

Primljen / Received: 5.11.2012.

Ispravljen / Corrected: 18.2.2013.

Prihvaćen / Accepted: 21.2.2013.

Dostupno online / Available online: 10.4.2013.

# A bilevel model for transit vehicle allocation

## Authors:



**Krunoslav Perić**, PhD. CE  
Institute IGH d.d.  
[krunoslav.peric@igh.hr](mailto:krunoslav.peric@igh.hr)



**Maria Boilé**, PhD. CE  
Hellenic Institute of Transport  
Thessaloniki, Greece  
[boile@certh.gr](mailto:boile@certh.gr)

Preliminary note

**Krunoslav Perić, Maria Boilé**

## A bilevel model for transit vehicle allocation

The formulation and resolution of a bi-level transit network design problem in an intermodal network environment is presented in the paper. The lower level of this bi-level problem is a combined mode-choice/assignment intermodal network equilibrium with asymmetric link interactions and variable transit frequencies. The upper level is a transit ridership maximization problem, with the number of transit vehicles allocated toward transit routes as a design variable. A sensitivity based algorithm is proposed for resolution of the bi-level problem.

### Key words:

transport planning, intermodal transport, transit networks, bi-level model

Prethodno priopćenje

**Krunoslav Perić, Maria Boilé**

## Dvorazinski model za raspodjelu tranzitnih vozila

Ovaj rad predstavlja formuliranje i rješenje projektantskih problema za dvorazinske tranzitne mreže u intermodalnom mrežnom okruženju. Niža razina dvorazinskog problema sastoji se od kombiniranog izbora i raspoređivanja intermodalne mrežne ravnoteže s asimetričnim interakcijskim povezivanjem i promjenjivim tranzitnim frekvencijama. Gornja razina je problem maksimizacije broja putnika koji koriste određeni oblik javnog prijevoza s brojem tranzitnih vozila dodijeljenih prema tranzitnim pravcima kao projektnih varijabli. Kao rješenje za dvorazinski problem predlaže se algoritam temeljen na osjetljivosti.

### Ključne riječi:

planiranje prijevoza, intermodalnost, projektno rješenje tranzitne mreže, dvorazinski model, SAB algoritam

Vorherige Mitteilung

**Krunoslav Perić, Maria Boilé**

## Zweistufenmodell für die Verteilung von Transitverkehrsmitteln

Die vorliegende Arbeit befasst sich mit der Formulierung und Lösung von Aufgabenstellungen im Entwurf zweistufiger Transitnetze in intermodalen Netzumgebungen. Die niedrigere Stufe dieses Zweistufenproblems besteht in der kombinierten Auswahl und Verteilung des intermodalen Netzgleichgewichts mit asymmetrischen Interaktionsverbindungen und variablen Transitfrequenzen. Die obere Stufe des Problems bezieht sich auf die Maximierung der Anzahl von Fahrgästen, die eine bestimmte Form öffentlichen Verkehrs nutzen, so dass die Anzahl der entsprechenden Transitfahrzeuge den Transitrichtungen als Entwurfsvariable zugeteilt wird. Als Lösung des Zweistufenproblems wird ein auf Empfindlichkeit beruhender Algorithmus vorgeschlagen.

### Schlüsselwörter:

Transportplanung, intermodaler Transport, Transitnetze, Zweistufenmodell

## 1. Introduction

Network design deals with modifications of transport networks that seek to optimize some objective function subject to equilibrium constraints. These modifications can be of operational or infrastructural nature. Operational changes include optimizing traffic signal settings, calculating toll pattern, and changing transit settings, while infrastructural changes include adding new lanes and expanding roadway capacities, introducing new highway lanes or rail lines, and other changes to the characteristics of the transport system. Some indicative examples of network design applications include the following. LeBlanc and Boyce [1] and Chiou [2] determined optimal link capacity improvements. Meng, Yang, and Bell [3] proposed a solution to the bi-level continuous network design problem. Yang and Lam [4], Yang and Bell [5], and Brotcorne, Labbe, Marcotte, and Savard [6], dealt with optimal toll finding. Clegg, Smith, Xiang, and Yarrow [7] developed a bi-level model based on a tool used to simultaneously find the toll pattern and signal timings in a transport network. Guan, Yang and Wirasinghe [8] presented a transit line configuration and passenger line assignment model formulated as a linear binary integer program. Yang and Yagar's [9] bi-level model optimizes signal splits in saturated networks. Mauttone and Urquhart [10] presented a solution algorithm for the transit network design problem that produces a set of transit routes under demand covering constraints. Zhao and Zeng [11] developed a method to optimize transit routes, headways, and vehicle scheduling. Marin and Jaramillo [12] proposed a model for capacity expansion at a rapid transit network as a mixed integer programming model. Guihaire and Hao [13] gave a global review of the most important strategic and tactical steps of transit network design and scheduling, classifying 69 different approaches dealing with the design, frequency setting, and timetabling of transit lines and their combinations.

The changes or improvements considered in this paper seek to optimize transit service within the context of an intermodal network and are constrained by available resources (i.e. number of transit vehicles). Based on the response of the users to these changes, the efficiency of each of the proposed improvements is evaluated, and the best improvement strategy is further developed. This framework answers the question of how to use available transit resources in the most efficient way.

The problem is formulated using a bi-level programming approach. This bi-level approach is based on sensitivity analysis, similar to the one implemented in [4, 9, 14, 15]. Yang and Yagar [9] and Yang and Lam [4] sought to find optimal signal timings and tolls, respectively, whereas Noriega and Florian [14] and Gao, Sun, and Shan [15] focused on optimizing transit frequencies. The latter ones assumed that cost interactions are asymmetric, which is an assumption also considered in the proposed formulation.

