Transportation Research Part B 45 (2011) 724–742

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Intermodal hub-and-spoke network design: Incorporating multiple stakeholders and multi-type containers

Qiang Meng *, Xinchang Wang

Department of Civil Engineering, National University of Singapore, Singapore 117576, Singapore

ARTICLE INFO

Article history:
Received 29 December 2009
Received in revised form 6 November 2010
Accepted 6 November 2010

Keywords:
Intermodal transportation
Hub-and-spoke network design
User equilibrium
Multi-type containers
Genetic algorithm

ABSTRACT

This paper develops a mathematical program with equilibrium constraints (MPEC) model for the intermodal hub-and-spoke network design (IHSND) problem with multiple stakeholders and multi-type containers. The model incorporates a parametric variational inequality (VI) that formulates the user equilibrium (UE) behavior of intermodal operators in route choice for any given network design decision of the network planner. The model also uses a cost function that is capable of reflecting the transition from scale economies to scale diseconomies in distinct flow regimes for carriers or hub operators, and a disutility function integrating actual transportation charges and congestion impacts for intermodal operators. To solve the MPEC model, a hybrid genetic algorithm (HGA) embedded with a diagonalization method for solving the parametric VI is proposed. Finally, the comparative analysis of the HGA and an exhaustive enumeration algorithm indicates a good performance of the HGA in terms of computational time and solution quality. The HGA is also applied to solve a large-scale problem to show the applicability of the proposed model and algorithm.

1. Introduction

Intermodal freight transportation involving multiple modes of transport, such as rail, road and sea, and multi-type containers provides an economical solution for the long-haul international/transcontinental cargo delivery. Multi-type containers refer to the containers serving various specific purposes, having different functions or sizes, and requiring different handling operations, costs and times. Examples include 20-foot equivalent units (TEUs), refrigerated TEUs, and tank containers. As suggested by Crainic and Kim (2007), an intermodal freight transportation system can be essentially formulated as a hub-and-spoke network. A typical container delivery in the network can be achieved in the following steps: containers are initially collected from shippers by carriers using drayage, transferred via a sequence of transfer terminals such as ports, rail–road terminals and border crossing terminals between adjacent countries, and eventually distributed to destinations. Transfer terminals are referred as to hubs, at which containers are sorted and switched between various modes. The non-hub nodes allocated to hubs to transship containers are referred as to spoke nodes. An intermodal hub-and-spoke network may reap the advantages of economies of scale and scope exhibited in transport of multi-type containers between hubs, so as to provide carriers with a cost-effective transportation solution. In addition, as hubs normally act as the main engines of economic development of the regions served by them, the market areas of the hubs could be expanded by organizing an effective hub-and-spoke network. The intermodal hub-and-spoke network design (IHSND) for multi-type container transportation is therefore of importance for both carriers and government authorities. However, the IHSND problem is fundamentally

* Corresponding author. Tel.: +65 6516 5494; fax: +65 6779 1635.
E-mail address: cvemq@nus.edu.sg (Q. Meng).

0191-2615/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trb.2010.11.002

distinguished from the conventional hub-and-spoke network design (HSND) problem and its variants (O'Kelly, 1987; Campbell, 1994), due to its unique characteristics elaborated below.

In addition to locating hubs, as addressed in the conventional HSND problem, the IHSND problem needs to identify optimal mode-change transshipment lines at hubs. A transshipment line represents a collection of infrastructure facilities such as yard cranes, vehicles and straddle carriers, needed to accomplish container mode changes. Generally, containers will be handled due to mode changes in transshipment, and as a consequence, additional transshipment cost and time will be incurred. It must be noted that a container handling procedure may be required even if the transport mode does not change. For instance, containers are often unloaded, inspected and reloaded for customs clearance at a border crossing terminal without change of transport mode. Such a process is also regarded as one type of mode change.

Furthermore, the IHSND problem involves various kinds of stakeholders: the network planner, carriers, hub operators and intermodal operators. Carriers represent freight transportation service providers such as railway transport companies and liner shipping companies. Hub operators offer container transshipment services at transfer terminals, such as port operators. The planner represents a decision-making entity which makes network design decisions based on its own concern. The network planner may represent a local or national government agency, an association of several countries or an intermodal organization, or a private sector firm. It is worthwhile to notice that, when a private sector firm behaves as the network planner to design a cost-effective intermodal network for its own use, the firm may provide both transportation services and hub operations, i.e., all carriers and hub operators are the service providers belonging to the sector. Intermodal operators can be considered as users of intermodal infrastructure facilities and services, who make route selection decisions for a shipment through the whole intermodal network (Macharis and Bontekoning, 2004). Intermodal operators intend to transport multi-type containers between each origin–designation (O/D) pair by purchasing transportation and transshipment services from carriers and hub operators on the behalf of shippers, and coordinate and organize the transport process. An intermodal operator may represent a shipper himself, a third-party logistics company or an intermediary broker.

These stakeholders exhibit distinct decision behaviors in the network design. For instance, the United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) initiated Trans-Asian Railway and Asian Highway projects with the aims to develop an accessible and efficient transport network linking the countries in Asia and Pacific region and to facilitate the development of Eurasian container trade and economic exchanges (UNESCAP, 1999, 2003). UNESCAP can be considered as a network planner. Given a limited budget, the network planner intends to design and build an optimal intermodal hub-and-spoke network for multi-country container transportation with the aim to minimize the total transportation cost of carriers and operational cost of hub operators. To design the network, the planner needs to determine hub locations, enhance or establish transshipment lines, and design physical links based on container flow pattern in the network. The flow pattern, on the other hand, results from the route choices of intermodal operators given a feasible network design decision. As addressed by Guélat et al. (1990) in the freight flow assignment problem, an intermodal operator's route choice can be reasonably assumed to follow the user equilibrium (UE) principle (Wardrop, 1952), in accordance with transportation disutilities of available intermodal routes. The disutility of an intermodal route actually represents the aversion of an intermodal operator toward the route. It can be expressed by the sum of the actual rate and transportation time multiplied by value of time (VOT) along the route. The rate is the cost incurred in transporting and transshipping one container along the route. Practically, the actual rate (expenditure) is determined by carriers or hub operators and charged to an intermodal operator according to specific rate tables. This amount of money can be regarded as the reward for the carriers or hub operators providing transportation or transshipment services. Transportation time is another key factor influencing the route choices of intermodal operators and is incorporated into disutility functions using VOT.

1.1. Previous related studies

The conventional HSND problem has been extensively investigated in past studies (Alumur and Kara, 2008). It was originally formulated to be a quadratic integer programming model by O'Kelly (1987) and later referred to as a p-hub median problem by Campbell (1994). The p-hub median problem aims to select p hub facilities from a set of available candidates, to allocate spoke nodes to hubs and to simultaneously route freight flows. The HSND problem has been attracting numerous researchers and has been formulated as various mathematical models, including path-based mixed-integer linear programming models with four-dimensional variables (Campbell, 1996; Skorin-Kapov et al., 1996; Klincewicz, 1998; Podnar et al., 2002) and origin-based mixed-integer linear programming models with three-dimensional variables (Ernst and Krishnamoorthy, 1998; Ebery et al., 2000; Boland et al., 2004). Early hub location models were developed under the following four basic assumptions: (i) no direct links are allowed to connect spoke nodes, (ii) hubs are fully interconnected, (iii) scale economies are exhibited in transportation between hub pairs, and (iv) each path traverses at most two hubs to connect each origin–destination (O/D) pair. The above-mentioned assumptions, however, are sometimes unrealistic for intermodal transportation network design. For example, intermodal transportation is a widely used service for international or transcontinental container delivery across different countries. Some hubs located in a country may only be connected to those situated in the same country, which does not comply with assumption (ii). As opposed to assumption (iv), an intermodal route may traverse more than two hubs with mode changes.

Some studies have been conducted to relax these assumptions (Klincewicz, 1998; Nickel et al., 2001). Campbell et al. (2005a, 2005b) proposed a hub arc location model which does not impose restrictions on hub connections by incorporating bridge arcs. Yoon and Current (2008) developed a mixed integer programming model embedding a multi-commodity flow model to solve the hub location problem with no restrictions on network topology. They also considered direct links between spoke nodes. Alumur et al. (2009) provided a unified modeling method for single-allocation hub location problems with incomplete hub networks. Contreras et al. (2010) addressed a tree of hubs location problem with the particularity that all hubs are connected by means of a non-directed tree, which has the potential applications where the cost of establishing hub–hub links is significantly high. These studies provide inspirations on formulating a more generalized hub location problem; however, they do not deal with locating transshipment lines at hubs, which are indispensable infrastructure facilities for intermodal transportation networks. Meanwhile, these HSND studies do not involve the interactive decision process among different stakeholders, which is regarded as another important characteristic of intermodal freight transportation.

Two typical types of unit transportation cost functions have been widely employed to reflect scale economies in the HSND problem. Balakrishnan and Graves (1989) employed a piecewise-linear decreasing function to compute unit transportation cost with respect to freight flow. O'Kelly and Bryan (1998) and Horner and O'Kelly (2001) suggested a nonlinear decreasing unit cost function for hub-and-spoke network design. However, this decreasing property is not constantly true in freight transportation. Norman (1979) pointed out that a flexible unit transportation cost function will exhibit an increasing trend when the freight flow exceeds a threshold, as labor, material and machinery costs will sharply increase to cater with the additional demand that is beyond the safe operating capacity of a carrier. Friesz and Holguin-Veras (2005) further emphasized that the unit freight transportation cost function should have a "U" shape that can reflect the transition from economies of scale to diseconomies of scale in different flow regimes. Additionally, past cost functions were proposed for characterizing the cost structure in transport of uni-type commodities and unable to reflect scope economies that reflect cost savings resulting from simultaneously transporting multi-type containers. Hence, we have to seek a more realistic transportation cost function which can depict the cost structure of multi-type container transportation for the IHSND problem.

As compared to the traditional HSND problem, the IHSND problem has received limited attention, presumably because intermodal freight transportation is a recently emerging research field and has not been so far explored in detail (Bontekoning et al., 2004). A few studies have been carried out for addressing the problem. Arnold et al. (2004) formulated an integer programming model for the rail–road hub location problem. Racunica and Wynter (2005) gave an optimal hub location model aiming to increase the market share of rail mode in a hub-and-spoke network. Limbourg and Jourquin (2007) located the best potential rail–road hubs in an intermodal network according to the traffic flow distribution and its geographical dispersion over the network. Although these models extend the conventional HSND problem to involve partial characteristics of intermodal transportation operations, they neglect the interactions among different stakeholders and cargo transfers between multiple modes besides rail and road. Moreover, these models are inadequate to formulate the multi-type container transportation problem, which is widely seen in practical intermodal freight operations.

Loureiro and Balston (1996) proposed a bilevel multi-commodity network design model for determining investment policies for intercity freight transportation networks by assuming that the behavior of shippers (network users) in route choice follows the user equilibrium (UE) principle. Yamada et al. (2009) established a bilevel programming model for freight transportation network design, in which the lower-level problem is a multimodal multi-class user equilibrium traffic assignment problem for network users. Each of these two models is a straightforward extension of the discrete network design problem with user equilibrium constraints for urban road networks (Yang and Bell, 1998; Meng et al., 2001). However, these two models are not concerned with locating hubs and designing transshipment lines. The unit transportation cost functions used in the two models are unable to reflect scale economies or to describe the cost structure of multi-type container transportation.

1.2. Contributions and objectives

The IHSND problem aims to design a given intermodal network as an optimal hub-and-spoke network for multi-type container transportation with multiple stakeholders: the network planner, carriers, hub operators and intermodal operators. To formulate the problem, we first define two correlated networks: physical network and operational network. The physical network is a collection of network elements including nodes, links and transshipment lines. The operational network, which is derived based on the physical network, explicitly reflects the container transfer processes at hubs and will be used as the basis for intermodal operators to select intermodal routes. The planner intends to design the former one, and based on the latter one, intermodal operators will route multi-type containers via available intermodal routes following the Wardropian UE principle. The container transfer processes at hubs in the operational network are modeled as a series of transfers that traverse through specific mode-change transshipment lines.

Next, a cost function that is capable of reflecting the transition from scale economies to scale diseconomies in distinct flow regimes and the relation between multi-type container flows is suggested for carriers and hub operators. A disutility function that integrates actual transportation rates and congestion impacts is proposed for intermodal operators. By using these two functions, a mathematical program with equilibrium constraints (MPEC) model is developed to formulate the IHSND problem with the aim of minimizing the total transportation cost of carriers and operational cost of hub operators. The model incorporates a parametric variational inequality (VI) that reflects the UE behavior of intermodal operators in route choice given any network design decision. To solve the MPEC model, a hybrid genetic algorithm (HGA) with an embedded diagonalization method for solving the asymmetric freight flow assignment problem is developed. The HGA and an exhaustive enumeration algorithm (EEA) are compared to assess the effectiveness of the proposed solution algorithm. The HGA is also applied to solve a large-scale IHSND problem.

