and B by truck, whereas BC involves transportation between B and C by train. While AC involves two different transport modes and may be referred to as a potential intermodal market. We assume there are two separate firms, firm *I* and firm *II*, that either control the transport infrastructure or provide complementary freight transportation service in AB and BC transport market separately. After market research, the two firms wish cooperatively in AC market to develop a long haul intermodal freight service, in which firm *I* is the first segment carrier and firm *II* is the second segment carrier. The two firms would make transport plan before providing the intermodal service. We discuss the two decision systems, that is, decentralized and centralized systems. In a centralized system—the first best case as a benchmark which contrasts with decentralized system, the two firms decide together on the quantities to distribute transport capacity by optimizing the total expected profits. In a decentralized system, firm *I* and firm *II* distribute transport capacity separately by maximizing their individual profit. That is a Stackelberg game actually. Firm *I*, firstly, as the first segment carrier, forecasts, demands, and decides the transport capacity. Firm *II* makes corresponding transport capacity decision based on the above decision. Because of different cost structure and opportunistic behavior of two firms, some discordant decisions would be made subsequently. In the game, firm *I* has overestimate motivation based on the restriction of service level and total cost, and then, by considering the opportunity cost, firm *II* always distributes less transport capacity in order to prevent capacity waste. (In China, e.g., the railway transport capacity is always in short supply. After considering factors such as types of car, stations, directions, among others, which is opportunity cost actually, the railway company often distributes less transport capacity than demand.) Their decisions are coupled. In order to solve the problem, a penalty scheme is designed to coordinate behaviors of two firms, that is, an overestimate penalty is charged to firm *I*, and an undersupply penalty is charged to firm *II*. The penalty scheme would be set up by the third party with authority.

To capture uncertainty in market demand, we assume the demand is a random variable, and the demand distribution is assumed to be known to both two firms. The timing of events is as follows. First, the third party with authority sets the overestimate penalties to firm *I* and the undersupply penalties to firm *II*. Second, a Stackelberg game is played between two firms. Firm *I* is the Stackelberg leader and making decision of distributing transport capacity and then reveal it to firm *II*. The firm *II*, the follower, determines its own capacity assignment plan based on this message. The final determinate intermodal transport capacity realization follows the minimum quantities of two firms' capacity assignment plan and the firms will be penalized if necessary.

The notation used in this paper is as follows: D —demand, assumed to be an absolutely continuous random variable; $f(x)$, $F(x)$ —density function and cumulative distribution function of D , and $f(x) > 0$; $F^{-1}(x)$ —inverse function of $F(x)$; P_{II}^u —under-supply penalty of transport capacity per item for firm *II*; P_I^o —over-estimate penalty per item for firm *I*; p_i —transport price per item for firm i , $i = I, II$; c_i —variable cost for firm i , $i = I, II$; c_{II}^w —opportunity cost per over-supply item for firm *II*; q_i —optimal transport capacity by firm i in the decentralized system, $i = I, II$; q_c —joint optimal capacity in the centralized system; α —given service level objective, $0 \leq \alpha \leq 1$; $I(\cdot)$ —0-1 indicator function, when (\cdot) is satisfied, then $I(\cdot) = 1$, otherwise $I(\cdot) = 0$; $E(\cdot)$ —expectation operator.

In this paper, we adopt the common assumption that all parameters of demand functions are common knowledge to both firms, and $p_i > c_i$ ($i = I, II$) is always satisfied.

2.2. Basic Model

In a decentralized system, the two separate firms distribute optimal transport capacity sequentially based on the objective of maximizing individual profit. We assume two firms with different cost structure. Firm *I*, as the first segment intermodal service provider, the variable cost is taken into consideration, and a given service level must be satisfied at the same time. For firm *II*, as the second carrier, except the variable cost, the waste cost from oversupply should be considered, which is an opportunity cost virtually. If not penalized, firm *I* will give an overestimate transport capacity to firm *II*, and firm *II* will assign less capacity than the given overestimate transport capacity. The final realized intermodal shipping volume which is decided on the minimum value among two firms' decision on transport capacity and the demand is correspondingly reduced. Therefore, a proper penalty scheme for two firms is designed necessarily. Firm *I*'s objective is given as follows:

$$\begin{aligned} \max R_I(q_I) &= (p_I - c_I) \min\{q_I, q_{II}(q_I), D\} - P_I^o [q_I - D]^+ I(D \leq q_{II}) \\ \text{s.t. } pr\{q_I \geq D\} &\geq \alpha. \end{aligned} \quad (2.1)$$

$R_I(q_I)$ is the profit function to firm *I*. The profit function consists of two parts. The first part is the revenue for providing intermodal service, where $\min\{q_I, q_{II}, D\}$ is the final realized intermodal shipping volume, and q_{II} is the transport capacity amount decided by firm *II* in response to firm *I*'s decision q_I . The second part is the penalty for overestimate, that means if the firm *II*'s decision q_{II} which is caused by firm *I*'s overestimate decision is greater than the realized demand D , then an overestimate penalty to firm *I* is given by $P_I^o [q_I - D]^+ I(D \leq q_{II})$, where $I(D \leq q_{II})$ is a 0-1 indicator function and $[q_I - D]^+ = \max\{q_I - D, 0\}$. Here, the firm *I* is penalized to the amounts $[q_I - D]^+$ only when the condition $I(D \leq q_{II}) = 1$ is satisfied.

That implies if the firm II 's capacity is always in short supply (less than the realized demand $q_{II} < D$), then the firm I should not be penalized for even an overestimate decision. The constraint means the probability that the transport demand should be satisfied is not less than the given service level α .

The firm II 's decision is to maximize the profit function itself:

$$\max R_{II}(q_{II}) = (p_{II} - c_{II}) \min\{q_I, q_{II}, D\} - c_{II}^w [q_{II} - \min\{D, q_I\}]^+ - P_{II}^u [\min\{q_I, D\} - q_{II}]^+. \quad (2.2)$$

The profit function consists of three parts. The first part is the revenue for providing intermodal service. The second part is the cost of waste transport capacity for making the oversupply transport plan, where $[q_{II} - \min\{D, q_I\}]^+$ are wastage. That implies if firm II distributes greater amount of transport capacity than the q_I or demand D , the waste cost would rise. It is the substantial opportunity cost. The third part is the penalty for the undersupply decision. If the decision q_{II} is less than the q_I or demand D , then q_{II} will be a bottleneck to the intermodal operation. Firm II , therefore, have to improve their decisions under the pressure of the penalty for undersupply.

Using backward induction algorithm to analyze the Stackelberg game. First, from problem (2.2), the following expect profit functions of R_{II} are derived. When $q_{II} \leq q_I$ is satisfied, we have

$$\begin{aligned} E(R_{II}(q_{II})) &= (p_{II} - c_{II}) \left[q_{II} \int_{q_{II}}^{\infty} f(x) dx + \int_0^{q_{II}} x f(x) dx \right] - c_{II}^w \int_0^{q_{II}} (q_{II} - x) f(x) dx \\ &\quad - P_{II}^u \left[\int_{q_I}^{\infty} (q_I - q_{II}) F(x) dx + \int_{q_{II}}^{q_I} (x - q_{II}) F(x) dx \right] \\ &= (p_{II} - c_{II}) \left[q_{II} - \int_0^{q_{II}} F(x) dx \right] - c_{II}^w \int_0^{q_{II}} F(x) dx - P_{II}^u \left[q_I - q_{II} - \int_{q_{II}}^{q_I} F(x) dx \right]. \end{aligned} \quad (2.3)$$

When $q_{II} \geq q_I$ is satisfied, the expect profit function is

$$\begin{aligned} E(R_{II}(q_{II})) &= (p_{II} - c_{II}) \left[\int_{q_I}^{\infty} q_I f(x) dx + \int_0^{q_I} x f(x) dx \right] \\ &\quad - c_{II}^w \left[\int_{q_I}^{\infty} (q_{II} - q_I) f(x) dx + \int_0^{q_I} (q_{II} - x) f(x) dx \right] \\ &= (p_{II} - c_{II}) \left[q_I - \int_0^{q_I} F(x) dx \right] - c_{II}^w \left[q_{II} - q_I + \int_0^{q_I} F(x) dx \right]. \end{aligned} \quad (2.4)$$

Based on the $f(x) > 0$, $p_i > c_i$, and $\partial^2 E(R_{II}) / \partial q_{II}^2 \leq 0$, it is easy to know $R_{II}(q_{II})$ is concave in q_{II} and the global optimal solution is existed.