The lower-level problem in [15] is the transit network equilibrium, defined as a variational inequality. The formulated lower-level problem proposed herein is defined as an intermodal network equilibrium with asymmetric cost functions, also defined as a variational inequality. The importance of this contribution lies in the realization of the interaction between modes that can either compete or complement each other for a specific service. As a result the change of service level in one mode has an impact on the demand, travel pattern and the service level of the other modes. Changes in the demand create a different travel pattern requiring different optimal system characteristics. Hence, to determine the effect of the changes in the network service and to find the optimal improvements, the optimization has to be accomplished within the overall transport network context.

Several authors' work on the combined modal split and assignment model in a multimodal network is noted. Of greatest interest to this study are the ones allowing for combined-mode (intermodal) paths, such as [16, 17, 18].

Fernandez, De Cea, Florian, and Cabrera [16] and Boile, Spasovic, and Bladikas [17] focused on symmetric cost interactions, whereas Wu and Lam [18] assumed asymmetric cost interactions and their model was formulated as a variational inequality, making it similar to the approach adopted here. Wu and Lam stated the problem in terms of path flows, whereas the problem suggested herein is stated in terms of link flows similar to [19, 20]. Because the intermodal network setting adopted here includes combined-mode paths and several transit modes, the network equilibrium formulation considers inverted demand functions for the mode split at every level, as with the problem suggested by Boile [21]. The difference is that more transit modes are considered here, including bus sharing the same roadway capacity with passenger cars. Asymmetric cost interactions are assumed, which makes the problem applicable to more complex and generalized situations.

The transit network representation proposed by [22] is adopted in this paper because it considers congestion on transit stops, is easy to implement, and does not require path enumeration. Transit link cost representation and assignment throughout the transit network adopted here is similar to the one suggested by Lam, Zhou, and Sheng [23], which is based on the formulation by DeCea and Fernandez [22]. The difference with the latter is that transit congestion is represented more realistically accounting for the delay caused by boarding and alighting. This approach is upgraded in this study by including bus travel time dependency on auto traffic over the same part of the highway. This contribution is very important because it takes into account the effect of car traffic on the bus-mode travel frequencies and cost. Dwell time functions included in the transit cost functions are adopted from [24].

The impact of transit-vehicle flow on auto costs is modelled by initiating the bus/auto conversion factor that converts

buses into the equivalent number of cars on the highway links. The proposed model accounts for both, car flow impact on the transit cost and transit vehicle flows on the cost of the car mode. This consideration, although it makes the model more realistic, it also makes it substantially more complex, especially in finding equilibrium and performing sensitivity analysis.

The objective function in the upper-level model is the transit-ridership-maximizing function, which includes total travel cost on all transit links. Based on the resulting proposed transit improvements the bi-level problem estimates the new equilibrium flows. A restricted sensitivity-analysis approach similar to that proposed by Friesz and Tobin [25] is adopted here to formulate and solve the bi-level problem.

The scientific merit of this paper lies in that it fills a gap in the intermodal network modelling literature, by considering asymmetric cost interactions and transit delays caused by congestion, in optimizing allocation of transit vehicles to transit routes such as to maximize transit ridership. From a practical point of view, the model presented herein allows planners to evaluate both the direct and the cross-effect of various pricing and operating policies, capturing the interrelations among various modes.

Section 2 presents the mode choice formulation. In Section 3, definition of the intermodal network and its parts is presented. Section 4 presents the bi-level model formulation and solution algorithm. Section 5 presents the results of model application on a medium-sized network. Section 6 concludes the paper.

## 2. Mode choice

Five modes are considered here, auto, bus, rail and their combinations (auto-bus and auto-rail), which are typically used along commuter corridors or along intercity corridors, to which this model mostly applies. We assume that when a traveller decides on a mode, he or she first chooses between transit and auto. If the transit mode is chosen, the traveller has the option to travel by bus or rail. When this decision has been made, a traveller can opt to start the trip by car or to directly access transit on foot. Every transit trip ends by bus or rail mode. Utility functions are used to determine the demand for each mode.

Utility for model  $M$  is defined as:

$$U_M^w = -\beta * Cost - \lambda_M \quad (1)$$

where:

- $\beta$  - coefficient estimated from the data
- $\lambda_M$  - indicates preference toward mode  $M$
- $Cost$  - s the total travel time cost.

$\lambda_M$  and  $\beta$  - are parameters that are calibrated based on observed data to adjust the model-observed behaviour regarding the choice process represented by the demand function.

Each choice (mode, transit option, and access type) is modelled by the logit demand function presented below:

$$g_c^w = g^w \frac{e^{U_c^w}}{\sum_{c'} e^{U_{c'}^w}} \quad (2)$$

where  $U_c^w$  is the utility of a choice  $c$  between an OD pair  $w$ , and  $c'$  is the set of all modes (choices) available at that OD. The ratio of the utility exponential to the sum of utility exponentials of all available choices gives the proportion of the demand selecting choice  $c$ . The nested logit model is the specific demand model implemented in this study. This model uses functions in the form presented above to determine the split of the total demand among all modes and travel options.

## 3. Modeling of transit network

### 3.1. Network definition

The network is represented by  $G(N, A)$ , where  $N$  is the set of nodes and  $A$  is the set of directed arcs  $A = PL \cup TL \cup WK \cup TRL$ .  $PL$  is the set of physical links, including highway  $HL$ , and links representing rail tracks  $RTL$ . The flow is not assigned over  $RTL$  links, but their impedances are used to obtain travel time over rail links.  $TL$  is the set of transit links which is derived from the transit network representation and which is divided into rail  $TLR$  and bus  $TLB$  transit links so that  $TL = TLR \cup TLB$ .  $WK = WKA \cup WKE$  is the set of walk access  $WKA$  and egress  $WKE$  links which are further subdivided into rail and bus access, so that  $WKA = WKAR \cup WKAB$  and  $WKE = WKER \cup WKEB$  is the set of transfer links, which includes transfer links to rail  $TRLR$  and transfer links to bus  $TRLB$ .  $W$  is the set of all OD pairs. Every link is associated with two nodes. If a node is the starting point of a link, the corresponding value in the link/node incidence matrix is set to 1. If a node is the end point of a link the value is set to -1. If a node is not associated with that link, the value is set to 0. This incidence matrix definition is used in the shortest path (Dijkstra's Algorithm) and the transit boarding and alighting time calculations.