The IHSND problem has the following unique characteristics, as compared to the conventional HSND problem: (i) an intermodal route may traverse more than two hubs and all hubs are less than fully interconnected, (ii) a flexible cost function reflecting the transition from economies of scale to diseconomies of scale and the relation between multi-type container flows is suggested for carriers and hub operators, (iii) transshipment lines need to be designed and container transfer processes at hubs are adequately modeled as transfers, and (iv) the network design is simultaneously determined by the interactive decisions of multiple stakeholders. By considering these four features, we take the first initiative to investigate the IHSND problem from a more realistic point of view, which contributes a new perspective to hub-and-spoke network design theories.

The remainder of this paper is organized as follows. Section 2 represents a given intermodal network as the physical and operational networks. Section 3 defines the IHSND problem, analyzes cost and disutility functions, and proposes an intermodal route choice model for intermodal operators. Section 4 formulates the IHSND problem as a MPEC model. Section 5 proposes a hybrid GA to solve the MPEC model. In Section 6, numerical examples are given to assess the proposed model and solution algorithm. Conclusions are drawn in Section 7.

2. Network representation

Consider an existing intermodal freight transportation network, comprising the set of spoke nodes denoted by $\mathcal{N}_0$, the set of hubs denoted by $\mathcal{N}_1$, the set of all direct links connecting two spoke nodes denoted by $\mathcal{A}_0$, and the set of hub links and spoke–hub links denoted by $\mathcal{A}_1$. A planner attempts to design the network as an intermodal hub-and-spoke network by locating new hubs from a set of candidate hubs, $\mathcal{N}_2$, and adding new links selected from a set of candidate links, $\mathcal{A}_2$. To facilitate the model formulation, we first represent the original network as physical and operational networks.

2.1. Physical network

Let directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ be a collection of possible physical network elements, where $\mathcal{N} = \mathcal{N}_0 \cup \mathcal{H}$ is the set of nodes with the set of hubs $\mathcal{H} = \mathcal{N}_1 \cup \mathcal{N}_2$ and the set of links $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1 \cup \mathcal{A}_2$. The sets of transportation modes and container types, are represented by $\mathcal{M}$ and $\mathcal{P}$, respectively. Let $\mathcal{U}$ be the set of carriers. Each link $a \in \mathcal{A}$ is represented by a triplet $(\tilde{h}_a, \tilde{t}_a, m_a)$, where $\tilde{h}_a, \tilde{t}_a \in \mathcal{N}$ are the head and tail of link a, respectively, and $m_a \in \mathcal{M}$ is the transportation mode on the link. Thus,

$$\mathcal{A}_0 \subseteq \mathcal{N}_0 \times \mathcal{N}_0 \times \mathcal{M} \tag{1}$$

$$\mathcal{A}_1 \subseteq (\mathcal{N}_0 \cup \mathcal{N}_1) \times \mathcal{N}_1 \times \mathcal{M} \tag{2}$$

$$\mathcal{A}_2 \subseteq \mathcal{N} \times \mathcal{N}_2 \times \mathcal{M} \tag{3}$$

Let Γ_a be the current traffic capacity of link a, the maximum number of standard vehicles that can pass though this link during some time unit such as a day or year. The traffic volume in terms of standard vehicles over a link can be practically estimated by converting various types of vehicles to a standardized vehicle type associated with the link mode. For example, we can convert buses and trucks to passenger car units for road mode, transform passenger and freight trains to rolling stocks for rail mode, and convert bulk, oil and container vessels to standard containerships for sea mode. This principle is consistent with the "passenger car equivalent" (p.c.e) concept that is now widely used and accepted by researchers and engineers (Daganzo, 1983; TRB, 2000). Note that current traffic capacity $\Gamma_a = 0$ for each link $a \in \mathcal{A}_2$ because it will be built in the future.

Each hub $h \in \mathcal{H}$ may possess one or several distinct directed mode-change transshipment lines that switch containers between different modes. Let $\mathcal{B}_h^1$ and $\mathcal{B}_h^2$ be the sets of existing and potential transshipment lines at hub $h \in \mathcal{H}$, respectively, and $\mathcal{B}_h = \mathcal{B}_h^1 \cup \mathcal{B}_h^2$ be the set of all transshipment lines at the hub. A transshipment line $b \in \mathcal{B}_h$ that can transfer containers from modes $m_1 \in \mathcal{M}$ to $m_2 \in \mathcal{M}$ is represented by a duplet $b = (m_1, m_2)$, and it follows that $\mathcal{B}_h \subseteq \Phi = \mathcal{M} \times \mathcal{M}$. All the containers to be transferred from modes m_1 to m_2 will traverse through the specific transshipment line b. Let $\mathcal{B}_1 = \cup_{h \in \mathcal{H}} \mathcal{B}_h^1,\ \mathcal{B}_2 = \cup_{h \in \mathcal{H}} \mathcal{B}_h^2$ and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ be the sets of existing, potential and all possible transshipment lines in graph $\mathcal{G}$, respectively.

In addition, each existing transshipment line $b \in \mathcal{B}_1$ has a current container transshipment capacity Γ_b, the maximum number of standardized containers allowed to be transferred via the transshipment line during some time unit. The current container transshipment capacity $\Gamma_b = 0$ for each $b \in \mathcal{B}_2$. The container volume in terms of standard containers loaded on any transshipment line $b \in \mathcal{B}$ can be calculated by converting flows of multi-type containers over the transshipment line to that of a 20-foot equivalent unit (TEU).

Let set $\mathcal{O} \subseteq \mathcal{N}$ consist of all origin nodes and $\mathcal{D} \subseteq \mathcal{N}$ comprise all destination nodes. Let (o, d) denote each O/D pair, where $o \in \mathcal{O},\ d \in \mathcal{D}$, between which intermodal operators route multi-type container flows. It is worthwhile to notice that the optimal container size (the containers with the same size are considered to belong to the same type) determined by shippers may impose a non-neglectable impact on transportation costs, disutilities and route choice (Abdelwahab, 1998; Holguín-Veras et al., 2009). We assume that the optimal container size and the transport demand of each container type have been predetermined by shippers as inputs of the IHSND problem. Let T_{od}^p be the transport demand of container $p \in \mathcal{P}$ between (o, d).

In addition to the above notations, we use letter n with a subscripted letter representing a set to denote cardinality of the set throughput this paper, e.g. n_P represents the number of container types in set $\mathcal{P}$.

2.2. Operational network

Based on physical network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, an operational network can be generated, where each carrier link or transshipment line is assumed to be operated by an individual carrier or hub operator. Let directed graph $G = (N, A, T)$ be the operational network derived from $\mathcal{G}$, where the set of nodes $N = \mathcal{N}$, the set of carrier links $A \subseteq \mathcal{A} \times \mathcal{U}$ and the set of transfers $T \subseteq A \times A$.

Let $\mathcal{U}_a \subseteq \mathcal{U}$ be the set of carriers on link $a \in \mathcal{A}$ and the link is further represented by a set of parallel carrier links in the operational network, denoted by A_a. Hence, carrier link $\bar{a} \in A_a$ is specified by the quadruplet $(\tilde{h}_{\bar{a}}, \tilde{t}_{\bar{a}}, m_{\bar{a}}, u_{\bar{a}})$, where $\tilde{h}_{\bar{a}}, \tilde{t}_{\bar{a}}$, and $m_{\bar{a}}$ are the head, tail and mode of the carrier link, respectively, with $\tilde{h}_{\bar{a}} = \tilde{h}_a, \tilde{t}_{\bar{a}} = \tilde{t}_a, m_{\bar{a}} = m_a$, and $u_{\bar{a}} \in \mathcal{U}_a$ represents the carrier over $\bar{a}$. We thus have $A = \bigcup_{a \in \mathcal{A}} A_a$. Each carrier link $\bar{a}$ has a transportation capacity $\Gamma^p_{\bar{a}}$ defined as the maximum number of containers of type $p \in \mathcal{P}$ that carrier $u_{\bar{a}}$ can deal with during some time unit as per its handling productivity.

At each hub $h \in \mathcal{H}$, the container transfer process from carrier links $\bar{a}_1 \in A$ to $\bar{a}_2 \in A$, can be represented by transfer $\bar{l} = (\bar{a}_1, \bar{a}_2)$ via transshipment line $b \in \mathcal{B}$. Cost and time will be correspondingly incurred on the transfer due to the utilization of the transshipment line. Let T_h be the set of all possible transfers at hub h and $T = \cup_{h \in H} T_h$ be the set of transfers in graph G.

Fig. 1 shows the process of deriving an operational network from a physical network. $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N} = \{1, 2, 3, 4\}$ and $\mathcal{A} = \{a_i | i = 1, 2, 3, 4\}$. We consider $\mathcal{M} = \{\text{rail}, \text{road}, \text{sea}\}, \mathcal{U} = \{u_i | i = 1, 2, 3, 4\}, \mathcal{H} = \{3\}$, and $\mathcal{B} = \{\text{road} - \text{road}, \text{road}- \text{rail}, \text{sea} - \text{road}, \text{sea} - \text{rail}\}$. Based on the physical network, an operational network $G = (N, A, T)$ can be generated, where $N = \{1, 2, 3, 4\}$ and $A = \{\bar{a}_i | i = 1, 2, \ldots, 6\}$. As two carriers u_1 and u_2 are operating on links a_1 or a_2, each of the physical links can be represented by two parallel carrier links i.e. $A_{a_1} = \{\bar{a}_1, \bar{a}_2\}$ and $A_{a_2} = \{\bar{a}_4, \bar{a}_5\}$. The transfers for hub $h = 3 \in \mathcal{H}$ are shown in the figure. Each transfer can only traverse a unique transshipment line. To reflect the incidence relation between a transfer and a transshipment line, let T^b_h be the set of transfers traversing through transshipment line $b \in \mathcal{B}_h$, where $h \in \mathcal{H}$.

$$T^b_h = \{\bar{l} = (\bar{a}_1, \bar{a}_2) \in T_h | m_{\bar{a}_1} = m_1, m_{\bar{a}_2} = m_2\}, \quad \forall b = (m_1, m_2) \in \mathcal{B}_h, \quad h \in \mathcal{H} \tag{4}$$

Let R^p_{od} be the set of all available intermodal routes for container type p between (o, d), where $o \in \mathcal{O}, d \in \mathcal{D}$. An intermodal route $r \in R^p_{od}$ may traverse a sequence of carrier links and transfers in G. Let $\delta^{\bar{a}p}_{odr} = 1$ if path r traverses carrier link $\bar{a} \in A$; zero otherwise, $\delta^{\bar{l}p}_{odr} = 1$ if path r traverses transfer $\bar{l} \in T$; zero otherwise. Correspondingly, route r passes through a sequence of physical links and transshipment lines in network $\mathcal{G}$. Let $\mathcal{A}_r$ and $\mathcal{B}_r$ be the sets of physical links and transshipment lines traversed by route r, respectively, namely,

$$\mathcal{A}_r = \left\{ a \in \mathcal{A} | \sum_{\bar{a} \in A_a} \delta^{\bar{a}p}_{odr} \geq 1 \right\}, \quad \forall r \in R^p_{od}, \quad o \in \mathcal{O}, \quad d \in \mathcal{D}, \quad p \in \mathcal{P} \tag{5}$$

$$\mathcal{B}_r = \left\{ b \in \mathcal{B}_h, h \in \mathcal{H} | \sum_{\bar{l} \in T^b_h} \delta^{\bar{l}p}_{odr} \geq 1 \right\}, \quad \forall r \in R^p_{od}, \quad o \in \mathcal{O}, \quad d \in \mathcal{D}, \quad p \in \mathcal{P} \tag{6}$$

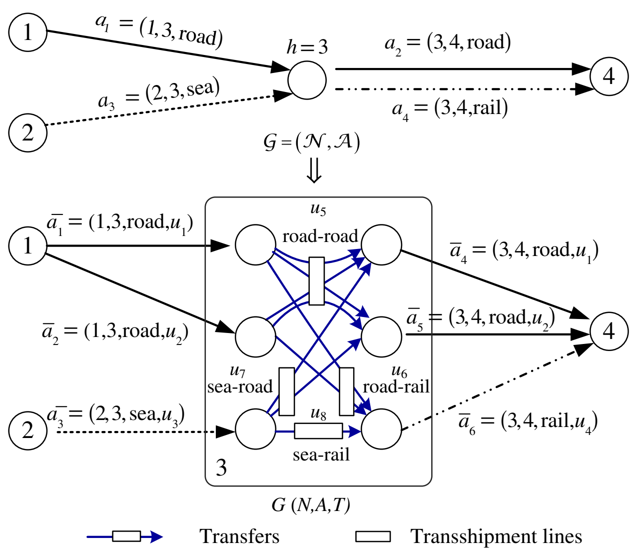

Fig. 1. An example for physical and operational network representations.