When $q_{II} \geq q_I$, we have $\partial E(R_{II}(q_{II}))/\partial q_{II} = -c_b^w < 0$. Therefore, q_{II} that maximizes R_{II} should satisfy $q_{II} \leq q_I$. For $q_{II} \leq q_I$, from the first-order condition,

$$\begin{aligned} \frac{\partial E(R_{II}(q_{II}))}{\partial q_{II}} &= (p_{II} - c_{II})[1 - F(q_{II})] - c_{II}^w F(q_{II}) + P_{II}^u [1 - F(q_{II})] = 0 \\ \implies F(q_{II}^*) &= 1 - \frac{c_{II}^w}{p_{II} - c_{II} + P_{II}^u + c_{II}^w} \implies q_{II}^* = F^{-1}\left(1 - \frac{c_{II}^w}{p_{II} - c_{II} + P_{II}^u + c_{II}^w}\right). \end{aligned} \quad (2.5)$$

If $q_{II}^* \leq q_I$, then $q_{II} = q_{II}^*$ to maximize R_{II} , otherwise, $q_{II} = q_I$ to maximize R_{II} . Therefore, the reaction function of q_{II} is given by

$$q_{II}(q_I) = \min\left\{F^{-1}\left(1 - \frac{c_{II}^w}{p_{II} - c_{II} + P_{II}^u + c_{II}^w}\right), q_I\right\}. \quad (2.6)$$

From (2.6), we have $q_{II} \leq q_I$ that implies it is never optimal to firm II to assign more transport capacity than firm I 's decision. The q_I is always considered as upper boundary while the firm II makes decision to q_{II} , and with the increase of punishment P_{II}^u , q_{II} will increase subsequently.

Firm I 's decision is affected by the service level α . Here, we define that the service level α is the given probability to meet the intermodal demand. Based on the reaction function (2.6), the Firm I 's decision, that is, problem (2.1) can be described as the solution of the following programming problem:

$$\begin{aligned} \max R_I(q_I) &= R_I(q_I) = (p_I - c_I) \min\{q_I, q_{II}, D\} - P_I^o [q_I - D]^+ I(D \leq q_{II}) \\ \text{s.t. } \begin{cases} q_{II} = \min\left\{q_I, F^{-1}\left(1 - \frac{c_{II}^w}{p_{II} - c_{II} + P_{II}^u + c_{II}^w}\right)\right\}, \\ q_I \geq F^{-1}(\alpha), \end{cases} \end{aligned} \quad (2.7)$$

where the constraint condition $q_I \geq F^{-1}(\alpha)$ is from this transformation: $pr\{q_I \geq D\} \geq \alpha \implies q_I \geq F^{-1}(\alpha)$.

From the reaction function (2.6), when $q_I \geq F^{-1}(1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w))$ is satisfied, we have $q_{II} = F^{-1}(1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w))$, and the expect profit function is

$$\begin{aligned} E(R_I(q_I)) &= (p_I - c_I) \left[q_{II} \int_{q_{II}}^{\infty} f(x) dx - \int_0^{q_{II}} x f(x) dx \right] - P_I^o \int_0^{q_{II}} (q_I - x) f(x) dx \\ &= (p_I - c_I) \left[q_{II} - \int_0^{q_{II}} F(x) dx \right] - P_I^o \left[(q_I - q_{II}) F(q_{II}) + \int_0^{q_{II}} F(x) dx \right]. \end{aligned} \quad (2.8)$$

When $q_I \leq F^{-1}(1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w))$ is satisfied, we have $q_I = q_{II}$, and the expect profit function is

$$\begin{aligned} E(R_I(q_I)) &= (p_I - c_I) \left[q_I \int_{q_I}^{\infty} f(x) dx - \int_0^{q_I} x f(x) dx \right] - P_I^o \int_0^{q_I} (q_I - x) f(x) dx \\ &= (p_I - c_I) \left[q_I - \int_0^{q_I} F(x) dx \right] - P_I^o \int_0^{q_I} F(x) dx. \end{aligned} \quad (2.9)$$

It is easy to prove that $R_I(q_I)$ is concave in q_I , and the global optimal solution is existed by the function (2.8) and (2.9). The constraint in problem (2.7) implies that if $\alpha \geq 1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w)$ is satisfied, then the constraint conditions are converted into only one item, that is, $q_I \geq F^{-1}(\alpha)$. If $\alpha \leq 1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w)$ is satisfied, then the constraint in problem (2.7) is converted into two items, that is, $q_I \geq F^{-1}(1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w))$ (and then $q_{II} = F^{-1}(1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w))$) or $F^{-1}(\alpha) \leq q_I \leq F^{-1}(1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w))$, and then $q_I = q_{II}$. After solving the problem (2.7), we summarize the following Proposition.

Proposition 2.1. *In a Stackelberg game on making intermodal transport capacity plan, firm I is penalized for overestimate by p_I^o and firm II for undersupply by p_{II}^u , and the subgame perfect Nash equilibrium is follows:*

- (i) *if the objective service level satisfies the condition $\alpha > 1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w)$, then $q_I = F^{-1}(\alpha)$ and $q_{II} = F^{-1}(1 - (c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w)))$;*
- (ii) *if the objective service level satisfies the condition $\alpha \leq 1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w)$, then $q_I = q_{II}$ and*

$$q_I = \left\{ \begin{array}{l} F^{-1}\left(\frac{p_I - c_I}{p_I - c_I + P_I^o}\right), \text{ if } F^{-1}(\alpha) \leq F^{-1}\left(\frac{p_I - c_I}{p_I - c_I + P_I^o}\right) \leq F^{-1}\mathcal{A} \\ F^{-1}\mathcal{A}, \text{ if } F^{-1}(\alpha) \leq F^{-1}\mathcal{A} < F^{-1}\left(\frac{p_I - c_I}{p_I - c_I + P_I^o}\right) \\ F^{-1}(\alpha), \text{ if } F^{-1}\left(\frac{p_I - c_I}{p_I - c_I + P_I^o}\right) < F^{-1}(\alpha) \leq F^{-1}\mathcal{A} \end{array} \right\}, \quad (2.10)$$

where \mathcal{A} denotes $(1 - (c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w)))$.

Note that, from the equilibrium transport capacity expression in Proposition 2.1, we know when a higher service level is given ($\alpha > 1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w)$), we always have $q_I > q_{II}$. Meanwhile, the firm I's decision is determined by the given service level, and the penalties mainly restrict undersupply to firm II. When a lower service level is given ($\alpha \leq 1 - c_{II}^w / (p_{II} - c_{II} + P_{II}^u + c_{II}^w)$), we always have $q_I = q_{II}$. The penalties do not only restrict overestimate to firm I but restrict undersupply to firm II.