### 3.2. Definition of subnetworks

The auto network is defined as a set of highway links  $HL$  considering that the "pure" auto mode uses only highway links. The "pure" bus mode subnetwork is defined as a set of bus walking access links, bus links and bus walking egress links  $TLB \cup WKAB \cup WKEB$ , considering that travellers walk to a bus station, board on transit vehicles which run over bus transit links and reach their destination through bus egress links. Similarly, the "pure" rail mode subnetwork is defined as a set of rail walking access links, rail links and rail walking egress links  $TLR \cup WKER \cup WKAR$ .

An "intermodal trip" is defined as a trip that contains more than one mode of travel. In this paper, intermodal trips start by auto and end with one of the two transit modes, bus or rail. The intermodal trip contains highway links, transfer links, transit links and egress links. Since, for the definitions used herein, the intermodal trips end by a transit mode, it was necessary to exclude highway links connected to the destination nodes. For every origin-destination pair **W**, an intermodal bus or rail subnetwork is thus defined as a set of bus or rail transfer, transit and egress links and highway links excluding those that are connected to a destination node (link-node incidence matrix element is -1). Highway links excluded for a certain origin-destination pair may be included in the intermodal subnetwork for another pair, given that they do not have the same destination.

### 3.3. Transit network representation

The basic concepts for transit network representation are as follows.

**A line** is a group of transit vehicles that runs from one node to another between two terminal nodes, the origin and the destination, and back through the transport network. The size, capacity, and operating characteristics of vehicles on the same line are typically considered to be identical, and the vehicles travel through the transit network using the same sequence of network links and nodes, which are addressed as itineraries.

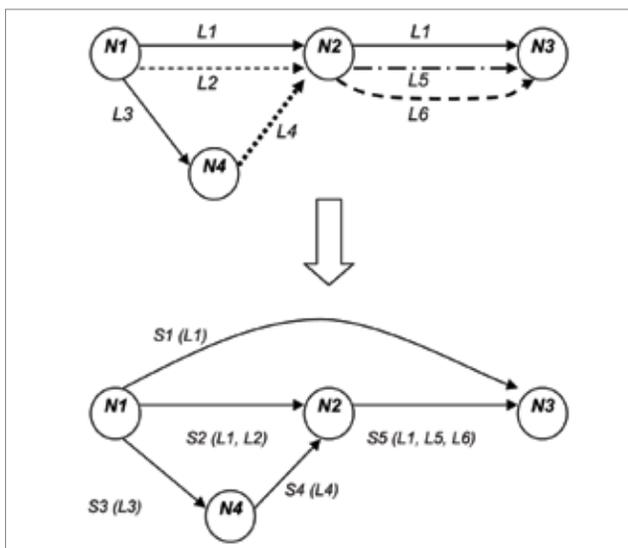


Figure 1. Transit network representation

**A transit route** is any path that a transit user can track on the transit network to reach a destination from an origin, generally represented by the origin and destination nodes, as well as transfer points.

**A transit link** is defined as the part of a route between two consecutive transfer points. Such links are associated with the set of transit lines that run between those two transfer nodes. A transit network example is presented in Figure 1.

L represents a transit line, while S represents a transit link with an associated set of transit lines contained within. All the transit lines that take a traveller from one point to the other point are contained in the corresponding transit link. If a transit line runs from N1 to N3 without stopping at N2, the line would be included in S1 but not in S2 and S5. Keeping track of transit lines included in transit links is important because of the calculation of waiting time for the transit link's costs.

### 3.4. Cost of transit network

On highway links, the cost depends on the flow of both private vehicles and transit vehicles. The flow of transit vehicles on the highway links is translated into car flow as a sum of frequencies of all transit lines running over that link, multiplied by a car-bus equivalency factor (bceq). The highway link cost function used herein is defined in terms of car vehicle flow. The total flow over a highway link is defined as:

$$v_{hl} = \sum_w \left( \sum_{p \in P_{tm,i}^W} \delta_{lp} h_p + \sum_{p \in P_{tm,i}^W} \delta_{lp} h_p \right) + \sum_{ptl} o_{ptl} bceq f_{ptl} \tag{3}$$

where  $v_{hl}$  is the flow over highway link  $hl$ ,  $\delta_{lp}$  is an element of the link/path incidence matrix,  $h_p$  is the flow on path  $p$  and  $o_{ptl}$  is the element of the link/transit line incidence matrix and takes the value of 1 if the link is contained in the public transit line and 0 otherwise.  $f_{ptl}$  is the frequency on the public transit line  $ptl$ .