3. Problem statement

3.1. Decision variables

In addressing the IHSND problem, the following assumptions are made throughout the paper: (i) spoke nodes cannot serve as transshipment points and (ii) no new direct links between spoke nodes will be established.

To design a given network as an intermodal hub-and-spoke network, the planner needs to make the following decisions: (i) expanding the current traffic capacities of existing links in set $\mathcal{A}_1$ by adding more lanes or conducting road maintenance or rehabilitation, (ii) establishing new links chosen from set $\mathcal{A}_2$ with specified traffic capacities, (iii) enhancing existing transshipment lines in set $\mathcal{B}_1$ by increasing their current container transshipment capacities, (iv) selecting new hubs from set $\mathcal{N}_2$, and (v) building new transshipment lines with specified capacities for the selected hubs. These decisions can be expressed by variables:

$$x_a = \begin{cases} 1, & \text{if traffic capacity of link } a \text{ is expanded} \\ 0, & \text{otherwise} \end{cases}, \forall a \in \mathcal{A}_1 \tag{7}$$

$$x_{\hat{a}} = \begin{cases} 1, & \text{if link } \hat{a} \text{ is established} \\ 0, & \text{otherwise} \end{cases}, \forall \hat{a} \in \mathcal{A}_2 \tag{8}$$

$$y_b = \begin{cases} 1, & \text{if capacity of transshipment line } b \text{ is enhanced} \\ 0, & \text{otherwise} \end{cases}, \forall b \in \mathcal{B}_1 \tag{9}$$

$$y_{\hat{b}} = \begin{cases} 1, & \text{if transshipment line } \hat{b} \text{ is established} \\ 0, & \text{otherwise} \end{cases}, \forall \hat{b} \in \mathcal{B}_2 \tag{10}$$

$$z_h = \begin{cases} 1, & \text{if hub } h \text{ is selected} \\ 0, & \text{otherwise} \end{cases}, \forall h \in \mathcal{N}_2 \tag{11}$$

In addition, let $x_a = 0$, $z_h = 1$ and $z_n = 0$ for any existing spoke link $a \in \mathcal{A}_0$, existing hub $h \in \mathcal{N}_1$ and spoke node $n \in \mathcal{N}_0$, respectively. These decision variables are grouped into three vectors of decision variables: $\mathbf{x} = (x_a, a \in \mathcal{A})$, $\mathbf{y} = (y_b, b \in \mathcal{B})$ and $\mathbf{z} = (z_n, n \in \mathcal{N})$.

Various investment actions are coupled with network design solutions. No investments are allocated for existing spoke links. It costs B_a to enhance the traffic capacity of link $a \in \mathcal{A}_1$ from Γ_a to Θ_a, and $B_{\hat{a}}$ to build a new link $\hat{a} \in \mathcal{A}_2$ with capacity $\Theta_{\hat{a}}$. Cost F_b is required to enhance the capacity of an existing transshipment line $b \in \mathcal{B}_1$ from Γ_b to Θ_b. When a new transshipment line $\hat{b} \in \mathcal{B}_2$ is established with capacity $\Theta_{\hat{b}}$, cost $F_{\hat{b}}$ will be incurred. Once a hub $h \in \mathcal{N}_2$ is selected, at least one transshipment line should be established at the hub; it will degenerate to a spoke node otherwise.

3.2. Intermodal route choice model for intermodal operators: UE principle

Given a feasible network design solution $\mathbf{s} = (\mathbf{x}, \mathbf{y}, \mathbf{z})$, intermodal operators will route multi-type containers over the operational network derived from $\mathbf{s}$, and the network flow distribution will result from their behavior in intermodal route choice. According to the UE principle, no intermodal operator could unilaterally decrease his/her transportation disutility by shifting to choose another route when he/she is faced with route choice for transporting one container of type $p \in P$ between each O/D pair in G.

The disutility along an intermodal route is the sum of disutilities along the carrier links and transfers traversed by the route. The disutility of a carrier link or transfer perceived by an intermodal operator is composed of two portions: the actual rate charged by the carrier on the carrier link or hub operator providing the transfer service according to a particular freight rate table, and the monetary value of transportation or transfer time that equals the actual container handling time multiplied with value of time (VOT). In reality, the second portion reflects the impact of congestion on the intermodal operator's route choice. Congestion phenomena can be observed on physical links and transshipment lines due to the limit of capacity.

Over each carrier link $\bar{a} \in A_a$ where $a \in \mathcal{A}$, transportation time for one container of type p is associated with multi-type container flows of the carriers who share the same physical link with $\bar{a}$. It is also associated with container flows loaded on other carrier links and transfers, since congestion at hubs or physical links may result in traffic delays in the whole network. Additionally, the transportation or transfer time is associated with network design solution $\mathbf{s}$ which determines the capacities of physical links and transshipment lines. The transportation or transfer time can thus be considered as a function of container flow vector $\mathbf{v}$ and network design solution $\mathbf{s}$, where

$$\mathbf{v} = (v^p_{\bar{a}}, \bar{a} \in A; v^p_{\bar{l}}, \bar{l} \in T; p \in \mathcal{P}) \tag{12}$$

in which $v^p_{\bar{a}}$ and $v^p_{\bar{l}}$ are the flows of container $p \in \mathcal{P}$ loaded on carrier link $\bar{a} \in A$ and transfer $\bar{l} \in T$, respectively.

Let $t^p_{\bar{a}}(\mathbf{v},\mathbf{s})$ and $t^p_{\bar{l}}(\mathbf{v},\mathbf{s})$ denote the time incurred on carrier link $\bar{a}$ and transfer $\bar{l}$, respectively. The disutilities of $\bar{a}$ and $\bar{l}$ perceived by intermodal operators can be expressed by,

$$g^p_{\bar{a}}(\mathbf{v},\mathbf{s}) = r^p_{\bar{a}} + \phi^p_{\bar{a}} t^p_{\bar{a}}(\mathbf{v},\mathbf{s}) \tag{13}$$
$$g^p_{\bar{l}}(\mathbf{v},\mathbf{s}) = r^p_{\bar{l}} + \phi^p_{\bar{l}} t^p_{\bar{l}}(\mathbf{v},\mathbf{s}) \tag{14}$$

where $r^p_{\bar{a}}$ and $r^p_{\bar{l}}$ are the actual freight rates charged by the carriers on physical link $\bar{a}$ and hub operators operating the transshipment line traversed by $\bar{l}$, respectively. $\phi^p_{\bar{a}}$ and $\phi^p_{\bar{l}}$ are the VOTs perceived by intermodal operators for handling containers of type p on $\bar{a}$ and $\bar{l}$, respectively.

One should calibrate the functional forms of transportation time $t^p_{\bar{a}}(\mathbf{v},\mathbf{s})$ and $t^p_{\bar{l}}(\mathbf{v},\mathbf{s})$ by using historical data. The BPR-form time function developed by the US Bureau of Public Roads (BPR) has been employed in many freight transportation studies (Fernádez et al., 2003; Yamada et al., 2009), since it can reflect the congestion impact in a steady-state transportation system by involving the volume/capacity ratio. The function has also been used in many countries without much effort to calibrate the parameters (Suh et al., 1990). We assume that the transportation or handling time function in multi-type container transportation has a BRP form. Let β^p_a be the factor that converts container type p into a standard container vehicle unit on physical link a. Let κ_a be factor that converts a standard container vehicle into a standardized vehicle equivalent on link a. Assume that the transportation time for transporting one container of type p over carrier link $\bar{a} \in A_a$ where $a \in \mathcal{A}$ will be affected by both container and non-container vehicles that share the same physical link a. Let η_a denote the proportion of the flow of standard vehicles brought about by container vehicles to the total traffic volume of standard vehicles on physical link a. The BPR-form transportation time function can be expressed by,

$$t^p_{\bar{a}}(\mathbf{v},\mathbf{s}) = t^{p0}_{\bar{a}} + \alpha^p_{\bar{a}} t^{p0}_{\bar{a}} \left(\frac{\kappa_a \beta^p_a v^p_{\bar{a}} + k^p_{\bar{a}}(w_a - \kappa_a \beta^p_a v^p_{\bar{a}})}{\Gamma_a(1-x_a) + \Theta_a x_a} \right)^{\omega^p_{\bar{a}}} \tag{15}$$

where $t^{p0}_{\bar{a}}$ is the free-flow time for transporting one container of type p over $\bar{a}$, and $k^p_{\bar{a}}$ is the asymmetry factor that reflects the interaction among multi-type container flows on physical link a. $\alpha^p_{\bar{a}}$ and $\omega^p_{\bar{a}}$ are two parameters to be calibrated. The traffic volume in terms of standard vehicles loaded over link a is expressed by,

$$w_a = \kappa_a \sum_{\bar{a}\in A_a} \sum_{p\in\mathcal{P}} \beta^p_a v^p_{\bar{a}} / \eta_a \tag{16}$$

We assume that a transfer $\bar{l} \in T^b_h$ where $b \in \mathcal{B}_h$, $h \in \mathcal{H}$ is only correlated with the transfers that traverse the same transshipment line with $\bar{l}$. Let β^p_b be the factor converting container type p to one TEU over b. The time for handling one container of type p over $\bar{l}$ can be thus evaluated by,

$$t^p_{\bar{l}}(\mathbf{v},\mathbf{s}) = t^{p0}_{\bar{l}} + \alpha^p_{\bar{l}} t^{p0}_{\bar{l}} \left(\frac{\beta^p_b v^p_{\bar{l}} + k^p_{\bar{l}}(w_b - \beta^p_b v^p_{\bar{l}})}{(1-y_b)\Gamma_b + y_b \Theta_b} \right)^{\omega^p_{\bar{l}}} \tag{17}$$

where $t^{p0}_{\bar{l}}$ is the free-flow handling time per container p and $k^p_{\bar{l}}$ is a factor reflecting the interaction among all transfers via the transshipment line b. The standard container flow in terms of TEUs transferred via b is thus calculated by,

$$w_b = \sum_{\bar{l}\in T^b_h} \sum_{p\in\mathcal{P}} \beta^p_b v^p_{\bar{l}} \tag{18}$$

and $\alpha^p_{\bar{l}}$ and $\omega^p_{\bar{l}}$ are two parameters to be calibrated.

Let $\mathbf{v}_{\bar{a}} = (v^p_{\bar{a}}, p \in \mathcal{P})$ and $\mathbf{v}_{\bar{l}} = (v^p_{\bar{l}}, p \in \mathcal{P})$ be the vectors of multi-type container flows loaded on $\bar{a}$ and $\bar{l}$, respectively. The flow of container type p on path $r \in R^p_{od}$ is denoted by f^p_{odr}. Vector disutility functions for $\bar{a}$ and $\bar{l}$ are represented by

$$\mathbf{g}_{\bar{a}}(\mathbf{v},\mathbf{s}) = (g^p_{\bar{a}}(\mathbf{v},\mathbf{s}), p \in \mathcal{P}) \tag{19}$$
$$\mathbf{g}_{\bar{l}}(\mathbf{v},\mathbf{s}) = (g^p_{\bar{l}}(\mathbf{v},\mathbf{s}), p \in \mathcal{P}) \tag{20}$$

The UE principle implies that the multi-type container flow patterns $\mathbf{v}$ over G can be obtained by solving the parametric variational inequality (VI),

$$\sum_{\bar{a}\in A} \mathbf{g}_{\bar{a}}(\mathbf{v},\mathbf{s})(\hat{\mathbf{v}}_{\bar{a}} - \mathbf{v}_{\bar{a}}) + \sum_{\bar{l}\in T} \mathbf{g}_{\bar{l}}(\mathbf{v},\mathbf{s})(\hat{\mathbf{v}}_{\bar{l}} - \mathbf{v}_{\bar{l}}) \geq 0 \text{ for any } \hat{\mathbf{v}} \in \Omega(\mathbf{s}) \tag{21}$$

where $\Omega(\mathbf{s})$ is the set of feasible multi-type container flow patterns for any given network design solution $\mathbf{s}$ and can mathematically be expressed as below,

$$\Omega(\mathbf{s}) = \left\{ \mathbf{v} \,|\, v^p_{\bar{a}} = \sum_{o\in O} \sum_{d\in D} \sum_{r\in R^p_{od}} f^p_{odr} \delta^{\bar{a}p}_{odr}, \forall \bar{a} \in A_a, a \in \mathcal{A}_0 \cup \mathcal{A}_1, p \in \mathcal{P} \right. \tag{22}$$

$$v^p_{\bar{a}} = \sum_{o\in O} \sum_{d\in D} \sum_{r\in R^p_{od}} x_a f^p_{odr} \delta^{\bar{a}p}_{odr}, \forall \bar{a} \in A_a, a \in \mathcal{A}_2, p \in \mathcal{P} \tag{23}$$

$$v_{\bar{l}}^{p} = \sum_{o \in O} \sum_{d \in D} \sum_{r \in R_{od}^{p}} f_{odr}^{p} \delta_{odr}^{\bar{l}p}, \forall \bar{l} \in T_{h}^{b}, b \in B_{h}^{1}, h \in H, p \in P \tag{24}$$

$$v_{\bar{l}}^{p} = \sum_{o \in O} \sum_{d \in D} \sum_{r \in R_{od}^{p}} y_{b} f_{odr}^{p} \delta_{odr}^{\bar{l}p}, \forall \bar{l} \in T_{h}^{b}, b \in \mathcal{B}_{h}^{2}, h \in \mathcal{H}, p \in \mathcal{P} \tag{25}$$

$$T_{od}^{p} = \sum_{r \in R_{od}^{p}} f_{odr}^{p}, \forall o \in O, d \in D, p \in \mathcal{P} \tag{26}$$

$$\left. v_{\bar{a}}^{p}, v_{\bar{l}}^{p} \geq 0, \forall \bar{a} \in A, \bar{l} \in T, p \in \mathcal{P} \right\} \tag{27}$$