3. Penalty Scheme

In this section, we first analyze the centralized system—the first best case as a benchmark, wherein two firms maximize their joint profit. In a centralized system, firm I is still the first

segment carrier and firm II is the second segment carrier, and they join together to determine the amount q_c of intermodal transport plan with the objective of maximizing their combined profits. The objective for maximizing joint profit is given as

$$\begin{aligned} \max R_c &= (p_I + p_{II} - c_I - c_{II}) \min\{q_c, D\} - c_{II}^w [q_c - \min\{q_c, D\}]^+ \\ \text{s.t. } pr\{D \leq q_c\} &\geq \alpha, \end{aligned} \quad (3.1)$$

where $\min\{q_c, D\}$ is the final realized intermodal shipping volume, $[q_c - \min\{q_c, D\}]^+$ is the waste capacity for overestimate. In order to draw comparison with the decentralized system easily, the transport revenue, variable cost, and opportunity cost on overestimate are included in the joint profit objective.

From (3.1), we know the expect profit function of R_c is

$$E(R_c) = (p_I + p_{II} - c_I - c_{II}) \left[q_c - \int_0^{q_c} F(x) dx \right] - c_{II}^w \int_0^{q_c} F(x) dx. \quad (3.2)$$

The constraint condition is converted to $q_c \geq F^{-1}(\alpha)$, so the optimal solution of function (3.1) is

$$q_c = \max \left\{ F^{-1} \left(1 - \frac{c_{II}^w}{p_I + p_{II} - c_I - c_{II} + c_{II}^w} \right), F^{-1}(\alpha) \right\}. \quad (3.3)$$

Proposition 3.1. *In a centralized system, with the constraint of service level α , the joint transport capacity optimal decision is $q_c = \max\{F^{-1}(1 - (c_{II}^w / (p_I + p_{II} - c_I - c_{II} + c_{II}^w))), F^{-1}(\alpha)\}$.*

Proposition 3.1 shows that given $\alpha^* = 1 - c_{II}^w / (p_I + p_{II} - c_I - c_{II} + c_{II}^w)$ is a critical point. When $\alpha < \alpha^*$ is satisfied, that is, setting a lesser objective service level, the programming (3.1) is a no constraint problem actually, and the final transport capacity is $q_c = F^{-1}(1 - c_{II}^w / (p_I + p_{II} - c_I - c_{II} + c_{II}^w))$ which is the optimal solution to maximize the joint profit, but it is not optimal to shipper. Meanwhile, the carrier has greater welfare. When $\alpha > \alpha^*$ is held, the final transport capacity is $q_c = F^{-1}(\alpha)$. Though the whole transport capacity q_c increases, however, it is not optimal to maximize the joint profit. Meanwhile, the shipper has greater welfare for more consumer surplus.

Next, we develop penalty schemes so that the decentralized system performs as well as the centralized system. Based on the Proposition 3.1, we consider two cases with different constraint of service level. For the convenience to the analysis subsequently, the following notations are introduced:

$$\begin{aligned} F_1 &= F^{-1} \left(1 - \frac{c_{II}^w}{p_{II} - c_{II} + P_{II}^u + c_{II}^w} \right), & F_2 &= F^{-1} \left(\frac{p_I - c_I}{p_I - c_I + P_I^o} \right), \\ F_3 &= F^{-1} \left(1 - \frac{c_{II}^w}{p_I + p_{II} - c_I - c_{II} + c_{II}^w} \right). \end{aligned} \quad (3.4)$$

Case 1. If $\alpha \leq \alpha^*$, then the mathematical programming (3.1) is an unconstrained problem, and the optimal solution $q_c = F_3$. Let

$$\begin{aligned} F_1 = F_3, \quad F_2 = F_3 \\ \implies P_{II}^u = p_I - c_I = \Delta_1, \quad P_I^o = \frac{(p_I - c_I)c_{II}^w}{p_I + p_{II} - c_I - c_{II}} = \Delta_2. \end{aligned} \quad (3.5)$$

From Proposition 3.1, we have $q_I = q_{II} = q_c = F_3$, the transport capacity decisions in both decentralized system and centralized system are identical and the penalty scheme coordinates the two firms' behaviors. After further discussion, Proposition 3.2 is summarized subsequently.

Proposition 3.2. *When the condition $\alpha \leq \alpha^*$ is satisfied, if (i) $P_{II}^u \geq \Delta_1$, $P_I^o = \Delta_2$, or (ii) $P_{II}^u = \Delta_1$, $P_I^o < \Delta_2$, then the capacity decisions in both decentralized system and centralized system are identical. Specially, $q_I = q_{II} = q_c = F_3$.*

Proposition 3.2 implies when a smaller service level is given, the suitable penalty scheme is a strong deterrent to two firms. It can improve the final equilibrium in a decentralized system effectively.

Case 2. If $\alpha > \alpha^*$ then $q_c = F^{-1}(\alpha)$ from Proposition 3.1, let

$$\begin{aligned} F_1 = F^{-1}(\alpha); \quad F_2 = F^{-1}(\alpha) \\ \implies P_{II}^u = \frac{c_{II}^w - (1 - \alpha)(p_{II} - c_{II} + c_{II}^w)}{1 - \alpha} = \Delta_3, \quad P_I^o = \frac{(1 - \alpha)c_{II}^w}{\alpha} = \Delta_4. \end{aligned} \quad (3.6)$$

When $P_{II}^u = \Delta_3$, then the constraint in programming (2.7) is $q_I \geq F_1 = F^{-1}(\alpha) = q_c$ one item, after the similar analysis that is used in Proposition 2.1(i), we know if $\alpha > \alpha^*$, then in a decentralized system the actions of two firms would be coordinated by setting a penalty scheme $P_{II}^u = \Delta_3$, and then we will have $q_I = q_{II} = q_c = F_1 = F^{-1}(\alpha)$. Because the given service level α is an only constraint to the decision by firm I , and the given service level α is always consistent with the firm II 's optimal decision. Hence, the penalty P_I^o is insignificant. The fact lies behind the penalty is to urge firm II to assign more transport capacity. After further similar discussion, the main result is summarized in Proposition 3.3.

Proposition 3.3. *If $\alpha > \alpha^*$, then the actions of two firms would be coordinated by designing penalty schemes. Specially (i) $P_{II}^u > \Delta_3$ and $P_I^o \geq \Delta_4$ or (ii) $P_{II}^u = \Delta_3$.*

Proposition 3.3 implies that, when the service level α is relatively higher ($\alpha \geq \alpha^*$), the two firms would be coordinated only by giving $P_{II}^u = \Delta_3$ to firm II . Meanwhile, firm I has to prepare greater amount transport capacity to satisfy the constraint for a higher service level, and the P_I^o lost its significance. When setting lower penalties to firm II ($P_{II}^u < \Delta_3$) the two firms actions would not be coordinated. Meanwhile, firm II 's decisions are not improved by setting smaller P_{II}^u which lacks a strong deterrent to undersupply. Therefore, firm I 's decision is always larger than firm II 's, and it is impossible to coordinate the two firms' action by setting penalties.

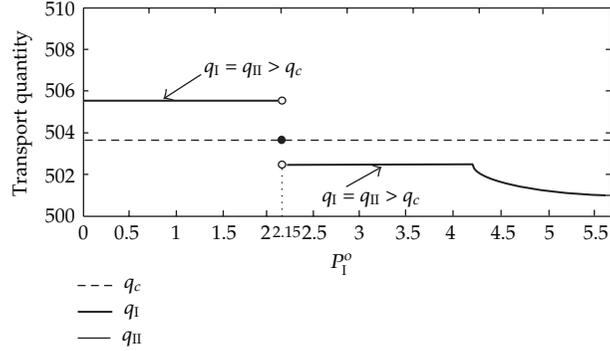


Figure 2: Coordination effects with $P_{II}^U > \Delta_1$ and $\alpha < \alpha^*$.