The highway link cost function is of the form::

$$c_{hl}(v) = t_{hl0} \left( 1 + \alpha \left( \frac{v}{k_{hl}} \right)^\beta \right) + gas_{hl} / vot \tag{4}$$

A transit link flow is the sum (over all  $w$ , transit modes  $tm$  and access types  $at$ ) of flow on pure transit and intermodal paths using that link.

$$v_{tl} = \sum_w \sum_{tm} \sum_{at} \sum_{p \in P_{tm,at}^W} \delta_{lp} h_p \tag{5}$$

The total cost for traversing a transit link is the sum of in-vehicle time and waiting time associated with that link:

$$c_{tl}(v_{tl}) = tt_{tl} + w_{tl}(v_{tl}) + fare_{tl} / vot - \text{for a rail link} \tag{6}$$

$$c_{tl}(v) = tt_{tl}(v) + w_{tl}(v) + fare_{tl} / vot - \text{for a bus link} \tag{7}$$

It is safe to assume that the in-vehicle time ( $tt_{tl}$ ) for a rail link is constant, since it has its own right-of-way, and it is not a function of flow. On the contrary, most of the bus lines do

not have exclusive right-of-way, therefore, they are subject to highway congestion and are highway flow dependent. The waiting time depends on the frequencies of the set of attractive transit lines running over that link.

$$w_{ij} = \frac{\rho}{\sum_{ptl \in A_{ij}} f_{ptl}} \quad (8)$$

where  $\rho$  is a wait time factor. A value of  $\rho=1$  corresponds to an exponential distribution of the interarrival times of the vehicles. A value  $\rho=0,5$  corresponds to a uniform distribution of the interarrival times. For simplicity, it is assumed that  $\rho=1$ .

The frequency of a transit line is defined as a number of transit vehicles dedicated to that transit line divided by the total journey time:

$$f_{ptl} = \frac{ntv_{ptl}}{T_{ptl}(v)} \quad (9)$$

The total journey time for line  $ptl$  is the sum of time spent at a terminal  $t_o$  plus the travel time over all links used by that transit line and dwell time at all the transit stops served by the transit line, which is a function of the number of boarding/alighting passengers:

$$T_{ptl}(v) = t_o + \sum_{pl \in ptl} tt_{pl}^{ptl} + \sum_{n \in tl} dt_n^{ptl}(v_{ij}) \quad (10)$$

Dwell time for line  $ptl$  at node  $n$  is a function of the total number of passengers alighting/boarding line  $ptl$  at node  $n$ :

$$dt_n^{ptl} = f(al_n^{ptl}, bo_n^{ptl}) \quad (11)$$

Number of travellers alighting/boarding line  $ptl$  at node  $n$ :

$$al_n^{ptl} = \sum_s \varphi_{ntl}^- \xi_{ptltl} X_{ij}^{ptl} v_{ij} \quad (12)$$

$$bo_n^{ptl} = \sum_s \varphi_{ntl}^+ \xi_{ptltl} X_{ij}^{ptl} v_{ij} \quad (13)$$

$$\varphi_{ntl}^- = \begin{cases} 1 & n = i^-(s) \\ 0 & n \neq i^-(s) \end{cases} \quad (14)$$

$$\varphi_{ntl}^+ = \begin{cases} 1 & n = i^+(s) \\ 0 & n \neq i^+(s) \end{cases} \quad (15)$$

where  $\varphi_{ntl}^-$  has the value of 1 if node  $s$  is the ending node of the transit link  $tl$ , and  $\varphi_{ntl}^+$  has the value of 1 if node  $s$  is the starting node of the transit link  $tl$ .  $\xi_{ptltl}$  is an element of the transit line/transit link incidence matrix.

The proportion of travellers on transit link  $tl$  that will board line  $ptl$  is in direct proportion to line frequency in the set of attractive lines on the transit links:

$$X_{ij}^{ptl} = \frac{f_{ptl}}{\sum_{l \in A_{ij}} f_l} \quad (16)$$

Note that according to the definition of a transit link and the corresponding transit lines used herein, the traveller will be indifferent as to which line within a link to choose, as all lines will serve the same trip purpose.

The total number of passengers boarding each line on the transit link  $tl$  is obtained by multiplying the proportion of the frequency of the transit line  $ptl$  in the total combined frequency of all attractive lines on that link  $tl(A_{ij})$ , with the total number of travellers  $v_{ij}$  using that link. The set of attractive lines is defined as a set of lines that solves the minimization problem [22]:

$$\min_{\{n_{ptl}\}} \frac{1 + \sum_{ptl=1}^k tt_{tl} \cdot f_{ptl} \cdot n_{ptl}}{\sum_{ptl=1}^k f_{ptl} \cdot n_{ptl}} \quad (17)$$

subject to:  $n_{ptl}$  is 0 ili 1, where if  $n_{ptl}=1$  then  $n_{ptl} \in A_{ij}$ .

**Notations used in equations (3) to (17)**

$c(v)$	- cost of traversing a link at flow level $v$
$G_1^{-1}$	- inverse demand function for total demand (between auto and transit)
$G_2^{-1}$	- inverse demand function for transit demand (between rail and bus)
$G_3^{-1}$	- inverse demand function for rail demand (choose access type- single or combined rail mode)
$G_4^{-1}$	- inverse demand function for bus demand (choose access type- single or combined bus mode)
$g^w$	- total demand between OD pair $w$
$g_a^w$	- auto demand between OD pair $w$
$g_t^w$	- transit demand between OD pair $w$
$g_{t,b}^w$	- bus transit demand between OD pair $w$
$g_{t,r}^w$	- rail transit demand between OD pair $w$
$g_{t,tm,i}^w$	- transit demand with auto access to transit between OD pair $w$ (intermodal)
$g_{t,tm,p}^w$	- transit demand with walk access to transit between OD pair $w$ (pure)
$v_{hl}$	- flow on highway link $l$
$v_{tl}$	- flow on transit link $tl$
$\delta_{lp}$	- element of the link/path incidence matrix
$O_{ ptl}$	- element of the link/transit-line incidence matrix
$bceq$	- bus-car equivalency factor
$vot$	- value of time
$f_{ptl}$	- frequency of transit line $ptl$
$fare_{tl}$	- fare on a transit link $tl$
$c_{hl}(v)$	- highway link cost