3.3. IHSND problem incorporating the planner, carriers, hub operators and intermodal operators

As explained previously, carrier link $\bar{a} \in A$ or transshipment line $b \in \mathcal{B}_h$, $h \in \mathcal{H}$ is operated by an individual carrier or hub operator, and let $c_{\bar{a}}(\mathbf{v}_{\bar{a}})$ and $c_b(\mathbf{v}_b)$ be the cost functions of them, respectively, where $\mathbf{v}_b$ is the vector multi-type container flow on transshipment line b, represented by

$$\mathbf{v}_b = \left(v_{\bar{l}}^{p} \bar{l} \in T_h^b, p \in \mathcal{P} \right) \tag{28}$$

These two cost functions should have the following properties in order to realistically reflect the cost structure of multi-type container transportation: (i) capable of evaluating the cost incurred in transporting multi-type containers, (ii) increasing with respect to the flow of each container type $p \in \mathcal{P}$, (iii) zero inputs of container flows are allowed; and (iv) the unit transportation cost function is U-shaped to reflect the transition from economies of scale to diseconomies of scale in different flow regimes.

A translog cost function which represents a local and second-order approximation to an arbitrary cost function is one cost function that fulfills the requirements of the above four properties. Translog cost functions have been extensively employed by economists to describe the cost structure of a multiproduct production system (Caves et al., 1980; Fuss and Waverman, 1981; Baumol et al., 1988). A translog cost function was also deemed to be appropriate for describing the cost structure of a freight transportation system by Winston (1983). Caves et al. (1981) applied a translog cost function to investigate the productivities of US railroads and Wang and Liao (2006) calibrated a translog cost function for examining the performance of Taiwan Railway in transporting heterogeneous commodities–passengers and cargo. However, a translog function possessing a logarithm operator does not allow zero outputs, which can be simply remedied by using a hybrid translog cost function (Baumol et al., 1988).

The multi-type container transportation or transshipment process operated by a carrier or hub operator can be regarded as a production system. Let $\mathcal{Q}$ be the set of all inputs needed to accomplish the operation over carrier link $\bar{a}$ or transshipment line b, such as machinery resources, labor, materials and fuel consumption. Let π_q be the price of input $q \in \mathcal{Q}$. Multi-type container flows handled by the carrier or hub operator are the outputs of this system. The hybrid translog cost function for the carrier on $\bar{a}$ can be written as:

$$\begin{aligned} c_{\bar{a}}(\mathbf{v}_{\bar{a}}) = \exp\Bigg(& \alpha_0 + \sum_{i=1}^{n_{\mathcal{P}}} \alpha_i((v_{\bar{a}}^i)^\theta - 1)/\theta + \sum_{i=1}^{n_{\mathcal{Q}}} \beta_i \ln \pi_i + \sum_{i=1}^{n_{\mathcal{P}}} \sum_{j=1}^{n_{\mathcal{P}}} \delta_{ij}((v_{\bar{a}}^i)^\theta - 1)((v_{\bar{a}}^j)^\theta - 1)/(2\theta^2) \\ & + \sum_{i=1}^{n_{\mathcal{Q}}} \sum_{j=1}^{n_{\mathcal{Q}}} \gamma_{ij} \ln \pi_i \ln \pi_j / 2 + \sum_{i=1}^{n_{\mathcal{P}}} \sum_{j=1}^{n_{\mathcal{Q}}} \rho_{ij} \ln \pi_j \left((v_{\bar{a}}^i)^\theta - 1 \right)/\theta \Bigg) \end{aligned} \tag{29}$$

where α_0 α_i, β_i, δ_{ij}, γ_{ij}, ρ_{ij} and θ are the parameters to be calibrated and these parameters should fulfill the following conditions:

$$\delta_{ij} = \delta_{ji}, \quad \gamma_{ij} = \gamma_{ji}, \quad i = 1, 2, \ldots, n_{\mathcal{P}}, \quad j = 1, 2, \ldots, n_{\mathcal{Q}} \tag{30}$$

$$\sum_{i=1}^{n_{\mathcal{Q}}} \beta_i = 1; \qquad \sum_{i=1}^{n_{\mathcal{Q}}} \gamma_{ij} = 0, \quad j = 1, 2, \ldots, n_{\mathcal{Q}}; \qquad \sum_{j=1}^{n_{\mathcal{Q}}} \rho_{ij} = 0, \quad i = 1, 2, \ldots, n_{\mathcal{P}} \tag{31}$$

$$\lim_{\theta \to 0} ((v_{\bar{a}}^p)^\theta - 1)/\theta = \ln v_{\bar{a}}^p \tag{32}$$

Similarly, the hybrid translog cost function for the hub operator on transshipment line b can be obtained by using the functions defined in Eqs. (29)–(32) with v replaced by w and $\bar{a}$ replaced by b. The flow of container p loaded over transshipment line b is expressed by

$$w_b^p = \sum_{\bar{l} \in T_h^b} v_{\bar{l}}^p \tag{33}$$

It should be pointed out that the hybrid translog cost function is sufficiently flexible for reflecting various properties of the multi-type container transportation or transshipment cost, such as scale and scope economies. Since it permits zero outputs, the function is capable of mimicking a realistic transportation operation process where a fixed cost would be incurred in the absence of outputs. On the other hand, the function involves several parameters which may result in a complex calibration process to specify the function form.

Given a physical network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, the IHSND problem aims to construct an intermodal hub-and-spoke network for multi-type container delivery. To achieve this goal, the network planer needs to make network design decisions with the aim to minimize the total transportation cost of carriers and operational cost of hub operators given the investment budget limit B. Note that a feasible network design solution $\mathbf{s}$ is obtained based on flow pattern $\mathbf{v}$. On the other hand, given a feasible $\mathbf{s}$, the flow pattern $\mathbf{v}$ results from the UE based route choice of intermodal operators.

4. Model formulation

The IHSND problem proposed in this study can be formulated as the mathematical program with equilibrium constraints (MPEC):

$$\min f(\mathbf{s},\mathbf{v}) = \sum_{a\in(\mathcal{A}_0\cup\mathcal{A}_1)}\sum_{\bar{a}\in\mathcal{A}_a} c_{\bar{a}}(\mathbf{v}_{\bar{a}}) + \sum_{a\in\mathcal{A}_2}\sum_{\bar{a}\in\mathcal{A}_a} x_a c_{\bar{a}}(\mathbf{v}_{\bar{a}}) + \sum_{b\in\mathcal{B}_1} c_b(\mathbf{v}_b) + \sum_{b\in\mathcal{B}_2} y_b c_b(\mathbf{v}_b) \tag{34}$$

subject to

$$\sum_{a\in\mathcal{A}} B_a x_a + \sum_{b\in\mathcal{B}} y_b F_b \le B \tag{35}$$

$$\sum_{b\in\mathcal{B}_h} y_b \ge z_h, \quad \text{for } \forall h\in\mathcal{N}_2 \tag{36}$$

$$\sum_{b\in\mathcal{B}_h} y_b \le M z_h, \quad \text{for } \forall h\in\mathcal{N}_2 \tag{37}$$

$$x_a \le z_{\tilde{h}_a} + z_{\tilde{t}_a}, \quad \text{for } \forall a\in\mathcal{A}_2 \tag{38}$$

$$\sum_{r\in R^p_{od}}\left(\prod_{a\in\mathcal{A}_r\cap\mathcal{A}_2} x_a\cdot\prod_{b\in\mathcal{B}_r\cap\mathcal{B}_2} y_b\right)\ge 1, \quad \text{for } \forall o\in\mathcal{O}, d\in\mathcal{D}, p\in\mathcal{P} \tag{39}$$

$$\kappa_a\sum_{\bar{a}\in A_a}\sum_{p\in\mathcal{P}}\beta^p_a v^p_{\bar{a}}/\eta_a \le (1-x_a)\Gamma_a + x_a\Theta_a, \forall a\in\mathcal{A} \tag{40}$$

$$\sum_{\bar{l}\in T^b_h}\sum_{p\in\mathcal{P}}\beta^p_b v^p_{\bar{l}} \le (1-y_b)\Gamma_b + y_b\Theta_b, \forall b\in\mathcal{B}_h, h\in\mathcal{H} \tag{41}$$

$$v^p_{\bar{a}} \le \Gamma^p_{\bar{a}}, \quad \text{for } \forall\bar{a}\in A, p\in\mathcal{P} \tag{42}$$

$$\sum_{\bar{a}\in A}\mathbf{g}_{\bar{a}}(\mathbf{v},\mathbf{s})(\hat{\mathbf{v}}_{\bar{a}}-\mathbf{v}_{\bar{a}}) + \sum_{\bar{l}\in T}\mathbf{g}_{\bar{l}}(\mathbf{v},\mathbf{s})(\hat{\mathbf{v}}_{\bar{l}}-\mathbf{v}_{\bar{l}})\ge 0, \hat{\mathbf{v}}\in\Omega(\mathbf{s}) \tag{43}$$

Objective function (34) minimizes the total transportation cost of carriers and operational cost of hub operators. Constraint (35) represents the budget limit. Constraints (36) and (37) ensure that once a new hub is selected, at least one transshipment line will be established at the hub; no transshipment line will be built otherwise. According to constraint (38), no new direct links between two spoke nodes are established. Constraint (39) ensures that at least one intermodal route exists between each O/D pair for each container type. Constraint (40) restricts that the standard traffic flow loaded on a physical link cannot exceed its capacity. Constraint (41) is the capacity constraint with respect to each container type for each transshipment line. Constraint (42) is the capacity constraint for each carrier. The parametric VI (43) reflects the interaction between network design decisions and network flow patterns. M is an arbitrarily big number.

In the MPEC model, the objective is to minimize the total transportation cost of carriers and operational cost of hub operators from the planner's perspective. The model can flexibly cater to the case that the planner aims to maximize the total profit of carriers and hub operators through replacing the objective (34) by the following expression,

$$\begin{aligned}\max f(\mathbf{s},\mathbf{v}) = &\sum_{a\in(\mathcal{A}_0\cup\mathcal{A}_1)}\sum_{\bar{a}\in A_a}\left[\sum_{p\in\mathcal{P}} r^p_{\bar{a}} v^p_{\bar{a}} - c_{\bar{a}}(\mathbf{v}_{\bar{a}})\right] + \sum_{a\in\mathcal{A}_2}\sum_{\bar{a}\in A_a} x_a\left[\sum_{p\in\mathcal{P}} r^p_{\bar{a}} v^p_{\bar{a}} - c_{\bar{a}}(\mathbf{v}_{\bar{a}})\right]\\ &+\sum_{h\in\mathcal{H}}\sum_{b\in\mathcal{B}^1_h}\left[\sum_{\bar{l}\in T^b_h}\sum_{p\in\mathcal{P}} r^p_{\bar{l}} v^p_{\bar{l}} - c_b(\mathbf{v}_b)\right] + \sum_{h\in\mathcal{H}}\sum_{b\in\mathcal{B}^2_h} y_b\left[\sum_{\bar{l}\in T^b_h}\sum_{p\in\mathcal{P}} r^p_{\bar{l}} v^p_{\bar{l}} - c_b(\mathbf{v}_b)\right]\end{aligned} \tag{44}$$

5. Solution algorithm

The discrete network design problem has been proved to be NP-hard (Johnson et al., 1978). The IHSND problem is even harder to solve due to the nonconvexity and nondifferential characteristics of the MPEC model including a parametric VI. We thus have to seek a heuristic method to solve the MPEC model (34)–(43). In recent decades, genetic algorithms have seen a number of applications in solving bilevel programming models (Ge et al., 2003), which is a special case of MPEC model. We propose a hybrid GA (HGA) with an embedded diagonalization algorithm to solve the IHSND problem. To penalize the candidate solutions violating capacity constraints (40)–(42), we first define the evaluation function as sum of the objective function and penalty terms as follows:

$$h(\mathbf{s},\mathbf{v}) = f(\mathbf{s},\mathbf{v}) + \sum_{\bar{a}\in A}\sum_{p\in\mathcal{P}} r_{\bar{a}}(\max(v_{\bar{a}}^p - \Gamma_{\bar{a}}^p, 0))^2 + \sum_{a\in\mathcal{A}} r_a(\max(w_a - (1-x_a)\Gamma_a - x_a\Theta_a, 0))^2 + \sum_{b\in\mathcal{B}} r_b(\max(w_b - (1-y_b)\Gamma_b - y_b\Theta_b, 0))^2 \quad (45)$$

where r_a, $r_{\bar{a}}$ and r_b are positive variable penalty coefficients. The hybrid GA for solving the MPEC model (34)–(43) is presented in the following four steps.