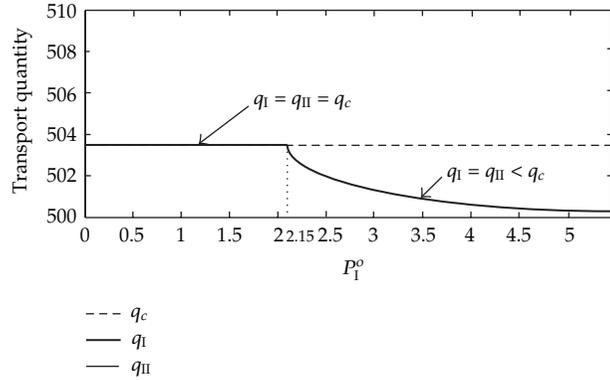


Figure 3: Coordination result under penalties (specially $P_I^U = \Delta_1$ and $\alpha < \alpha^*$).

4. Numerical Study

In this section, we introduce a numerical example to simulate the incentive of coordination mechanisms to separate carriers. As depicted in Figure 1, we adopt the same scenario as in Section 2.1, there are two transport firms with complementary transport model to develop AC intermodal service. A penalty scheme is designed to coordinate the actions between two firms. All parameters which used in example are given, where intermodal freight demand follows normal distribution which the mean value is 500 and the variance is 25, and $P_I = 10$, $P_{II} = 8$, $c_I = 3$, $c_{II} = 2$, $c_{II}^w = 4$. Without loss of generality, we assume α^* is the critical value of service level and then $\alpha^* = 0.765$ which is based on the equation $\alpha^* = 1 - c_{II}^w / (p_I + p_{II} - c_I - c_{II} + c_{II}^w)$. In this section, two cases which are represented $\alpha \geq \alpha^*$ and $\alpha < \alpha^*$ are studied separately to testify the conclusions in the paper.

In the case of $\alpha < \alpha^*$, let $\alpha = 0.665$, after calculation, we have $\Delta_1 = 7$, $\Delta_2 = 2.15$, $q_c = 503.6$, $\alpha c_{II}^w / (1 - \alpha) + c_{II} - p_{II} = 2$, $F^{-1}(\alpha) = 502.1$. When $P_{II}^u \geq \Delta_1$, let $P_{II}^u = 8$ and $P_{II}^u = 7$ simulate the variation of the game equilibrium, which are depicted in Figures 2 and 3. Let $P_{II}^u = 5$, and $P_{II}^u = 1$ denote the constraint $\alpha c_{II}^w / (1 - \alpha) + c_{II} - p_{II} < P_{II}^u < \Delta_1$ and $P_{II}^u < \alpha c_{II}^w / (1 - \alpha) + c_{II} - p_{II} < \Delta_1$ to simulate the variation of the equilibrium. The final results are shown in Figures 4 and 5.

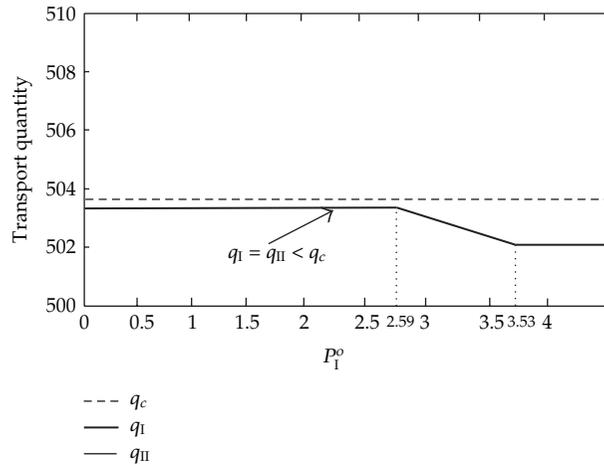


Figure 4: Coordination result under penalties (specially $\alpha c_{II}^w / (1 - \alpha) + c_{II} - p_{II} < P_{II}^u < \Delta_1$ and $\alpha < \alpha^*$).

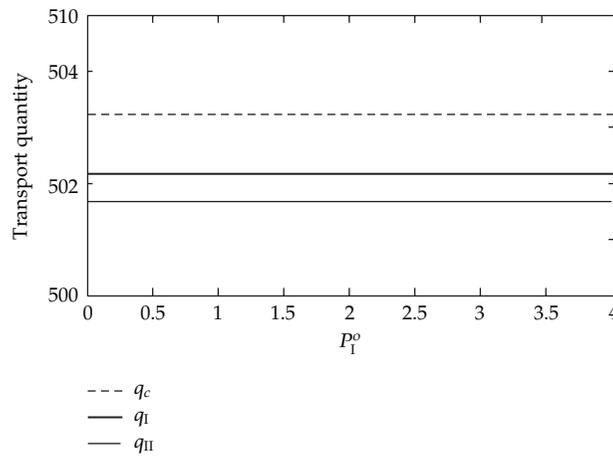


Figure 5: Coordination result under penalties (specially $P_{II}^u < \alpha c_{II}^w / (1 - \alpha) + c_{II} - p_{II} < \Delta_1$ and $\alpha < \alpha^*$).

From Figure 2, we find the three curves q_I, q_{II}, q_c meet at a point if $P_I^o = 2.215$, and the three curves are superposed if $P_I^o \leq 2.15$ in Figure 3. From Figures 5 and 6, the three curves are never superposed. That implies if $P_{II}^u \geq \Delta_1, P_I^o = \Delta_2$, or $P_{II}^u = \Delta_1, P_I^o < \Delta_2$ are satisfied, then $q_I = q_{II} = q_c$, the actions of two firms are coordinated.

In the case of $\alpha \geq \alpha^*$, we have $\Delta_3 = 15.8, \Delta_4 = 0.734, q_c = 505.5$. We also analyze the equilibrium with the variance of penalties. The final results are shown in Figures 6, 7, and 8, where Figure 6 depicts the equilibrium when $P_{II}^u = 16 (P_{II}^u > \Delta_3)$, Figure 7 depicts the equilibrium when $P_{II}^u = 15.8 (P_{II}^u = \Delta_3)$, and Figure 8 depicts the equilibrium when $P_{II}^u = 14 (P_{II}^u < \Delta_3)$, respectively.

From Figure 6, we find the three curves are superposed when $P_I^o \geq 0.734$, the three curves are superposed in Figure 7. In Figure 8, nevertheless, there are only two curves (q_I, q_c) coincident each other, and the curve q_{II} lies below other two curves. That means when $\alpha \geq \alpha^*$,

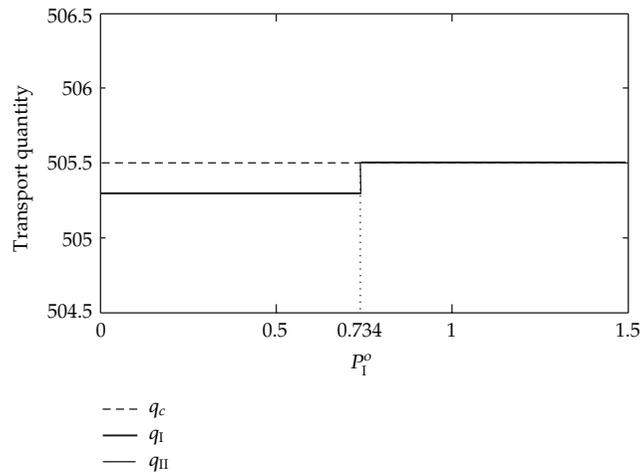


Figure 6: Coordination result under penalties (specially $P_{II}^U > \Delta_3$ and $\alpha \geq \alpha^*$).