- $c_{ij}(v)$  - transit link cost
- $t_{nlo}$  - free-flow time at link  $l$
- $\lambda, \beta$  - highway link cost coefficients
- $k_{hl}$  - capacity of link  $l$
- $T_{ptl}$  - total journey time for line  $ptl$
- $t_{tl}^{ptl}$  - traverse time over transit link  $tl$  contained in the transit line  $ptl$
- $dt_n^{ptl}$  - dwell time for transit line  $ptl$  at the stop  $n$
- $t_o$  - time transit vehicle spends in the terminal
- $\eta_{mtl}$  - element of transit-section/transit-link incidence matrix
- $x_{tl}^{ptl}$  - proportion of the flow at transit link  $tl$  that will board transit line  $ptl$
- $ntv_{ptl}$  - number of vehicles dedicated to transit line  $ptl$
- $al_n^{ptl}$  - number of passengers alighting line  $ptl$  at stop  $n$
- $bo_n^{ptl}$  - number of passengers boarding line  $ptl$  at stop  $n$
- $\Phi_{notl}^+$  - a 0-1 parameter, equal to 1 if node  $n$  is starting node of transit link  $tl$
- $\Phi_{nptl}^-$  - a 0-1 parameter, equal to 1 if node  $n$  is ending node of transit link  $tl$
- $\xi_{\geq ptl tl}$  - element of the transit-line/transit-link incidence matrix
- $w_{tl}$  - vwaiting time for transit link  $tl$
- $\rho$  - constant (1 if arrival distribution is exponential, 0.5 if arrival distribution is uniform)
- $tt_{tl}$  - travel time over transit link  $tl$
- $h_a^{wp}$  - flow on path  $p$  between OD pair  $w$  on the auto mode network at equilibrium
- $h_{t,tm,at}^{wp}$  - flow on path  $p$  between OD pair  $w$  on the transit network  $t$  mode  $tm$  and access type  $at$  at equilibrium
- $W$  - set of paths of mode  $m$  on OD pair  $w$
- $P_m^w$  - set of paths of mode  $m$  on OD pair  $w$
- $U_m^{wp}$  - perceived utility of traversing a path at equilibrium using mode  $m$  between origin-destination pair  $w$
- $v^n$  - vector of link flows at the  $n$ -th iteration
- $Z_a^n$  - pauxiliary demand of mode  $m$  at the  $n$ -th iteration
- $y^n$  - vector of auxiliary link flows at the  $n$ -th iteration
- $A_{tl}$  - set of attractive lines
- $c$  - vector of link cost functions
- $G^l$  - vector of inverse demand functions with dimension  $w$
- $v$  - vector of flows at links in the intermodal network
- $v^*$  - vector of equilibrium flows at links in the intermodal network
- $g$  - vector of demand between origin and destination in the intermodal network with dimension  $w$
- $g^*$  - vector of equilibrium modal demands between origin and destination in the intermodal network with dimension  $w$
- $\beta_i$  - parameter of the utility function for the choice  $i$
- $C_m$  - constant in the  $e$  value calculation for mode  $m$ .

### 4. Bi-level model

A general form of a bi-level model is:

$$(U) \max_x F(x, y(x)) \tag{18}$$

subject to:  $G(x) \leq 0$ ,

Where  $y(x)$  is defined:

$$(L) \min_y f(x, y) \tag{19}$$

subject to:  $g(x, y) \leq 0$ .

The upper level (U) (18) represents the objective function of the transit system managers and  $x$  is the decision vector. The lower level (L) (19) represents network users decisions, modelling the choices of the travellers, and determining the travel pattern throughout the network. It is a combined mode split/ assignment model over an intermodal network. The upper level model seeks transit settings that maximize bus fare revenue. The cost interactions are asymmetric and do not have equivalent optimization formulation. Therefore, the problem is expressed in the form of a variational inequality. The upper level formulates the objective that a network planner wishes to achieve in terms of providing optimal transit coverage with a certain number of vehicles available, which in this case aims to maximize the profit of bus lines. Taking into account the characteristics of the intermodal network, the formulation of a bi-level model of continuous transit network design is:

#### Upper level

$$\max \sum_{bptl} v_{bptl} \text{ fare}_{bptl} \tag{20}$$

subject to:

$$\sum_{bptl} ntv_{ptl} \leq tn_{bptl} \text{ } U1$$

$$\sum_{rptl} ntv_{ptl} \leq tn_{rptl} \text{ } U2$$

$$ntv_{bptl} + ntv_{rptl} \geq 1 \text{ } U3$$

#### Lower level

$$c(v^*)^T (v - v^*) - G_1^T(g, g^*) - G_2^T(g_{t,b}, g_{t,b}^*) - G_3^T(g_{t,r,p}, g_{t,r,p}^*) - G_4^T(g_{t,b,p}, g_{t,b,p}^*) \geq 0, \tag{21}$$

subject to for every feasible  $v$  and  $g$ :

$$g^w = g_t^w + g_a^w \tag{1} \quad g_{t,b,i}^w = \sum_p h_{t,b,i}^{wp} \tag{6}$$

$$g_t^w = g_{t,r}^w + g_{t,b}^w \tag{2} \quad g_{t,r,p}^w = \sum_p h_{t,r,p}^{wp} \tag{7}$$

$$g_{t,r}^w = g_{t,r,i}^w + g_{t,r,p}^w \tag{3} \quad g_{t,r,i}^w = \sum_p h_{t,r,i}^{wp} \tag{8}$$

$$g_{t,b}^w = g_{t,b,i}^w + g_{t,b,p}^w \tag{4}$$

$$g_{t,b,p}^w = \sum_p h_{t,b,p}^{wp} \tag{5} \quad g_a^w = \sum_p h_a^{wp} \tag{9}$$

$$h_{t,tm,at}^{wp}, h_a^{wp} \geq 0$$

$$\forall w \in W$$

$\sum_{opt} v_{bptl} \text{ fare}_{bptl}$  is the upper level objective function. It calculates total bus revenue in the network.