5.1. Hybrid genetic algorithm

Step 0. (*Initialization*) Randomly determine an initial population consisting of λ distinct chromosomes satisfying constraint conditions (35)–(39) and denote the population by set $\mathcal{S}^n = \{\mathbf{s}_i^n | i = 1, 2, \ldots, \lambda\}$, of which bit string $\mathbf{s}_i^n$ is used to represent chromosome i. Let the number of iterations $n = 0$.

Step 1. (*Crossover and mutation*) Perform one-cut-point crossover and mutation operations on the chromosomes selected from $\mathcal{S}^n$ based on specified crossover probability p_c and mutation probability p_m respectively. Let $\widetilde{\mathcal{S}}^n$ be the set of all resulting offspring chromosomes that satisfy constraint conditions (35)–(39) and μ_n be the cardinality of $\widetilde{\mathcal{S}}^n$. Replace parent chromosomes in $\mathcal{S}^n$ by their corresponding offspring in $\widetilde{\mathcal{S}}^n$ and go to next step if $\mu_n \geqslant 1$; re-execute Step 1 otherwise.

Step 2. (*Fitness evaluation*) Two substeps are executed to calculate fitness of each chromosome in population $\mathcal{S}^n$ as follows.

Step 2.1 (*Network flow calculation*) For a chromosome $\mathbf{s}_i^n \in \mathcal{S}^n (i = 1, 2, \ldots, \lambda)$, solve the relevant parametric VI (43) using a diagonalization algorithm to obtain the UE flow pattern $\mathbf{v}_i^n$.

Step 2.2 (*Fitness normalization*) Given $\mathbf{s}_i^n$ and $\mathbf{v}_i^n (i = 1, 2, \ldots, \lambda)$, calculate evaluation value $h_i^n = h(\mathbf{s}_i^n, \mathbf{v}_i^n)$ of chromosome i defined in Eq. (45). The fitness of each chromosome can be then computed by normalizing its evaluation value using the following equation:

$$\tilde{h}_i^n = [h_{\max}^n - h_i^n(\mathbf{s}_i^n, \mathbf{v}_i^n) + \gamma]/[h_{\max}^n - h_{\min}^n + \gamma], i = 1, 2, \ldots, \lambda \quad (46)$$

where $h_{\max}^n$ and $h_{\min}^n$ are maximum and minimum ones out of evaluation values of all chromosomes included in $\mathcal{S}^n$, respectively, and $\gamma \in (0, 1)$.

Step 3. (*Breed a new population*) Generate a population $\mathcal{S}^{n+1}$ of size λ by using a binary tournament selection method (Gen and Cheng, 1997) and the fitness of each chromosome defined in (46); set $n = n + 1$ and go to Step 1.

Step 4. (*Stopping test*) Stop if $h_{\min}^{(n+1)} - h_{\min}^n < \varepsilon_1$ where ε_1 is a positive tolerance error or $n > N$ where N is an arbitrary population generation number.

The IHSND problem has binary integer variables $\mathbf{x} = (x_1, x_2, \ldots, x_{n_\mathbf{x}}), \mathbf{y} = (y_1, y_2, \ldots, y_{n_\mathbf{y}}), \mathbf{z} = (z_1, z_2, \ldots, z_{n_\mathbf{z}})$, where $n_\mathbf{x}$, $n_\mathbf{y}$, and $n_\mathbf{z}$ are the number of entities in the three vectors. Each chromosome encodes a possible solution of $\mathbf{s} = (\mathbf{x}, \mathbf{y}, \mathbf{z})$ and comprises $(n_\mathbf{x} + n_\mathbf{y} + n_\mathbf{z})$ genes ranging according to the sequence $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ on the chromosome. Each gene corresponds with a binary decision variable with the same location as the gene. The chromosome is formed in such a way that each gene is assigned value of one if the investment action represented by the gene's corresponding decision variable is implemented; zero otherwise.

Constraint (39) actually guarantees the connectivity between an O/D pair for transporting each type of containers. However, it is somewhat problematic for examining the constraint condition since enumerating all intermodal routes between each O/D pair is unacceptably time consuming. Hence, the constraint condition (39) is validated by using link-label setting shortest path algorithm that has been employed to find the optimum intermodal/multimodal shortest path. The algorithm is sufficiently efficient with time complexity $O(n_A + n_T)^2$ and it has also been incorporated into the Frank-Wolf algorithm (LeBlanc et al., 1975) that is dedicated to solving the parametric VI in Step 2.1. Given solution $\mathbf{s}$, constraint (39) is satisfied if at least one intermodal shortest path can be found for each container type between each O/D pair.

The diagonalization algorithm used in Step 2.1 is described in the following substeps:

Step 2.1.1 (*Initialization*) Given a chromosome $\mathbf{s}_i^n \in \mathcal{S}^n (i = 1, 2, \ldots, \lambda)$, determine an initial flow pattern $\mathbf{v}^0 \in \Omega(\mathbf{s}_i^n)$ and set $k = 0$.

Step 2.1.2 (*Solve the diagonalized problem*) Apply Frank-Wolf algorithm to solve a general UE problem presented by VI (47) to obtain a network flow pattern $\mathbf{v}$ and let $\mathbf{v}^{k+1} = \mathbf{v}$.

$$\sum_{p\in\mathcal{P}}\sum_{\bar{a}\in A}(\hat{v}_{\bar{a}}^p - v_{\bar{a}}^p) g_{\bar{a}}^{(k)p}(v_{\bar{a}}^p, \mathbf{s}_i^n) + \sum_{p\in\mathcal{P}}\sum_{\bar{l}\in T}(\hat{v}_{\bar{l}}^p - v_{\bar{l}}^p) g_{\bar{l}}^{(k)p}(v_{\bar{l}}^p, \mathbf{s}_i^n) \geq 0, \forall \hat{\mathbf{v}} \in \Omega(\mathbf{s}_i^n) \quad (47)$$

where for $\forall e \in A \cup T, p \in \mathcal{P}$,

$$g_e^{(k)p} = g_e^p\left[v_e^p, v_e^{\hat{p}} = v_e^{(k)\hat{p}}(\hat{p} \in \mathcal{P} \neq p), v_j^{\hat{p}} = v_j^{(k)\hat{p}}(\forall j \in A \cup T \neq e, \hat{p} \in \mathcal{P}), \mathbf{s}_i^n\right] \tag{48}$$

Step 2.1.1 (*Stopping test*). If

$$\|\mathbf{v}^{(k+1)} - \mathbf{v}^k\|_2 / \|\mathbf{v}^k\|_1 \le \varepsilon_2 \tag{49}$$

where ε_2 is an small positive value, then stop and let $\mathbf{v}_i^n = \mathbf{v}^{k+1}$; set $k = k + 1$ and go to Step 2.1.3 otherwise.

The diagonalization algorithm used in Step 2.1 converges if the Jacobian matrix of disutility functions with respect to network flow pattern **v** is diagonally dominant (Dafermos, 1983).

6. Numerical examples

Fig. 2 shows an example intermodal freight transportation network used for assessing the applicability of the developed MPEC model and hybrid genetic algorithm. The network is compiled by referring to the working plans of the Trans-Asian Railway and Asian Highway projects initiated by UNESCAP (2003,2009), and involves the mode set $\mathcal{M} = \{\text{rail}, \text{truck}, \text{sea}\}$ and container type set $\mathcal{P} = \{1 = \text{general container}, 2 = \text{refrigerated container}\}$. Each link segment contains two physical

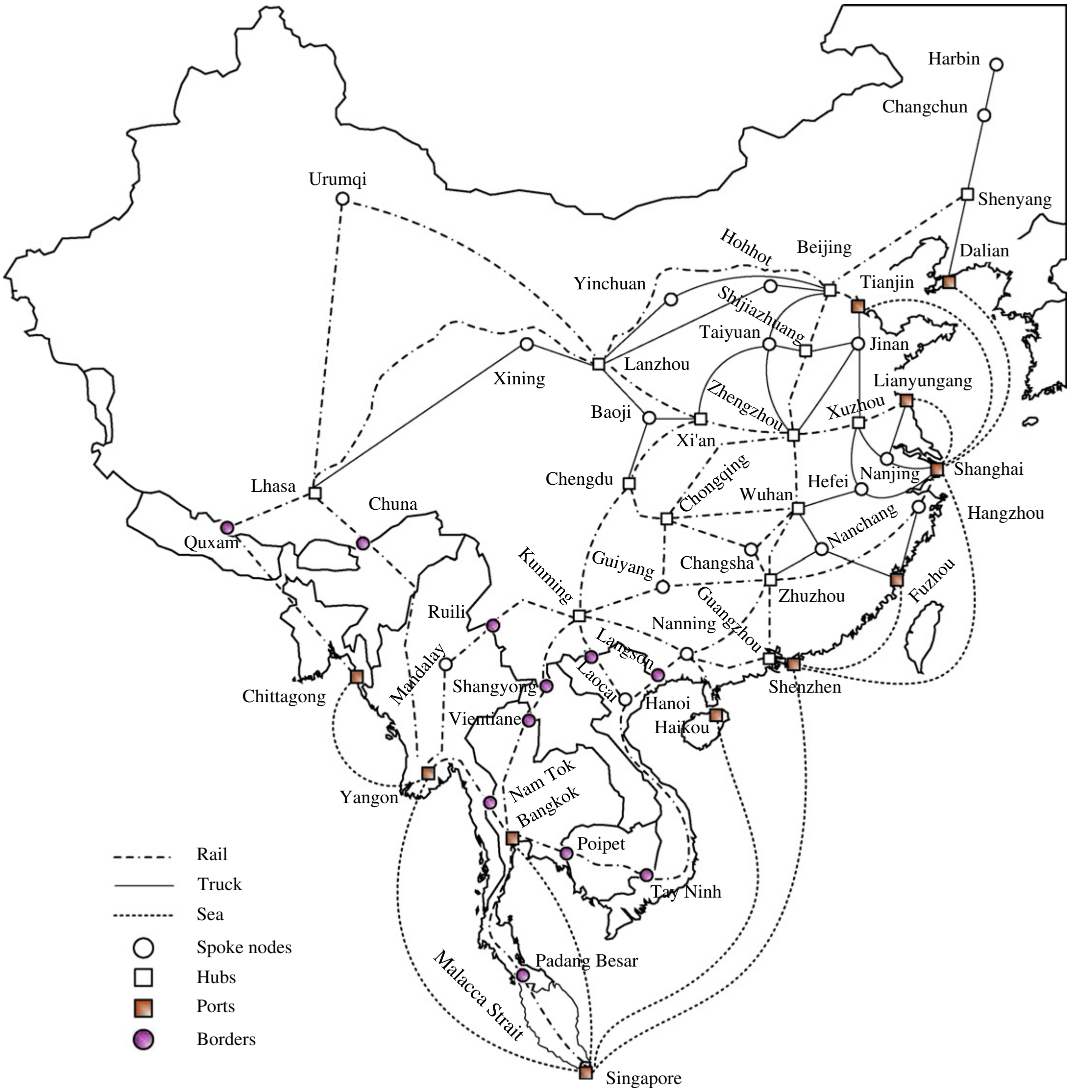

Fig. 2. An example intermodal transportation network.

links in opposite directions. It is further hypothesized that each physical link is operated by one carrier. A and B in Appendix A show the nodes and links in the network in Appendix A, respectively.