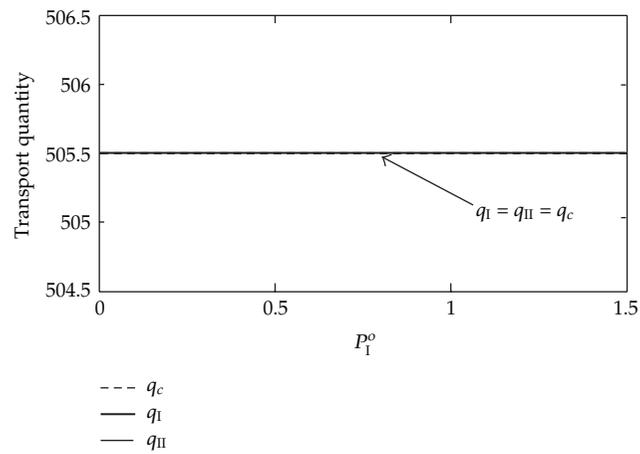


Figure 7: Coordination result under penalties (specially $P_{II}^U = \Delta_3$ and $\alpha \geq \alpha^*$).

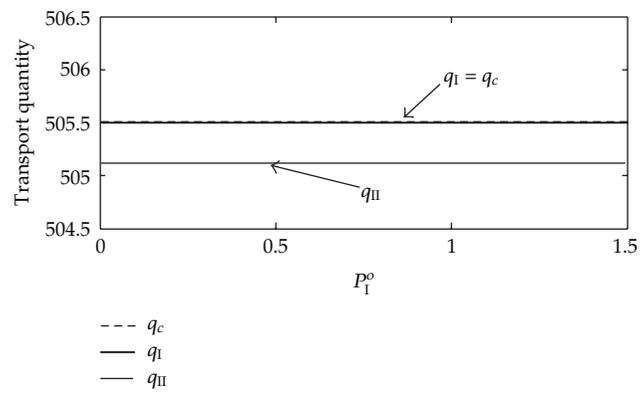


Figure 8: Coordination result under penalties (specially $P_{II}^U < \Delta_3$ and $\alpha \geq \alpha^*$).

if $P_{II}^u > \Delta_3$, $P_I^o \geq \Delta_4$, or $P_{II}^u = \Delta_3$ are satisfied, then $q_I = q_{II} = q_c$, the actions of two firms are coordinated.

The above two case studies show the discordant behaviors would be coordinated by setting suitable coordination mechanisms in a decentralized decision system, and all individual profits would be improved correspondingly. Therefore, all results are consistent with the Propositions in the paper, the validity of Propositions is consequently testified.

5. Conclusions

In this paper, a coordinate problem on making freight plan between two separate transport carriers which provide complementary transport service jointly to develop a long haul intermodal service is studied. Two possible decision systems—centralized and decentralized—are taken into consideration, our primary objective is to develop the coordination mechanisms through penalty schemes to coordinate the interactions for two firms in decentralized decision system. In the centralized case, two firms jointly decide on the transport capacity assignment. In the decentralized case, we model a single period problem as a Stackelberg game. Firm *I*, the leader, decides transport capacity to the intermodal service. Firm *II*, the follower, makes transport capacity assignment based on firm *I*'s action subsequently. Due to the different cost structure and opportunistic behavior by two firms, some discordant decisions would be made subsequently. After detailed models analysis by comparing the final equilibrium made in the two decision systems, some managerial insights are induced. Among other results, we show that one can generate the same result in a decentralized system as what obtained from a centralized system by setting suitable penalties, and that the service level restriction is a significant factor to setting the correct penalty scheme. We also discuss in details the feasible range of penalties to coordinate two firms' decision. All the study in this paper is under the framework of complete information. For the future research, the authors plan to extend the model to the incomplete information, and to make it closer to the real world.

Acknowledgment

The authors would like to thank the two anonymous referees for helpful comments. This research was supported in part by the Natural Sciences Foundation of China under Grant no. 50968009 and no. 61164003.

References

- [1] C. Macharis and Y. M. Bontekoning, "Opportunities for OR in intermodal freight transport research: a review," *European Journal of Operational Research*, vol. 153, no. 2, pp. 400–416, 2004.
- [2] Y. M. Bontekoning, C. Macharis, and J. J. Trip, "Is a new applied transportation research field emerging?—a review of intermodal rail-truck freight transport literature," *Transportation Research A*, vol. 38, no. 1, pp. 1–34, 2004.
- [3] F. Southworth and B. E. Peterson, "Intermodal and international freight network modeling," *Transportation Research C*, vol. 8, no. 1–6, pp. 147–166, 2000.
- [4] B. S. Boardman, E. M. Malstrom, D. P. Butler, and M. H. Cole, "Computer assisted routing of intermodal shipments," *Computers and Industrial Engineering*, vol. 33, no. 1–2, pp. 311–314, 1997.
- [5] A. Muriel and F. N. Munshi, "Capacitated multicommodity network flow problems with piecewise linear concave costs," *IIE Transactions*, vol. 36, no. 7, pp. 683–696, 2004.

- [6] A. Amiri and H. Pirkul, "New formulation and relaxation to solve a concave-cost network flow problem," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 278–287, 1997.
- [7] J. Holguín-Veras and E. Thorson, "Modeling commercial vehicle empty trips with a first order trip chain model," *Transportation Research B*, vol. 37, no. 2, pp. 129–148, 2003.
- [8] W. J. Hurley and E. R. Petersen, "Nonlinear tariffs and freight network equilibrium," *Transportation Science*, vol. 28, no. 3, pp. 236–245, 1994.
- [9] F. Xiao and H. Yang, "Three-player game-theoretic model over a freight transportation network," *Transportation Research C*, vol. 15, no. 4, pp. 209–217, 2007.
- [10] A. Zhang, C. Lang, Y. V. Hui, and L. Leung, "Intermodal alliance and rivalry of transport chains: the air cargo market," *Transportation Research E*, vol. 43, no. 3, pp. 234–246, 2007.
- [11] M. Nagarajan and G. Sošić, "Game-theoretic analysis of cooperation among supply chain agents: review and extensions," *European Journal of Operational Research*, vol. 187, no. 3, pp. 719–745, 2008.
- [12] L. A. Guardiola, A. Meca, and J. Timmer, "Cooperation and profit allocation in distribution chains," *Decision Support Systems*, vol. 44, no. 1, pp. 17–27, 2007.
- [13] M. Celikbas, J. George Shanthikumar, and J. M. Swaminathan, "Coordinating production quantities and demand forecasts through penalty schemes," *IIE Transactions*, vol. 31, no. 9, pp. 851–864, 1999.
- [14] J. S. Raju and A. Roy, "Market information and firm performance," *Management Science*, vol. 46, no. 8, pp. 1075–1084, 2000.
- [15] G. Cachon, "Supply chain coordination with contracts," in *Handbooks in Operations Research and Management Science*, pp. 229–246, North-Holland, Amsterdam, The Netherlands, 2003.