The constraint U1 states that the sum of transit vehicles assigned to bus transit lines is less than or equal to the total number of bus transit vehicles available to a transit agency.

The constraint U2 states that the sum of transit vehicles assigned to rail transit lines is less than or equal to the total number of rail transit vehicles available to the transit agency.

The constraint U3 states that at least one transit vehicle has to be assigned to every public transit line (bus and rail). The objective function of the lower level model is the variational inequality formulation of the intermodal network combined mode choice/traffic assignment. The above objective function is subject to the following demand constraints:

According to demand constraint (1), the total demand  $g$  is divided between demand for transit and demand for auto. According to demand constraint (2), the transit demand  $g_t$  is divided between demand for bus and demand for rail. According to demand constraint (3), the total demand  $g_r$  is divided between demand for pure rail and demand for intermodal rail. According to demand constraint (4), the total demand  $g_b$  is divided between demand for pure bus and demand for intermodal bus. According to constraints (5) through (9) the sum of flows on all paths for a particular mode and OD pair must equal the demand for that mode on that OD pair.

#### 4.1. Solution algorithm of lower level

The diagonalization algorithm is adopted and modified to solve the lower level problem. The algorithm is as follows:

##### Diagonalization Algorithm

**Step 0:** Find the free-flow minimum paths in each subnetwork (auto, pure bus, pure rail, rail intermodal, bus intermodal). Find a flow in each subnetwork according to its minimum cost paths. Assign the demand obtained for auto on the auto paths and the demand obtained for transit to the corresponding intermodal and pure transit paths, resulting in the travel pattern  $v^n$ .

**Step 1:** Update travel times.

**Step 2:** Direction finding. Perform the all-or-nothing assignment to the path that will create the largest drop in the objective function, yielding to the  $v^n$  travel pattern and (sub) mode demands  $z_a^n$  and  $z_{t,tm,at}^n, \forall t, tm, at$ .

**Step 3:** Finding step size  $\alpha_n$ . involves solution to the diagonalized problem.

**Step 4:** Update link flows and modal demands.

**Step 5:** Check the convergence.

If the condition  $\max_j \{(v_i^{n+1} - v_i^n) / v_i^n\} \leq \epsilon$  is satisfied then the solution is  $v^{n+1}$ . Otherwise, go to Step 1.

The descent direction is determined based on the minimum e value in the following expressions:

- If e1 is the minimum among all e values, then the descent direction is the all-or-nothing assignment to the shortest path of the auto network.
- If e2 is the minimum e value, then the descent direction is the all-or-nothing assignment to the shortest path of the pure bus network, and so on. This has to hold for every origin-destination pair.

$$e1 = C_a + \frac{1}{\beta_1} (\ln g_a + 1) \tag{22}$$

$$e2 = C_{bp} + \frac{1}{\beta_4} \ln \frac{g_{bp}}{g_b} + \frac{1}{\beta_2} \ln \frac{g_b}{g_t} + \frac{1}{\beta_1} (\ln g_t + 1) \tag{23}$$

$$e3 = C_{br} + \frac{1}{\beta_4} \ln \frac{g_{br}}{g_b} + \frac{1}{\beta_2} \ln \frac{g_b}{g_t} + \frac{1}{\beta_1} (\ln g_t + 1) \tag{24}$$

$$e4 = C_{rp} + \frac{1}{\beta_3} \ln \frac{g_{rp}}{g_r} + \frac{1}{\beta_2} \ln \frac{g_r}{g_t} + \frac{1}{\beta_1} (\ln g_t + 1) \tag{25}$$

$$e5 = C_{ri} + \frac{1}{\beta_3} \ln \frac{g_{ri}}{g_r} + \frac{1}{\beta_2} \ln \frac{g_r}{g_t} + \frac{1}{\beta_1} (\ln g_t + 1) \tag{26}$$

#### 4.2. Solution algorithm for Bi-level model

The Sensitivity Analysis Based (SAB) algorithm [25, 26] was used to solve the bi-level transit optimization problem. Because the upper-level objective function is implicit and is a non-linear function of decision variable  $ntv_{ptl}$ , local linear approximations using Taylor's formula are implemented based on the derivatives of the reaction functions with respect to the number of transit vehicles. Thus, the derivative information was obtained by implementing sensitivity analysis for a given solution of the intermodal network equilibrium problem. The number of transit vehicles is used as the decision variable for transit information because it is the only element of the transit service that stays constant inside the equilibrium calculation, and therefore can be determined exogenously. On the contrary, transit frequency, which is usually used as the decision variable in transit optimization procedures, is elastic and subject to network congestion. Thus, it cannot be used as the decision variable. The SAB algorithm is presented below.

##### SAB Algorithm

**Step 0:** Determine an initial solution of number of vehicles assigned to transit lines  $ntv_{ptl}$

**Step 1:** Solve the lower-level intermodal network equilibrium problem for the given  $ntv_{ptl}$  using the diagonalization algorithm, and obtain  $v^n$ .

- Step 2:** Calculate derivatives  $\delta v^n / \delta ntv_{ptl}$  using the sensitivity analysis method.
- Step 3:** Formulate local linear approximations of the upper-level objective function and constraints using the derivative information, and solve the resulting problem to get an auxiliary solution  $y$ .
- Step 4:** Compute  $ntv_{ptl}^{n+1}$  on the interval  $[ntv_{ptl}^n, y]$  that maximizes the upper-level objective function.
- Step 5:** If  $|ntv_{ptl}^{n+1} - ntv_{ptl}^n| \leq \epsilon$  on all public transit lines  $ptl$  for predetermined tolerance  $\epsilon$ , then stop. Otherwise, set  $n := n + 1$  and go to Step 1.