As reported by UNCTAD (2008), the intra-Asia trade is estimated at 40 million TEUs in 2007. We distribute the intra-Asia container traffic over 25 O/D pairs by using a doubly constraint gravity model with transportation time as deterrence, and the resulting O/D demand matrix is tabulated in Table C in Appendix A. Furthermore, the demands of general containers and refrigerated containers between each O/D pair are estimated at 10% and 90% of the total container demand, respectively. Table 1 shows the parameter values for disutility functions based on Wang et al. (2009) and Meng and Wang (2010). In the table, $s_{\bar{a}}^1$ and $s_{\bar{a}}^2$ represent the free-flow speeds for transporting one general container and refrigerated container on carrier link $\bar{a}$, respectively, and the free-flow transportation time $t_{\bar{a}}^{10}$ and $t_{\bar{a}}^{20}$ defined in Eq. (15) can be calculated by using link length divided by $s_{\bar{a}}^1$ and $s_{\bar{a}}^2$ respectively.

Table 2 shows the assumed parameter values for the hybrid translog cost functions of carriers and hub operators. As explained earlier, these functions can flexibly reflect various properties of multi-type container transportation cost or transshipment cost, such as scale and scope economies. They also bring difficulties in calibrating their realistic forms owing to having many parameters. To apply them to realistic applications, a careful calibration of the functions is indeed important, since they may significantly impact the results of IHSND. This is a research topic accompanied with this study. Currently, we take assumed values for the parameters and expect to calibrate the function and analyze the sensitivity of the IHSND results to these parameter values in the future work.

6.1. Comparison of two algorithms

We use two numerical examples to validate the effectiveness of the proposed MPEC model and HGA based on the example intermodal network shown in Fig. 2. In these two examples, the proposed HGA with a binary tournament selection method and exhaustive enumeration algorithm (EEA) are employed to solve the IHSND problem with 16 possible projects to be planned for network design, respectively. Table 3 shows the projects to be planned in the comparison analysis. In reality, it is somewhat unrealistic to involve too many network design projects in intermodal freight transportation network design, since most network elements currently exist and only key links and transshipment lines need to be established or enhanced (Yamada et al., 2009). Besides, network design projects are highly cost-consuming, and the budget of a network planner normally allows only a limited number of projects to be planned.

The implementations of the two algorithms are described below.

Table 1
Parameter values for disutility functions.

$\bar{a} \in A$	$(s_{\bar{a}}^1, s_{\bar{a}}^2)$ (km/h)	$(r_{\bar{a}}^1, r_{\bar{a}}^2)$ ($/km)	$(\phi_{\bar{a}}^1, \phi_{\bar{a}}^2)$ ($/h)	$(\omega_{\bar{a}}^1, \omega_{\bar{a}}^2)$	$(\alpha_{\bar{a}}^1, \alpha_{\bar{a}}^2)$	$k_{\bar{a}}^p$
Rail	(33, 33)	(0.1, 0.2)	(0.1, 0.1)	(2.0, 2.0)	(2.5, 2.5)	0.5
Road	(23, 23)	(0.8, 1.2)	(1.5, 1.5)	(2.0, 2.0)	(2.5, 2.5)	0.5
Sea	(22, 22)	(0.2, 0.4)	(0.1, 0.1)	(2.0, 2.0)	(2.5, 2.5)	0.5
$\bar{l} \in T$	$(t_{\bar{l}}^1, t_{\bar{l}}^2)$ (h)	$(r_{\bar{l}}^1, r_{\bar{l}}^2)$ ($)	$(\phi_{\bar{l}}^1, \phi_{\bar{l}}^2)$ ($/h)	$(\omega_{\bar{l}}^1, \omega_{\bar{l}}^2)$	$(\alpha_{\bar{l}}^1, \alpha_{\bar{l}}^2)$	$k_{\bar{l}}^p$
Ports	(24, 24)	(387, 580)	(5.0, 5.0)	(2.0, 2.0)	(2.5, 2.5)	0.5
Hubs	(12, 12)	(27, 40)	(5.0, 5.0)	(2.0, 2.0)	(2.5, 2.5)	0.5
Borders	(48, 48)	(293, 493)	(5.0, 5.0)	(2.0, 2.0)	(2.5, 2.5)	0.5

$a \in \mathcal{A}$	Θ_a (std vehicles/year)	κ_a	η_a	(β_a^1, β_a^2)	B_a
Rail	5000	1.0	0.5	(0.25, 0.25)	4.0
Road	3000	2.0	0.1	(0.5, 0.5)	2.5
Sea	1000	1.0	0.6	(10^{-3}, 10^{-3})	1.0

$b \in \mathcal{B}$	Θ_b (1000 TEUs/year)	Γ_b (1000 TEUs/year)	(β_b^1, β_b^2)	F_b
$b \in \mathcal{B}_1$	1000	500	(1.0, 1.0)	1
$b \in \mathcal{B}_2$	1000	0	(1.0, 1.0)	2

Table 2
Parameter values for cost functions.

Parameters	α_0	(α_1, α_2)	$(\delta_{11}, \delta_{12}, \delta_{22})$	θ
Rail links	−2	(0.01, 0.01)	(0.01, 0.0359, 0.01)	0.001
Road links	−2	(0.01, 0.01)	(0.01, 0.0543, 0.01)	0.05
Sea links	−4	(0.01, 0.01)	(0.01, 0.0978, 0.01)	0.001
Transshipment lines	6	(0.01, 0.01)	(0.01, 0.0359, 0.01)	0.057

Table 3
Network design projects to be planned in comparison analysis.

Actions	Links or location	Optimal solution	
		HGA	EEA
Hub location	Kunming	1	1
Hub location	Lhasa	1	1
Transshipment line establishment	{rail–rail} at Kunming	1	1
Transshipment line establishment	{rail–truck} at Lhasa	1	0
Transshipment line establishment	{rail–rail} at Lhasa	1	1
Transshipment line establishment	{truck–rail} at Lhasa	1	1
Rail link establishment	Urumqi → Lhasa	1	1
Rail link establishment	Ruili → Kunming	0	1
Rail link establishment	Nanning → Langson	1	0
Rail link establishment	Kunming → Shangyong	1	0
Rail link establishment	Kunming → Laocai	1	1
Rail link establishment	Lhasa → Urumqi	0	1
Rail link establishment	Kunming → Ruili	0	1
Rail link establishment	Langson → Nanning	0	0
Rail link establishment	Shangyong → Kunming	0	0
Rail link establishment	Laocai → Kunming	1	1

(i) *HGA*: the hybrid GA is described earlier. Initialization, crossover, mutation and fitness evaluation are the same as those in a general genetic algorithm. After mutation, the algorithm draws two chromosomes *N* times. At each time, the two individuals are compared in terms of fitness and the best one will be kept in the next population.
(ii) *EEA*: a binary search tree with 2^{16} nodes at the final depth is used to obtain the optimal solution. The tree contains all feasible solutions, each of which is represented by a particular branch of the tree. All feasible solutions are consecutively tested and the optimal solution is finally obtained.

The algorithms are coded in C++ and executed by using a desktop with CPU of Core 2 Duo 3.00 GHz and 4G RAM. Let parameters $r_{\bar{a}} = r_a = r_b = 10$ for $\bar{a} \in A, a \in \mathcal{A}$ and $b \in \mathcal{B}$, crossover probability $p_c = 0.25$, mutation probability $p_m = 0.01$, the total budget $B = 30$, the number of generations $N = 100$, population size $\lambda = 50$, $\varepsilon_2 = 10^{-6}$ and $\gamma = 0.05$.

The computational results by using the HGA and EEA are compared in terms of computational time and objective value in Table 4. The optimal network design decisions obtained by using the two algorithms are shown in Table 3. As indicated by Table 4, undoubtedly, the EEA can find the best solution – the global optimal solution by searching the whole feasible space. Meanwhile, the HGA can provide a reasonably good solution for running 100 generations, which gives a 0.6% bigger objective value than the globally minimum objective value. The small difference results from two facts. First, the objective value is strongly associated with the SUE-based network flow, which is obtained by solving a VI based on a given network structure. In the numerical analysis, 174 physical links and 88 transshipment lines are assumed to always exist in the network, and only 16 network elements are affected by the network decision. Although EEA and HGA give different optimal decisions for the 16 projects to be planned, the network structure seems not to change much. Thus, no matter what the network decision is, the SUE-based network flow by solving the VI does not change much. This explains, at least to some extent, why the objective value does not vary significantly when different network decisions are concerned. Second, the parameter values used for hybrid translog cost functions are assumed in numerical analysis. The resulting cost functions may not be sensitive to the change in flows, i.e., flow changes resulting from different network decisions may give rise to an indistinguishable change in transportation cost. To deal with this limitation, a realistic hybrid translog cost function needs to be calibrated based on historical data and practical transportation operations.

The EEA can provide an optimal solution, however, since the IHSND problem is a NP-hard problem, a small increase in the current number of binary variables will give rise to an exponentially increasing computational time, which makes the EEA unacceptable for solving the a larger IHSND problem. Additionally, the computational time per iteration by using the HGA is mainly attributed to executing the diagonalization algorithm for solving the parametric VI. The total computation time would not change much even if a larger number of possible projects are planned in the example network shown in Fig. 2. The results indicate that the HGA provides a reasonable good solution in terms of both computation time and solution quality.

Table 4
Comparison of the two algorithms.

	HGA	EEA
CPU time	236.64 h	1084.8 h
Min. objective value	1107,468,713	1106,710,821
Max. objective value	1107,468,713	1575,785,221

6.2. Application of the HGA in a large-scale IHSND problem

We apply the proposed HGA to a large-scale IHSND problem based on the example network shown in Fig. 2, assuming that all physical links need to be established in addition to planning the transshipment lines and hubs shown in Table 5. Let $B = 650$ units and $N = 50$, and all remaining parameter values and the operating environment are the same as those used in the above example. The resulting IHSND problem has 209 possible projects (0–1 variables) to be planned, which consists of 184 potential physical links shown in Table B and 25 hub and transshipment line establishment projects shown in Table 5. Fig. 3 shows the change in the minimum objective value with respect to the number of generation Table A.

The solutions for hub locations and transshipment line establishments by using HGA are shown in Table 5, and those for physical links are given in Table B in Appendix A. In the Table B, x_{a^*} and $x_{\hat{a}^*}$ represent the optimal solutions with respect to physical links $a \in \mathcal{A}$ and $\hat{a} \in \mathcal{A}$ in the opposite direction of a, respectively Table C.

As implied by the results, the {truck–truck} transshipment line should not be established at Zhengzhou. It indicates that container flows originated from Taiyuan and Jinan are not switched via Zhengzhou, and they may be transshipped to each other via Shijiazhuang due to the shorter distance by traversing Shijiazhuang. As shown in Table B, most of physical links need to be established to maintain the connectivity of the intermodal network for container delivery between 25 O/D pairs. Since many links in Southeast Asian countries are currently still missing, the Trans-Asian Railway and Asian Highway projects initiated by UNESCAP should be given a high priority in order to adapt to the growing containerized cargo trade and economic exchanges in China and Mainland Southeast Asia.

In addition, the network design solution in Table 5 suggests one-way links between some pairs of cities. The solution implies that only unidirectional rail tracks, road lanes, or maritime passages need to be established between these city pairs.

Table 5
Network design projects to be planned in the large-scale example.

Hubs/transshipment lines	Optimal solution	Hubs/transshipment lines	Optimal solution
Zhengzhou	1	{truck–rail} at Lanzhou	1
Lanzhou	1	{rail–rail} at Kunming	1
Kunming	1	{rail–truck} at Lhasa	1
Lhasa	1	{rail–rail} at Lhasa	1
Xuzhou	1	{truck–rail} at Lhasa	1
Zhuzhou	1	{truck–rail} at Xuzhou	1
{truck–rail} at Zhengzhou	1	{truck–truck} at Xuzhou	1
{truck–truck} at Zhengzhou	0	{rail–rail} at Xuzhou	1
{rail–rail} at Zhengzhou	1	{rail–truck} at Xuzhou	1
{rail–truck} at Zhengzhou	1	{rail–rail} at Zhuzhou	1
{rail–truck} at Lanzhou	1	{rail–truck} at Zhuzhou	1
{rail–rail} at Lanzhou	1	{truck–rail} at Zhuzhou	1
{truck–truck} at Lanzhou	1		

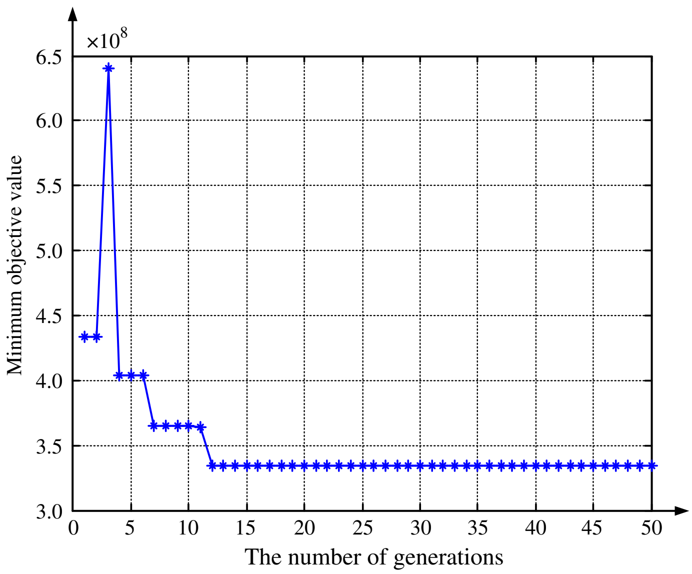

Fig. 3. Minimum objective value in each generation for the large-scale example.