Research Article

Research on the Behavior Characteristics of Pedestrian Crowd Weaving Flow in Transport Terminal

Liya Yao,¹ Lishan Sun,² Zhiyong Zhang,² Shuwei Wang,² and Jian Rong²

¹ Department of Transport Engineering, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China

² Key Laboratory of Traffic Engineering, Beijing University of Technology, 100 Pingleyuan, Chaoyang District, Beijing 100124, China

Correspondence should be addressed to Liya Yao, yaoliya@bit.edu.cn

Received 27 April 2012; Revised 19 July 2012; Accepted 4 August 2012

Academic Editor: Wuhong Wang

Copyright © 2012 Liya Yao 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.

Due to the poor transfer organization in urban public transport terminal, pedestrian crowd are often forced to weaving in their transfer flow lines. Frequent weaving behaviors not only decrease passengers' transfer comfort, but may also trigger serious crowd disaster such as trampling. In order to get accurate understanding of the weaving features of pedestrian crowd and analyze the relevant evolution law, researches have been conducted on the basis of field investigation. First, the typical weaving phenomenon were defined and classified, and a microscopic parameters system of pedestrian crowd weaving flow was constructed. The detection and quantification methods of multiple indicator parameters were also given. Then, correlation between different behavioral parameters was analyzed based on the survey data of weaving pedestrian crowd on the stairs of DongZhiMen (DZM) hub. The basic characteristics and evolution law of the weaving behaviors were then discussed, and conclusions were drawn.

1. Introduction

Along with the increasing volume of transfer pedestrians, the transferring service level in some existing urban public transport terminals is greatly decreased by the crowding and interweaving pedestrian flow. How to improve the transfer efficiency of transport terminal through optimizing the organization of pedestrian flow has become an interesting thing.

In the past years, scholars at home and abroad have achieved great progress in researches about characteristics and evolution law of the weaving behaviors of pedestrian flow in transport terminals.

In 1971, Henderson analyzed the statistical characteristics of high density pedestrian flow [1]. Helbing and Molnár introduced the principle of fluid dynamics into pedestrian characteristics analysis in 1995 and carried out the numerical simulation research of pedestrian flow [2]. Since then, scholars over all countries started to conduct extensive and thorough research in pedestrian flow characteristics detection and regularity analysis [3–9]. Ukkusuri et al. [10], Laxman et al. [11] and Lam et al. [12] studied the characteristics of pedestrian flow at certain transportation facilities. Siram et al. [13], Zhou et al. [14], and Wang et al. [15] explained the particular phenomenon in elderly pedestrians and paired pedestrians. Hughes [16] gave the maximum speed and density of pedestrian flow. The relationship between pedestrian speed and density was studied by Ando et al. [17], Thompson and Marchant [18], Hughes [19], Hankin and Wright [20], and so on. Cheung and Lam [21] and Tanaboriboon et al. [22] studied the relationship between pedestrian volume and density. On this basis, the 2010 issue of the American Capacity Manual summarized the characteristics of most types of pedestrian flow [23].

With the analysis and brief history above as a backdrop, it is clear that the conventional pedestrian flow studies have made progress. In recent years, scholars began to focus on several special phenomenon analysis of pedestrian flow [24–26]. As an important phenomenon of transfer activities, pedestrian crowd weaving flow generally exists in transport terminals of China [27]. Frequent intertwined behaviors in areas such as entrance, ticket hall, platform, and stairs not only decrease the transfer comfort of pedestrian crowd, but also may trigger serious crowd disaster such as trampling. It is still a problem of how to grasp the characteristics of weaving transfer pedestrian crowds and analysis the relevant evolution law accurately.

This paper describes a survey and analysis method for typical intertwined flow of pedestrian crowd. The remainder of the paper is structured as follow. Section 2 discusses the definition and classification of typical weaving phenomenon. Section 3 constructs a microscopic parameters system of weaving crowd. Section 4 presents the survey method for weaving pedestrian crowd and demonstrates its application to DZM hub in Beijing, China. Analysis results are also presented in Section 4, followed with the conclusions in Section 5.

2. Passenger Weaving Flow Categorization

Due to the limited space in transport terminal, there are often many conflict points among the pedestrian flow lines. Pedestrian crowd are forced to weaving at the conflict points frequently during their transfer process. Summarizing the weaving law of pedestrian crowd is the foundation to carry out analysis of various types of weaving phenomenon. In this study, weaving pedestrian crowd can be defined as a traffic phenomenon that more than two pedestrian crowd flows with transfer purpose confluence or shunt continuously in a short distance.

According to the definition and characteristics analysis of weaving pedestrian crowd, combined with field investigation, weaving phenomenon of pedestrian crowd can be classified from three aspects as follows.

- (1) According to the weaving angle, weaving phenomenon can be divided into forward weaving behavior, cross weaving behavior, and lateral weaving behavior, as shown in Figure 1.
- (2) According to the law of weaving occurrence, weaving phenomenon can be divided into regular interweave and sudden interweave.

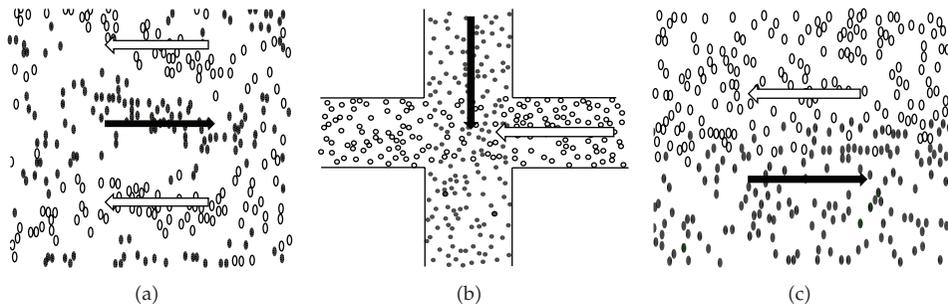


Figure 1: Schematic diagram of weaving pedestrian crowd—forward weaving flow (a), cross weaving flow (b), and lateral weaving flow (c).

- (3) According to the boundary condition, weaving phenomenon can be divided into boundary interweave and no boundary interweave.

3. Characteristics of Weaving Pedestrian Crowd

Different with the conventional of pedestrian crowd, the weaving pedestrian crowd show special characteristics as follows.

- (1) Spatial constraint. In a transport terminal with good guidance system, pedestrian crowd usually have relatively fixed directions. But due to the limited internal space, most of the pedestrians are forced to move laterally.
- (2) Time constraint. Due to the impact of subway departure time, passengers arrive periodically, with high walking speed, which leading to a more frequent acceleration, deceleration, and way finding phenomenon in the weaving process.

Due to the spatial-time constraints above, pedestrian flow show more frequent crowding and collision, leading to a fluctuation of traffic density and speed.

Trajectory tracing and interaction analysis are required in the pedestrian crowd's numerical simulation. For pedestrian crowd tracks and interaction, behavior indicators which can accurately describe the pedestrian behavior are needed to explain the weaving process. Considering the spatial-time restriction of pedestrian crowd interweaving phenomenon, speed variation and lateral movement were selected together with conventional indicators such as passenger flow, average speed, average density, stride length, and frequency [28]. Definition and quantify method for indicators are as follows.

- (1) Pedestrian crowd volume (person/meter/second): the quantity of pedestrian crowd that going through the weaving section within a unit time period as (3.1)

$$Q = D \times V, \quad (3.1)$$

where Q represents the pedestrian crowd flow, D represents the average density of weaving area, and V represents the average speed of pedestrian crowd in weaving area.

- (2) Average walking speed (meter/second): the average time that pedestrian crowd walk through a certain distance within the weaving area as follows:

$$V = \frac{L}{\sum_{i=1}^N t_i / N}, \quad (3.2)$$

where L represents walking distance, N represent the number of pedestrian crowd observed, and t_i represents the walking time of pedestrian i .