At Step 1, the lower level intermodal equilibrium problem is solved using the diagonalization algorithm. As described in the previous chapter, the variational inequality representing network equilibrium is split into a series of minimization problems. As the final output, the algorithm generates information necessary for the sensitivity analysis, including the complete link flow pattern  $v^n$ , the subset of minimum cost paths between each OD pair and the derivative of link costs with respect to link flows and to the number of transit vehicles per transit line.

Using the information obtained in Step 1, the sensitivity analysis is performed at Step 2 to obtain the derivatives of equilibrium link flows with respect to the number of transit vehicles for a given solution to the intermodal network equilibrium problem. At Step 3, the information obtained from sensitivity analysis is used to formulate the linear approximations of the reaction functions and hence the upper-level objective function and constraints. The approximations will result in a mathematical programming problem to which the auxiliary solution  $y$  is in the farthest point in the descent direction of the upper-level objective function. At Step 4, the golden section method is used on the interval  $[ntv_{ptl}^n, y]$  to find the solution that gives the greatest increase in the upper-level objective function. Some authors have used a predetermined step that monotonically decreases with the number of iterations, because one-dimensional searches (the golden section method being one of them) require equilibrium assignments and can be computationally extensive. At Step 5, the convergence criteria are checked, and if satisfied, the solution is found. Otherwise, the algorithm returns to Step 1.

Since the upper-level objective function is pretty regular monotonic (or ditonic depending on the observed interval), the golden section search worked well and found solutions close to the global optimum. In addition, the number of equilibrium assignments to be solved did not differ significantly from the case when predetermined step size was used.

For more information about SAB algorithm see [25, 26].

## 5. Case study

### 5.1. Test network

The model presented in the previous section will be applied on a test network. The intermodal network to be analyzed is presented in Figure 1. This network is a realistic representation of a commuter corridor located between I-80 and I-78 in Morris, Ocean, and Union County, New Jersey. Its four origins are Morristown, Madison, Summit, and Maplewood, and the destination is Newark (Figures 2 and 3). It consists of 102 links and 52 nodes.



Figure 2. Test network



Figure 3. Schematic presentation of test network

The demand for the peak period from origins to destination is shown in Table 1.

Table 1. Origin-Destination demand

OD* Pair	Origin	Trips to destination (Newark)
1	Morristown	6550
2	Madison	605
3	Summit	310
4	Maplewood	330

\*OD = Origin-Destination

The purpose of this case study is to demonstrate the application of the model. It is assumed that 20 vehicles are available for 3 transit lines. The model should allocate them in order to maximize bus usage/profit on all paths and from all origins. At the start, it is assumed that buses are distributed among transit lines as follows in Table 2:

Table 2. Starting bus allocation

Transit Line	Number of buses
70	4
73	10
MCM3	6

## 5.2. Results

The solution suggested by the model is shown in Table 3. The model allocated buses from lines 70 and MCM3 to line 73.

Table 3. Model Solution

No.	Transit Line	Initial	Model Optimized
1	70	4	3,14
2	73	10	12,29
3	MCM3	6	4,57

That change in vehicle allocation produced increase in peak period revenue for those three lines from \$5854.4 to \$5931.2. The impacts of the changes on the bus ridership are presented in the following tables.

It can be seen from Table 4 that the total bus ridership has increased by 51.7 passengers or around 3%. Most of newly attracted passengers were from the most distant, from the destination, origins (origins 1 and 2), served by line 70. Thus, the largest number of vehicles has been allocated towards it, attracting mostly intermodal passengers. On the other hand, origins closer to the destination have lost bus passengers because of the decreased level-of-service.

Table 4. Total bus ridership between each OD pair

OD Pair No.	Initial	Optimized	Difference	Change [%]
1	1663,67	1722,43	58,76	3,5
2	156,95	162,15	5,20	3,3
3	96,54	90,68	-5,86	-6,1
4	107,35	100,94	-6,41	-6,0
Total	2024,51	2076,20	51,69	2,6

Tables 5 and 6 show how the travel (path) cost from origins to the destination has changed with changed transit settings for bus modes. Costs from OD pairs 1 and 2 have experienced decrease in cost for intermodal bus mode. This is expected since most of the bus demand comes from origins 1 and 2 using intermodal bus mode. In addition, pure bus mode was less affected for OD pairs 1 and 2, while passengers

from nodes 3 and 4 for all bus modes experienced a more significant decrease in the level-of-service because of the reduced number of buses on line that was serving them.

Table 5. Intermodal bus path cost from every origin (in minutes)

OD Pair No.	OD Pair	Initial	Optimized	Difference	Change [%]
1	(1,8)	83,88	79,05	-4,83	-5,76
2	(2,8)	87,93	83,15	-4,78	-5,43
3	(4,8)	59,83	66,25	6,42	10,73
4	(19,8)	53,09	59,49	6,40	12,05

Table 6. Pure bus path cost from every origin (in minutes)

OD Pair No.	OD Pair	Initial	Optimized	Difference	Change [%]
1	(1,8)	135,47	135,88	0,410	0,30
2	(2,8)	119,22	119,63	0,410	0,34
3	(4,8)	73,18	79,601	6,421	8,77
4	(19,8)	55,417	61,817	6,400	11,55

The average cost for each bus mode and bus mode in general are presented in Table 7. The intermodal bus mode experienced reduction in travel cost over the case study area, while an average pure bus mode travel time was higher than the previous one. An overall bus travel time improved by 3.3%.

Table 7. Bus travel cost over the case study area (in minutes)

Mode	Initial	Optimized	Difference	Change [%]
Pure bus	122,33	123,82	1,48	1,21%
Intermodal bus	85,23	81,47	-3,77	-4,42%
Total bus	95,78	92,61	-3,17	-3,31%

## 5.3. Effect on overall transport network

By optimizing bus service, the total network cost experienced reduction (Table 8), although it was not the primary goal of the case study. Taking travellers off the highway and attracting them to bus service was not beneficial only to the transit agency, but to the overall transport network.