Theoretically, it is possible to obtain such a solution by solving the MPEC model. We model the road lanes, rail tracks or maritime passages heading in the same direction as one physical link between each pair of nodes, and consider those heading in the opposite direction as a different physical link. Each of these two physical links is assigned a binary variable representing the link establishment decision. Any solution can possibly be obtained by using the HGA. Practically, the solution reflects the phenomenon that the optimal intermodal route from one city A to another city B may not traverse the same set of nodes as that from B to A, as far as the above-mentioned physical link modeling method is concerned. For instance, $A \to C \to B$ is the best route from cities A to B in terms of disutility; the best route from cities B to A may be $B \to D \to A$ instead of $B \to C \to A$. As a result, a large amount of containers flows will be motivated to traverse link AC and link AC needs to be established. On the other hand, link CA is not necessarily established and the budget may be allocated to other physical links which are located in the optimal intermodal routes. Thus, one-way link establishment decisions may happen.

7. Conclusions

In this paper we proposed a novel intermodal hub-and-spoke network design problem for multi-type container flows in the context of intermodal freight transportation operations. The problem is fundamentally different from the conventional hub-and-spoke network design problem and its variants by incorporating multiple stakeholders, multi-type containers and container transfer processes at hubs.

To formulate the problem, we first represented a given intermodal network as two directed networks-a physical network that a network planner attempts to design and an operational network on which intermodal operators make intermodal route choice. Next, a joint transportation cost function that possesses a U-shaped unit cost function to reflect the transition from scale economies to scale diseconomies in distinct flow regimes was suggested to describe a carrier/hub operator's cost structure in transport of multi-type containers. A disutility function consisting of actual transportation charges and congestion impacts was proposed to reflect the preference of an intermodal operator in route choices. A mathematical program with equilibrium constraints model incorporating a parametric variational inequality was then developed to formulate the intermodal hub-and-spoke network design problem with the objective to minimize the total transportation cost of carriers and operational cost of hub operators. To solve the mathematical program with equilibrium constraints model, a hybrid genetic algorithm embedded with a diagonalization iterative scheme for user equilibrium based multi-type container flow assignment was proposed. The proposed hybrid genetic algorithm was found to have a reasonably good performance in terms of computational time and solution quality compared to the exhaustive enumeration algorithm. The genetic algorithm was also applied to solve a large-scale intermodal hub-and-spoke network design problem to assess the applicability of the proposed model and algorithm.

Table A
Nodes in the example network.

Id	Name	Type	Id	Name	Type
1	Changchun	Hub	28	Chongqing	Hub
2	Shenyang	Hub	29	Kunming	Hub
3	Shijiazhuang	Hub	30	Nanning	Spoke
4	Jinan	Spoke	31	Lhasa	Hub
5	Hefei	Spoke	32	Dalian	Port
6	Nanjing	Spoke	33	Lianyungang	Port
7	Wuhan	Hub	34	Shenzhen	Port
8	Nanchang	Spoke	35	Xuzhou	Hub
9	Fuzhou	Port	36	Zhuzhou	Hub
10	Hangzhou	Spoke	37	Baoji	Spoke
11	Hohhot	Spoke	38	Quxam	Border
12	Yinchuan	Spoke	39	Chuna	Border
13	Xian	Hub	40	Ruili	Border
14	Chengdu	Hub	41	Shangyong	Border
15	Guiyang	Spoke	42	Laocai	Border
16	Guangzhou	Hub	43	Langson	Border
17	Tianjin	Port	44	Vientiane	Border
18	Harbin	Spoke	45	Hanoi	Spoke
19	Beijing	Hub	46	Tay Ninh	Border
20	Shanghai	Port	47	Poipet	Border
21	Urumqi	Spoke	48	Bangkok	Port
22	Xining	Spoke	49	Nam Tok	Border
23	Haikou	Port	50	Yangon	Port
24	Taiyuan	Spoke	51	Mandalay	Spoke
25	Zhengzhou	Hub	52	Chittagong	Port
26	Changsha	Spoke	53	Padang Besar	Border
27	Lanzhou	Hub	54	Singapore	Port

It must be pointed out that this study has not fully addressed the issues involved in the IHSND. For instance, we assumed parameter values for hybrid translog cost functions and a realistic form of the functions needs to be calibrated. In addition, the sensitivity of the network design results to these parameter values needs to be analyzed. It is also our future research work to develop a branch-and-bound based approximation method.

Acknowledgement

This paper is supported by the NOL Fellowship Programme of Singapore (WBS No.: R-264-000-243-720).

Appendix A

Tables A–C.

Table B
Links in the example network.

Segment	Mode	Length	$(x_{a^*}, x_{\hat{a}^*})$	Segment	Mode	Length	$(x_{a^*}, x_{\hat{a}^*})$
21–27	Rail	1178.2	(1, 1)	28–15	Rail	406.9	(1, 0)
21–31	Rail	560.1	(1, 1)	7–26	Rail	397.3	(1, 1)
31–22	Truck	751.9	(1, 1)	26–36	Rail	49.2	(1, 1)
22–27	Truck	344.8	(1, 1)	28–7	Rail	1034.6	(1, 1)
31–27	Rail	309.8	(1, 1)	26–28	Rail	1084.0	(1, 1)
27–12	Truck	385.7	(1, 1)	7–36	Rail	445.6	(1, 1)
12–19	Truck	487.4	(1, 1)	8–7	Truck	354.7	(1, 1)
27–19	Rail	207.9	(1, 1)	8–9	Truck	638.1	(1, 1)
27–11	Truck	283.1	(1, 1)	8–36	Truck	348.1	(1, 1)
11–19	Truck	118.9	(1, 1)	40–29	Rail	778.1	(1, 1)
19–2	Rail	682.5	(1, 1)	29–15	Rail	572.2	(1, 1)
18–1	Truck	511.1	(1, 1)	15–36	Rail	808.5	(1, 1)
1–2	Truck	1229.0	(1, 1)	36–10	Rail	954.0	(1, 1)
2–32	Truck	1038.0	(1, 1)	10–9	Truck	728.6	(0, 1)
19–24	Truck	354.3	(0, 1)	29–30	Rail	909.1	(1, 1)
24–3	Truck	315.0	(1, 1)	30–36	Rail	951.0	(1, 1)
19–3	Rail	298.0	(1, 1)	30–16	Rail	711.6	(1, 1)
19–17	Rail	410.0	(0, 1)	30–23	Rail	516.8	(1, 1)
24–13	Truck	474.5	(1, 1)	30–43	Rail	244.9	(1, 0)
24–25	Truck	175.0	(1, 1)	36–16	Rail	698.1	(1, 1)
17–20	Sea	607.9	(1, 1)	16–34	Rail	130.0	(1, 1)
32–20	Sea	691.0	(1, 1)	9–34	Sea	795.6	(1, 1)
17–4	Truck	503.5	(0, 0)	40–51	Rail	416.2	(1, 1)
4–35	Truck	424.8	(1, 1)	29–41	Rail	721.0	(1, 1)
4–3	Truck	365.9	(1, 1)	29–42	Rail	489.8	(1, 1)
4–25	Truck	232.5	(1, 1)	51–50	Rail	661.8	(1, 1)
27–37	Truck	516.0	(1, 1)	41–44	Rail	456.0	(1, 1)
37–13	Truck	1220.6	(1, 1)	44–48	Rail	684.0	(1, 1)
27–37	Rail	563.8	(1, 1)	50–49	Rail	623.0	(0, 1)
37–14	Truck	351.9	(1, 0)	49–48	Rail	210.0	(1, 1)
13–25	Rail	343.0	(0, 1)	42–45	Rail	306.5	(1, 1)
3–25	Rail	344.8	(1, 1)	43–45	Rail	140.8	(0, 1)
25–35	Rail	503.6	(0, 1)	34–54	Sea	2621.7	(0, 1)
35–33	Rail	485.4	(1, 1)	23–54	Sea	2377.2	(0, 1)
33–20	Sea	193.0	(1, 1)	45–46	Rail	1391.5	(1, 1)
25–28	Rail	1531.7	(1, 1)	46–47	Rail	530.7	(0, 1)
25–7	Rail	865.2	(1, 1)	47–48	Rail	274.8	(1, 1)
35–5	Truck	319.7	(1, 1)	48–54	Sea	1557.1	(1, 1)
35–6	Truck	1178.2	(1, 1)	48–53	Rail	972.4	(1, 1)
6–20	Truck	560.1	(1, 1)	53–54	Rail	809.9	(1, 1)
5–20	Truck	751.9	(1, 1)	54–50	Sea	1234.3	(1, 1)
5–7	Truck	344.8	(1, 1)	50–52	Sea	735.5	(1, 1)
10–20	Rail	309.8	(1, 1)	52–38	Rail	1311.7	(1, 1)
20–34	Sea	385.7	(1, 1)	38–31	Rail	707.3	(1, 1)
14–29	Rail	487.4	(1, 1)	31–39	Rail	400.0	(1, 1)
14–28	Rail	207.9	(1, 1)	39–50	Rail	1737.1	(1, 1)

Note: the integer numbers representing segments are node Id's.

Table C
The total O/D demand matrix (1000 TEUs/year.