- (3) Average density (person/square meter): the quantity of pedestrian crowd within a unit area of the weaving area as (3.3)

$$D = \frac{Q}{S}, \quad (3.3)$$

where D represents average density, Q represents the number of pedestrian crowd within the weaving area, and S represents the measure of the weaving area.

- (4) Stride length (m): the distance that pedestrian go through by one step in weaving situation. Stride length is influenced mainly by the flow density, psychological status of pedestrian crowd, and other factors. The stride length can be estimated through the length of the floor tiles or a grid drawn on the weaving area.
- (5) Stride frequency (number/second): the numbers of steps pedestrian walk during a unit time period in weaving situation. Stride frequency is influenced mainly by the flow density, transfer facility status, and other factors.
- (6) Speed variation (meter/second): the maximum increase or decrease value of pedestrian crowd's walking speed within weaving area.
- (7) Lateral movement probability: the percentage of pedestrian that are forced to deviated from their original path due to the weaving behavior.

4. Investigation Method of Weaving Phenomenon

Forward-boundary-regular weaving phenomenon appears very frequently in a transport terminal, which is also the main factor reducing the hub's operation efficiency. Choose DZM station of Beijing subway as investigation object, illustrate the acquire method of weaving behavior indicators, and then analyze the characteristics of pedestrian crowd weaving flow.

(1) Investigation Time

Weaving phenomenon in a transport terminal occurs most frequently during peak hours, which has the maximum interference to terminal operation. So the investigation time was selected as working day 8:00-9:00 (A.M) and 17:00-18:00 (P.M).

(2) Investigation Area

Weaving phenomenon of pedestrian crowd tends to occur in areas with high density and complex streamlines, such as stairs and platform areas and so forth. The connecting area

Table 1: Investigate methods for behavior indicators of weaving pedestrian crowd.

Investigation method	Observing point number	Indicator			
Manual	Fixed point observation	1, 2	Pedestrian crowd volume	Stride length	—
	Following investigation	—	Average volume	Stride frequency	—
Video	Vertical recording	3	Average density	Speed variation	Lateral movement probability
	Parallel recording	4	Stride length	—	—

Table 2: Behavior indicator data of weaving pedestrian crowd flow.

	Average stride length (m)	Average stride frequency (n/s)	Maximum acceleration (m/s ²)	Speed variation (m/s)	Lateral movement probability (%)
Without weaving behavior	0.6	1.262854	-0.31	0.17	0.089
With weaving behavior	0.2	1.213844	-5.81	2	0.55

between stairs and platform of DZM hub was chosen as the investigation area. Weaving behaviors within this area can be regarded as a forward-boundary-regular one.

(3) Data Collection and Analysis Method

Aiming at the characteristics of short weaving duration, high walking speed, and so forth, manual recording and video recording were combined in the investigation. Manual recording method was used to get pedestrian volume, average speed, stride length, and frequency; video recording method was used to get average density, speed variation, and lateral movement probability, as shown in Table 1. Figure 2 shows the distribution of data collection points.

A statistical analysis of the walking speed and pedestrian crowd density were conducted, results show that the weaving phenomenon takes 140 seconds for a periodic regularity of occurrence and dissipation, as shown in Figures 3 and 4. This meets the actual arrival and departure regularity of metro vehicle.

On these basis, record the related parameters of interweave phenomena by the cycle of 140 seconds according to weaving phenomenon occurrence regularity, get the relationship between pedestrian crowd volume, speed, and density of interweaving phenomenon, as shown in Figures 5, 6 and 7, the value of other related indicators are shown in Table 2.

5. Conclusions

Aiming at the pedestrian weaving behavior in transport terminal, this paper defined and classified the pedestrian crowd weaving flow and constructed a characteristic indicator system. The investigation and quantization methods for indicators were also given. Then summarize the basic regularity of pedestrian crowd interweaving phenomenon through field

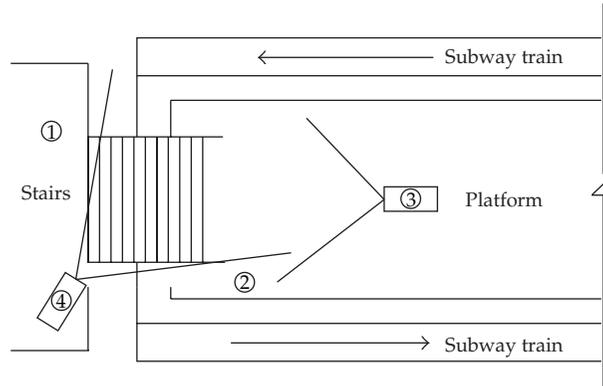


Figure 2: Investigate scheme for behavior indicators of weaving pedestrian flow.

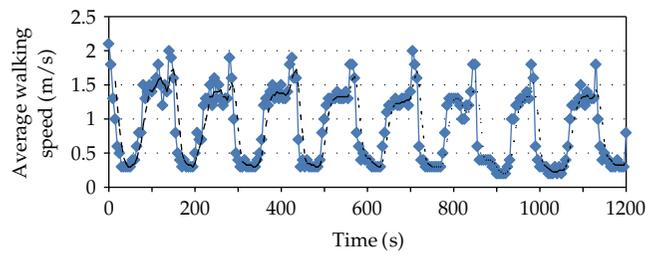


Figure 3: Walking speed change of weaving pedestrian crowd.

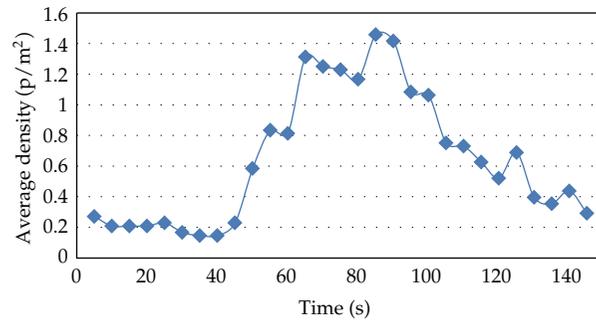


Figure 4: Average density change of weaving pedestrian crowd.

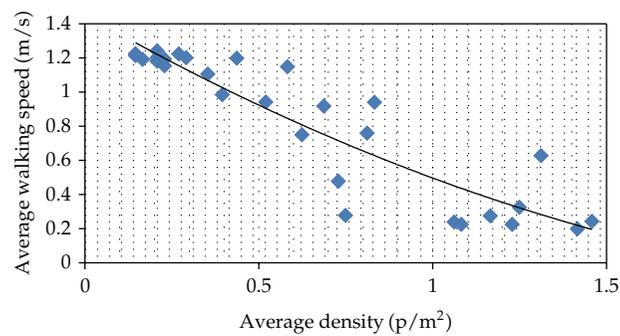


Figure 5: Correlation between the walking speed and density of weaving pedestrian crowd flow.

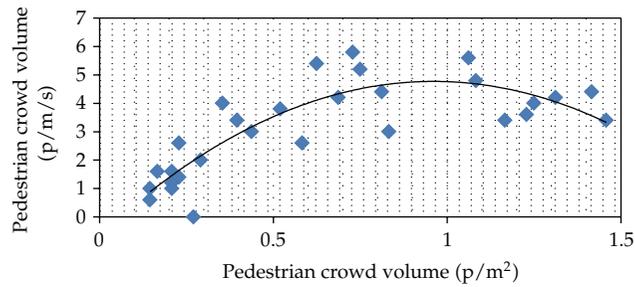


Figure 6: Correlation between the pedestrian crowd volume and average density of pedestrian crowd.