Table 8. Total network travel cost (in minutes)

	Initial	Optimized	Difference	Change [%]
Total network cost	558,986	552,900	-6,096	-1,1

## 6. Conclusion

The work presented in this study achieves several objectives. A combined mode choice / assignment model to represent general intermodal commuter corridors was developed; a diagonalization algorithm for the solution of the network equilibrium was proposed; and descent direction for this

algorithm was derived. Second, a bi-level approach for finding optimal transit vehicle reallocation was developed and the SAB algorithm for this approach was suggested. Third, the approach was tested on a real-world network, which showed the advantages and justification for applying the proposed approach. Further research will be directed towards inclusion of different trip purposes, user classes and time of day travel.

## REFERENCES

- [1] Leblanc, L.J., Boyce, D. E.: A Bi-level Programming Algorithm for Exact Solution of the Network Design Problem with User-Optimal Flows, *Transportation Research*, vol. 20B, p. 259-265, 1986.
- [2] Chiou, S.W.: Bilevel programming for the continuous transport network design problem, *Transportation Research*, 39B, p. 361-383, 2005.
- [3] Meng, Q., Yang, H., Bell, M.G.H.: An Equivalent Continuously Differentiable Model and Locally Convergent Algorithm for the Continuous Network Design Problem, *Transportation Research*, vol. 35B, p. 83-105, 2001.
- [4] Yang, H., Lam, W.H.K.: Optimal Road Tolls under Conditions of Queueing and Congestion, *Transportation Research*, 30A, p. 319-332, 1996.
- [5] Yang, H., Bell, M.G.H.: Traffic Restraint, Road Pricing and Network Equilibrium, *Transportation Research*. 31B, p. 303-314, 1997.
- [6] Brotcorne, L., Labbe, M., Marcotte, P., Savard, G.: A Bilevel Model for Toll, 2001.
- [7] Clegg, J., Smith, M., Xiang, Y., Yarrow, R.: Bilevel programming applied to optimising urban transportation, *Transportation Research Part B* 35, p. 41-70, 2001.
- [8] Guan, J.F., Yang, H. and Wirasinghe, S.C.: Simultaneous optimization of transit line configuration and passenger line assignment, *Transportation Research*, vol. B 40, p. 885-902, 2006.
- [9] Yang, H. & Yagar, S.: Traffic Assignment and Signal Control in Saturated Road Networks, *Transportation Research*, 29A, p. 125-139, 1995.
- [10] Mauttone, A. and Urquhart, M.E.: A route set construction algorithm for the transit network design problem, *Computers & Operations Research*, vol. 36, p. 2440 - 2449, 2009.
- [11] Zhao, F. and Zeng, X.: Optimization of transit route network, vehicle headways and timetables for large-scale transit networks, *European Journal of Operational Research* vol. 186, p. 841-855, 2008.
- [12] Marín, Á. and Jaramillo, P.: Urban Rapid Transit Capacity Expansion, *European Journal of Operational Research*, no. 191, p. 45-60, 2008.
- [13] Guihaire, V. and Hao, J. K.: Transit network design and scheduling: A global review, *Transportation Research Part*, 42A, p. 1251-1273, 2008.
- [14] Noriega, Y., Florian, M.: L'Optimisation Des Frequencies D'un Reseau De Transport en Commun Moyennement Congestionne, *INFOR*, vol. 41, no. 2, 2003.
- [15] Gao, Z., Sun, H., and Shan, L.L., A Continuous Equilibrium Network Design Model and Algorithm for Transit Systems, *Transportation Research* 38B, p. 235-250, 2004.
- [16] Fernandez, E., De Cea, J., Florian, M., and Cabrera, E.: Network Equilibrium Models with Combined Modes, *Transportation Science*, vol. 28, no. 3, p.182-192, 1994.
- [17] Boile, M.P., Spasovic L.N., and Bladikas, A.K.: Modeling Intermodal Auto-Rail Commuter Networks, *Transportation Research Record*, no. 1516, p. 38-47, 1995.
- [18] Wu, Z. X., Lam, W.H.K.: A Combined Modal Split and Stochastic Assignment Model for Congested Networks with Motorized and Non-Motorized Transport Modes, *TRB 2003 Annual meeting*, 2003.
- [19] Florain, M., Speiss, H.: On Binary Mode Choice/Assignment Models, *Transportation Science*, vol. 17, no. 1, p. 32-47, 1983.
- [20] Dafermos, S.: Relaxation Algorithms for the General Asymmetric Traffic Equilibrium Problem, *Transportation Science*, vol. 16, no. 2, p. 231-240, 1982.
- [21] Boile, M.: *Intermodal Commuter Network Planning* Ph.D. Dissertation, New Jersey Institute of Technology, 1995.
- [22] De Cea, J., Fernandez, E.: Transit Assignment for Congested Public Transport Systems: An Equilibrium Model, *Transportation Science*, vol. 27, no. 2, 1993.
- [23] Lam, W.H.K., Zhou, J., and Sheng, Z.: A Capacity Restraint Transit Assignment with Elastic Line Frequency, *Transportation Research Part B*, p. 919-938, 2002.
- [24] Aashtiani, H.Z., Iravani, H.: Application of Dwell Time Functions in Transit Assignment Model, *Transportation Research Record*, no. 1817, p. 88-92, 2002.
- [25] Tobin, R.L., Friesz, T.L.: Sensitivity Analysis for Equilibrium Network Flow, *Transportation Science*, vol. 22, no. 4, 1988.
- [26] Tobin, R.L.: Sensitivity Analysis for Variational Inequalities, *Journal of Optimization Theory and Applications*, vol. 48, no. 1, p. 191-204, 1986.