Origin *o*	Destination *d*																								
	1	4	5	6	8	10	11	12	15	16	18	19	20	21	22	24	26	30	37	45	48	50	51	52	54
1	0.0	26.3	8.4	18.8	5.8	13.0	9.0	0.9	2.8	22.7	6.6	8.9	12.8	3.0	0.7	7.7	8.2	5.2	5.4	1.9	3.7	0.4	0.2	2.5	107.8
4	26.3	0.0	46.3	103.6	31.8	71.9	49.3	5.2	15.2	125.0	35.0	48.7	70.4	16.6	3.8	42.1	45.3	28.9	29.5	10.3	20.6	2.5	1.2	13.7	592.0
5	8.4	46.3	0.0	33.5	10.3	23.3	15.7	1.7	4.9	40.4	11.2	15.5	22.8	5.3	1.2	13.5	14.7	9.3	9.5	3.3	6.7	0.8	0.4	4.4	192.0
6	18.8	103.6	33.5	0.0	23.0	52.2	35.2	3.7	11.0	90.4	25.0	34.8	51.1	11.9	2.7	30.1	32.7	20.9	21.2	7.4	14.9	1.8	0.9	9.8	428.9
8	5.8	31.8	10.3	23.0	0.0	16.1	10.8	1.1	3.4	28.2	7.7	10.7	15.7	3.7	0.8	9.3	10.2	6.5	6.5	2.3	4.6	0.6	0.3	3.1	133.6
10	13.0	71.9	23.3	52.2	16.1	0.0	24.5	2.6	7.6	63.1	17.4	24.1	35.7	8.3	1.9	20.9	22.8	14.5	14.7	5.2	10.4	1.2	0.6	6.8	299.2
11	9.0	49.3	15.7	35.2	10.8	24.5	0.0	1.8	5.2	42.6	12.0	16.7	23.9	5.7	1.3	14.5	15.5	10.0	10.1	3.5	7.1	0.8	0.4	4.7	202.8
12	0.9	5.2	1.7	3.7	1.1	2.6	1.8	0.0	0.6	4.5	1.3	1.8	2.5	0.6	0.1	1.5	1.6	1.0	1.1	0.4	0.7	0.1	0.0	0.5	21.5
15	2.8	15.2	4.9	11.0	3.4	7.6	5.2	0.6	0.0	13.5	3.7	5.1	7.5	1.8	0.4	4.5	4.9	3.2	3.2	1.1	2.3	0.3	0.1	1.5	64.9
16	22.7	125.0	40.4	90.4	28.2	63.1	42.6	4.5	13.5	0.0	30.1	42.0	61.7	14.4	3.3	36.4	40.2	25.9	25.8	9.2	18.5	2.2	1.1	12.2	532.5
18	6.6	35.0	11.2	25.0	7.7	17.4	12.0	1.3	3.7	30.1	0.0	11.9	17.0	4.0	0.9	10.2	10.9	7.0	7.1	2.5	5.0	0.6	0.3	3.3	143.3
19	8.9	48.7	15.5	34.8	10.7	24.1	16.7	1.8	5.1	42.0	11.9	0.0	23.6	5.6	1.3	14.2	15.2	9.7	9.9	3.5	6.9	0.8	0.4	4.6	199.1
20	12.8	70.4	22.8	51.1	15.7	35.7	23.9	2.5	7.5	61.7	17.0	23.6	0.0	8.1	1.9	20.4	22.3	14.2	14.4	5.1	10.1	1.2	0.6	6.7	291.7
21	3.0	16.6	5.3	11.9	3.7	8.3	5.7	0.6	1.8	14.4	4.0	5.6	8.1	0.0	0.5	4.9	5.2	3.4	3.5	1.2	2.4	0.3	0.1	1.6	69.4
22	0.7	3.8	1.2	2.7	0.8	1.9	1.3	0.1	0.4	3.3	0.9	1.3	1.9	0.5	0.0	1.1	1.2	0.8	0.8	0.3	0.6	0.1	0.0	0.4	15.9
24	7.7	42.1	13.5	30.1	9.3	20.9	14.5	1.5	4.5	36.4	10.2	14.2	20.4	4.9	1.1	0.0	13.2	8.4	8.7	3.0	6.0	0.7	0.4	4.0	173.0
26	8.2	45.3	14.7	32.7	10.2	22.8	15.5	1.6	4.9	40.2	10.9	15.2	22.3	5.2	1.2	13.2	0.0	9.3	9.3	3.3	6.6	0.8	0.4	4.4	190.8
30	5.2	28.9	9.3	20.9	6.5	14.5	10.0	1.0	3.2	25.9	7.0	9.7	14.2	3.4	0.8	8.4	9.3	0.0	6.0	2.2	4.3	0.5	0.3	2.8	124.8
37	5.4	29.5	9.5	21.2	6.5	14.7	10.1	1.1	3.2	25.8	7.1	9.9	14.4	3.5	0.8	8.7	9.3	6.0	0.0	2.1	4.3	0.5	0.3	2.9	123.0
45	1.9	10.3	3.3	7.4	2.3	5.2	3.5	0.4	1.1	9.2	2.5	3.5	5.1	1.2	0.3	3.0	3.3	2.2	2.1	0.0	1.6	0.2	0.1	1.0	44.9
48	3.7	20.6	6.7	14.9	4.6	10.4	7.1	0.7	2.3	18.5	5.0	6.9	10.1	2.4	0.6	6.0	6.6	4.3	4.3	1.6	0.0	0.4	0.2	2.1	92.4
50	0.4	2.5	0.8	1.8	0.6	1.2	0.8	0.1	0.3	2.2	0.6	0.8	1.2	0.3	0.1	0.7	0.8	0.5	0.5	0.2	0.4	0.0	0.0	0.3	10.9
51	0.2	1.2	0.4	0.9	0.3	0.6	0.4	0.0	0.1	1.1	0.3	0.4	0.6	0.1	0.0	0.4	0.4	0.3	0.3	0.1	0.2	0.0	0.0	0.1	5.4
52	2.5	13.7	4.4	9.8	3.1	6.8	4.7	0.5	1.5	12.2	3.3	4.6	6.7	1.6	0.4	4.0	4.4	2.8	2.9	1.0	2.1	0.3	0.1	0.0	59.9
54	107.8	592.0	192.0	428.9	133.6	299.2	202.8	21.5	64.9	532.5	143.3	199.1	291.6	69.4	15.9	173.0	190.8	124.8	123.0	44.9	92.4	10.9	5.4	59.9	0.0

References

Abdelwahab, W.M., 1998. Elasticities of mode choice probabilities and market elasticities of demand: evidence from a simultaneous mode choice/shipment size freight transport models. Transportation Research Part E 34 (4), 257–266.

Alumur, S., Kara, B.Y., 2008. Network hub location problems: the state of the art. European Journal of Operational Research 190 (1), 1–21.

Alumur, S.A., Kara, B.Y., Karasan, O.E., 2009. The design of single allocation incomplete hub networks. Transportation Research Part B 43 (10), 936–951.

Arnold, P., Peeters, D., Thomas, I., 2004. Modelling a rail/road multimodal transportation system. Transportation Research Part E 40 (3), 255–270.

Balakrishnan, A., Graves, S.C., 1989. A composite algorithm for a concave-cost network flow problem. Networks 19 (2), 175–202.

Baumol, W.J., Panzar, J.C., Willig, R.D., 1988. Contestable markets and the theory of industry structure, second ed. Harcourt Brace Jovanovich, Inc., NY.

Boland, N., Krishnamoorth, M., Ernst, A.T., Ebery, J., 2004. Preprocessing and cutting for multiple allocation hub location problems. European Journal of Operational Research 155 (3), 638–653.

Bontekoning, Y.M., Macharis, C., Trip, J.J., 2004. Is a new applied transportation research field emerging? A review of intermodal rail–truck freight transport literature. Transportation Research Part A 38 (1), 1–34.

Campbell, J.F., 1994. Integer programming formulations of discrete hub location problems. European Journal of Operational Research 72 (2), 387–405.

Campbell, J.F., 1996. Hub location and the p-hub median problem. Operations Research 44 (6), 923–935.

Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2005a. Hub arc location problems: part I – introduction and results. Management Science 52 (10), 1540–1555.

Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2005b. Hub arc location problems: part II – formulations and optimal algorithms. Management Science 52 (10), 1556–1571.

Caves, D.W., Christensen, L.R., Tretheway, M.W., 1980. Flexible cost functions for multiproduct firms. Review of Economics and Statistics 62 (3), 477–481.

Caves, D.W., Christensen, L.R., Swanson, J.A., 1981. Productivity growth, scale economies, and capacity utilization in U.S. railroads, 1955–74. The American Economic Review 71 (5), 994–1002.

Contreras, I., Fernández, E., Marín, A., 2010. The tree of hubs location problem. European Journal of Operations Research 202 (2), 390–400.

Crainic, T.G., Kim, K.H., 2007. Multimodal transportation. In: Barnhart, C., Laporte, G. (Eds.), Transportation. Handbooks in Operations Research and Management Science, vol. 12. North-Holland, Amsterdam, pp. 467–532.

Dafermos, S., 1983. An iterative scheme for variational inequalities. Mathematical Programming 26 (1), 40–47.

Daganzo, C.F., 1983. Stochastic network equilibrium with multiple vehicle types and asymmetric, indefinite link cost Jacobians. Transportation Science 17 (3), 282–300.

Ebery, J., Krishnamoorthy, M., Ernst, A., Boland, N., 2000. The capacitated multiple allocation hub location problem: formulations and algorithms. European Journal of Operational Research 120 (3), 614–631.

Ernst, A.T., Krishnamoorthy, M., 1998. Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. European Journal of Operational Research 104 (1), 100–112.

Fernádez, J.E., De Cea, J., Soto, A., 2003. A multi-model supply–demand equilibrium model for predicting intercity freight flows. Transportation Research Part B 37 (7), 615–640.

Friesz, T.L., Holguin-Veras, J., 2005. Dynamic game theoretic models of urban freight: formulation and solution approach. In: Reggiani, A., Schintler, L.A. (Eds.), Methods and Models in Transportation and Telecommunications. Springer-Verlag, New York, pp. 143–162.

Fuss, M.A., Waverman, L., 1981. The regulation of telecommunications in Canada. Report to the Economic Council of Canada, No. 9. March, 1981.

Ge, Y.E., Zhang, H.M., William, H.K.L., 2003. Network reserve capacity under influence of traveler information. Journal of Transportation Engineering 129 (3), 262–270.

Gen, M., Cheng, R., 1997. Genetic Algorithms and Engineering Design. John Wily and Sons, New York.

Guélat, H., Florian, M., Crainic, T.G., 1990. A multimode and multiproduct network assignment model for strategic planning of freight flow. Transportation Science 24 (1), 25–39.

Holguín-Veras, J., Xu, N., de Jong, G., Maurer, H., 2009. An experimental economics investigation of shipper–carrier interactions on the choice of mode and shipment size in freight transport. Networks and Spatial Economics. doi:10.1007/s11067-009-9107-x.

Horner, M., O'Kelly, M.E., 2001. Embedding economy of scale concepts for hub network design. Journal of Transportation Geography 9 (4), 255–265.

Johnson, D.S., Lenstra, J.K., Rinnoy Kan, A.H.G., 1978. The complexity of the network design problem. Networks 8 (4), 279–285.

Klincewicz, J.G., 1998. Hub location in backbone/tributary network. Location Science 6 (1), 307–335.

LeBlanc, L.J., Morlok, E.K., Pierskalla, W.P., 1975. An efficient approach to solving the road network equilibrium traffic assignment problem. Transportation Research 9 (5), 309–318.

Limbourg, S., Jourquin, B., 2007. Rail–road terminal locations: aggregation errors and best potential locations on large networks. European Journal of Transportation and Infrastructure Research 7 (4), 317–334.

Loureiro, C.F.G., Balston, B., 1996. Investment selection model for multicontainer multimodal transportation networks. Transportation Research Record 1522, 38–46.

Macharis, C., Bontekoning, Y.M., 2004. Opportunities for OR in intermodal freight transport research: a review. European Journal of Operational Research 153 (2), 400–416.

Meng, Q., Wang, X., 2010. Utility-based probabilistic port hinterland estimation for intermodal freight transportation networks. In: Transportation Research Board 89th Annual Meeting Compendium of Papers DVD, Washington, DC, January 10–14, 2010.

Meng, Q., Yang, H., Bell, M.G.H., 2001. An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. Transportation Research Part B 35 (11), 83–105.

Nickel, S., Schöbel, A., Sonneborn, T., 2001. Hub location problems in urban traffic networks. In: Niittymaki, J., Pursula, M. (Eds.), Mathematics Methods and Optimization in Transportation Systems. Kluwer Academic Publishers.

Norman, G., 1979. Economies of Scale, Transportation Costs, and Location. Springer, New York.

O'Kelly, M.E., 1987. A quadratic integer program for the location of interacting hub facilities. European Journal of Operational Research 32 (3), 393–404.

O'Kelly, M.E., Bryan, D., 1998. Hub location with flow economies of scale. Transportation Research Part B 32 (8), 605–616.

Podnar, H., Skorin-Kapov, J., Skorin-Kapov, D., 2002. Network cost minimization using threshold-based discounting. European Journal of Operational Research 137 (2), 371–386.

Racunica, I., Wynter, L., 2005. Optimal location of multimodal freight hubs. Transportation Research Part B 39 (5), 453–477.

Skorin-Kapov, D., Skorin-Kapov, J., O'Kelly, M.E., 1996. Tight linear programming relaxations of uncapacitated p-hub median problems. European Journal of Operational Research 94 (3), 582–593.

Suh, S., Park, C.-H., Kim, T.J., 1990. A highway capacity function in Korea: measurement and calibration. Transportation Research Part A 24 (3), 177–186.

Transportation Research Board (TRB), 2000. Highway Capacity Manual 2000. <http://144.171.11.107/Main/Blurbs/Highway_Capacity_Manual_2000_152169.aspx> (accessed 06.07.10).

United Nations Conference on Trade and Development (UNCTAD), 2008. Review of Maritime Transport. <http://www.unctad.org/en/docs/rmt2008_en.pdf> (accessed 29.05.10).

United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP), 1999. Development of Trans-Asian Railway: Trans-Asian Railway in the Southern Corridor of Asian-Europe Routes. <http://www.unescap.org/ttdw/common/tis/tar/ncorridor.asp> (accessed 06.04.10).

United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP), 2003. Asian Highway Handbook. <www.unescap.org/ttdw/Publications/TIS_pubs/pub_2303/Full%20version.pdf> (accessed 06.02.10).

United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP), 2009. About the Trans-Asian Railway. <www.unescap.org/ttdw/index.asp?MenuName =TheTrans-AsianRailway> (accessed 06.2010).
Wang, S.E., Liao, C.H., 2006. Cost structure and productivity growth of the Taiwan Railway. Transportation Research Part E 42 (4), 317–339.
Wang, X., Meng, Q., Miao, L., Fwa, T.F., 2009. Impact of landbridge on port market area: model development and scenario analysis. Transportation Research Record 2097, 78–87.
Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proceeding Institution Civil Engineers Part II 1 (2), 325–378.
Winston, C., 1983. The demand for freight transportation: models and applications. Transportation Research Part A 17 (6), 419–427.
Yamada, T., Russ, B.F., Castro, J., Taniguchi, E., 2009. Designing multimodal freight transportation networks: a heuristic approach and applications. Transportation Science 43 (2), 129–143.
Yang, H., Bell, M.G.H., 1998. Models and algorithms for road network design: a review and some new developments. Transportation Reviews 18 (3), 257–278.
Yoon, M.-G., Current, J., 2008. The hub location and network design problem with fixed and variable arc costs: formulation and dual-based solution heuristics. Journal of the Operational Research Society 59 (1), 80–99.