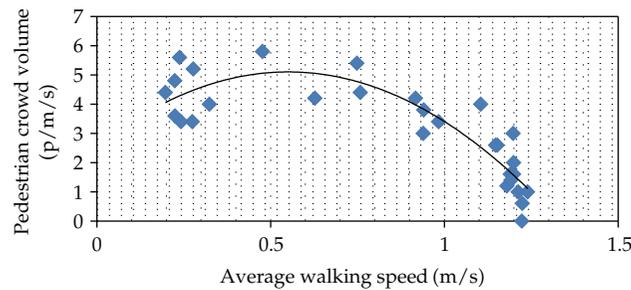


Figure 7: Correlativity between the pedestrian crowd volume and average walking speed.

investigation of Beijing DZM transport terminal, which provides a reference for the practical work of high density pedestrian crowd control and transfer facility design. Compared with the conventional pedestrian crowd, pedestrian crowd weaving behaviors show different regularities as follows.

- (1) Pedestrian crowd weaving phenomenon appears relatively apparent periodic with three basic states: begin, spread, and dissipated. And the cycle length is closely related to the arrive and departure time of mass transit way.
- (2) Due to the limitation of space and time, pedestrian of different flow line need to share space and are forced to adjust their behavior frequently. This is the reason why collisions among weaving pedestrians occurred more frequently.
- (3) Although the indicators of pedestrian weaving behavior showed strong complexity and variability, but indicators such as pedestrian crowd volume, speed, and density still has a strong correlation in accordance with the basic characteristics of fluid mechanics.

Acknowledgments

This research was supported by National Nature Science Foundation of China (no. 51108028) and Beijing Municipal Natural Science Foundation (no. 8122009) and Programme of Introducing Talents of Discipline to Universities (no. B12022) and 973 Program (no. 2012CB725403).

References

- [1] L. F. Henderson, "The statistics of crowd fluids," *Nature*, vol. 229, no. 5284, pp. 381–383, 1971.
- [2] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995.
- [3] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research Part C*, vol. 50, no. 8, pp. 1704–1714, 2011.
- [4] J. K. K. Yuen and E. W. M. Lee, "The effect of overtaking behavior on unidirectional pedestrian flow," *Safety Science*, vol. 50, no. 8, pp. 1704–1714, 2012.
- [5] H. Gotoh, E. Harada, and E. Andoh, "Simulation of pedestrian contra-flow by multi-agent DEM model with self-evasive action model," *Safety Science*, vol. 50, no. 2, pp. 326–332, 2012.
- [6] B. Piccoli and A. Tosin, "Time-evolving measures and macroscopic modeling of pedestrian flow," *Archive for Rational Mechanics and Analysis*, vol. 199, no. 3, pp. 707–738, 2011.
- [7] M. Di Francesco, P. A. Markowich, J. F. Pietschmann, and M. T. Wolfram, "On the Hughes' model for pedestrian flow: the one-dimensional case," *Journal of Differential Equations*, vol. 250, no. 3, pp. 1334–1362, 2011.
- [8] C. Dogbe, "On the Cauchy problem for macroscopic model of pedestrian flows," *Journal of Mathematical Analysis and Applications*, vol. 372, no. 1, pp. 77–85, 2010.
- [9] B. Steffen and A. Seyfried, "Methods for measuring pedestrian density, flow, speed and direction with minimal scatter," *Physica A*, vol. 389, no. 9, pp. 1902–1910, 2010.
- [10] S. Ukkusuri, S. Hasan, and H. M. A. Aziz, "Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency," *Transportation Research Record*, no. 2237, pp. 98–106, 2011.
- [11] K. K. Laxman, R. Rastogi, and S. Chandra, "Pedestrian flow characteristics in mixed traffic conditions," *Journal of Urban Planning and Development*, vol. 136, no. 1, pp. 23–33, 2010.
- [12] W. H. K. Lam, J. Y. S. Lee, and C. Y. Cheung, "A study of the bi-directional pedestrian flow characteristics at Hong Kong signalized crosswalk facilities," *Transportation*, vol. 29, no. 2, pp. 169–192, 2002.
- [13] S. M. Siram, V. Sonaike, O. B. Bolorunduro et al., "Does the pattern of injury in elderly pedestrian trauma mirror that of the younger pedestrian?" *Journal of Surgical Research*, vol. 167, no. 1, pp. 14–18, 2011.
- [14] J. W. Zhou, H. Kuang, M. R. Liu, and L. J. Kong, "Paired behavior effect on pedestrian evacuation dynamics," *Acta Physica Sinica*, vol. 58, no. 5, pp. 3001–3007, 2009.
- [15] W. H. Wang, Q. Cao, K. Ikeuchi, and H. Bubb, "Reliability and safety analysis methodology for identification of drivers' erroneous actions," *International Journal of Automotive Technology*, vol. 11, no. 6, pp. 873–881, 2010.
- [16] R. L. Hughes, "A continuum theory for the flow of pedestrians," *Transportation Research Part B*, vol. 36, no. 6, pp. 507–535, 2002.
- [17] K. Ando, H. Ota, and T. Oki, "Forecasting the flow of people," *Railway Research Review*, vol. 45, no. 8, pp. 8–14, 1998.
- [18] P. A. Thompson and E. W. Marchant, "A computer model for the evacuation of large building populations," *Fire Safety Journal*, vol. 24, no. 2, pp. 131–148, 1995.
- [19] R. L. Hughes, "The flow of large crowds of pedestrians," *Mathematics and Computers in Simulation*, vol. 53, no. 4–6, pp. 367–370, 2000.
- [20] B. D. Hankin and R. A. Wright, "Passenger flow in subways," *Operational Research Quarterly*, vol. 9, no. 2, pp. 81–88, 1958.
- [21] C. Y. Cheung and W. H. K. Lam, "Pedestrian route choices between escalator and stairway in MTR stations," *Journal of Transportation Engineering*, vol. 124, no. 3, pp. 277–285, 1998.
- [22] Y. Tanaboriboon, Sim Siang Hwa, and Chin Hoong Chor, "Pedestrian characteristics study in Singapore," *Journal of Transportation Engineering, ASCE*, vol. 112, no. 3, pp. 229–235, 1986.
- [23] National Research Council (NRC), *Highway Capacity Manual*, National Research Council (NRC), Washington, DC, USA, 2010.
- [24] X. H. Cui, Q. Li, J. Chen, and C. X. Chen, "Study on MA-based model of occupant evacuation in public facility," *Journal of System Simulation*, vol. 20, no. 4, pp. 1006–1023, 2008.

- [25] J. Wu, J. Feng, and J. Ye, "Capacity analysis of magnetic card and contactless IC card automatic fare gate in railway stations," *Journal of Tongji University*, vol. 38, no. 1, pp. 85–91, 2010.
- [26] A. El Kacimi and O. Laghrouche, "Numerical analysis of two plane wave finite element schemes based on the partition of unity method for elastic wave scattering," *Computers and Structures*, vol. 88, no. 23-24, pp. 1492–1497, 2010.
- [27] L. Sun, J. Rong, and L. Yao, "Measuring transfer efficiency of urban public transportation terminals by data envelopment analysis," *Journal of Urban Planning and Development*, vol. 136, no. 4, pp. 314–319, 2010.
- [28] H. F. Jia, L. L. Yang, and M. Tang, "Pedestrian flow characteristics analysis and model parameter calibration in comprehensive transport terminal," *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 5, pp. 117–123, 2009.