

# **Transit Network Optimization – Minimizing Transfers and Optimizing Route Directness**

*Fang Zhao, Florida International University  
Ike Ubaka, Florida Department of Transportation*

## **Abstract**

*This paper presents a mathematical methodology for transit route network optimization. The goal is to provide an effective computational tool for the optimization of large-scale transit route networks. The objectives are to minimize transfers and optimize route directness while maximizing service coverage. The formulation of the methodology consists of three parts: (1) representation of transit route network solution space; (2) representation of transit route and network constraints; and (3) solution search schemes. The methodology has been implemented as a computer program and has been tested using previously published results. Results of these tests and results from the application of the methodology to a large-scale realistic network optimization problem in Miami-Dade County, Florida are presented.*

## **Introduction**

Transit route network (TRN) design is an important component in the transit planning process, which also includes transit network schedule (TNS) design. A TRN optimization process attempts to find the route network structure with optimal transfer, route directness, and ridership coverage. Unfortunately, TRN design optimization processes suffer from combinatorial intractability, and thus far for practical transit network problems of large scales, TRN designs seem to be

limited to the use of various heuristic approaches where the solution search schemes are based on a collection of design guidelines, criteria established from past experiences, and cost and feasibility constraints. A systematic mathematical methodology applicable to large-scale transit networks for TRN optimization design seems to be missing.

The quality of a TRN may be evaluated in terms of a number of network parameters, such as route directness, service coverage, network efficiency, and number of transfers required. Route directness refers to the difference between the trip lengths,<sup>1</sup> if the trip is to be made by transit or by a car following the shortest path. Service coverage refers to the percentage of the total estimated demand (measured by transit trips) that potentially can be satisfied by the transit services based on a given transit route network. In this study, if the origin and destination of a potential transit trip are within walking distance of a transit stop and are connected by transit routes, the trip is considered served by the network or "covered." Network efficiency reflects the cost of providing transit services within a given network, other things being equal. Transfers are a result of the inability of a given network to provide direct service between all pairs of origins and destinations. Stern (1996) conducted a survey of various transit agencies in the United States, and about 58% of the respondents believed that transit riders were only willing to transfer once per trip. This suggests that the ridership of a transit system may be increased by merely reducing required transfers through the optimization of a TRN configuration. In addition to increasing ridership, an improved TRN configuration may also reduce transit operating cost and allow more services to be provided.

For transit systems with small bus route networks, a seasoned planner may be able to obtain near optimal bus route network results based on personal knowledge, experience, and certain guidelines. For large transit systems, intuition, experiences, and simple guidelines may be insufficient to produce even near-optimal transit route network configurations, due to the problem complexity. Therefore, systematic methodologies are needed to obtain better TRN configurations. This paper presents a methodology for TRN structure optimization based on a mathematical approach with the objectives of minimizing transfers, optimizing route directness, and maximizing service coverage (Zhao 2003). The methodology has been implemented as a computer program and has been tested using previously published results and a large-scale realistic network optimization problem in Miami, Florida.

### Formulation of A TRN Optimization Problem

A TRN optimization problem may be stated as the determination of a set of transit routes, given a transit demand distribution in a transit service area and subject to a set of feasibility constraints, to achieve objectives that optimize the overall quality of a TRN. Mathematically, a typical network optimization process may be stated as: optimize an objective function  $f(\mathbf{x}, \mathbf{y}, \mathbf{O}) \forall \mathbf{x} \in \mathbf{X}$  and  $\mathbf{y} \in \mathbf{Y}$ , subject to certain constraints, where  $\mathbf{x}$  is a real vector,  $\mathbf{y}$  is an integer vector (or a set of vectors), and  $\mathbf{O}$  is a matrix defined on the network's node set.  $\mathbf{X}$  is a space of real vectors, and  $\mathbf{Y}$  is a set of integer vectors

$$\mathbf{Y} = \mathbf{Y} \left\{ \mathbf{y}(i_1, i_2, \dots, i_s) \mid i_j \in \mathbf{N}, j = 1, 2, \dots, s \right\}$$

where  $\mathbf{N}$  is an integer set. A combinatorial optimization problem is a special case of integer optimization problems and refers to an integer optimization problem where the integer vector's component set in vector  $\mathbf{y}(i_1, i_2, \dots, i_s)$  is an ordered subset of a larger integer base set  $\mathbf{N}\{n_1, n_2, \dots, n_n\}$ , i.e.,  $(i_1, i_2, \dots, i_s) \subset \mathbf{N}\{n_1, n_2, \dots, n_n\}$  and  $n \geq s$  (in this paper, an ordered set is enclosed in parentheses while an unordered set is enclosed in brackets). TRN design is a typical combinatorial optimization problem, where the base set  $\mathbf{N}\{n_1, n_2, \dots, n_n\}$  is the set of all street nodes suitable to serve as transit stops, and the combinatorial set  $\mathbf{P}_N$  is the set of all paths in the street network suitable for transit vehicle operations. The matrix  $\mathbf{O} = \mathbf{O}(o_{ij})$  represents the transit demand at street nodes and is the OD matrix as  $o_{ij}$  represents the number of transit trips between street node  $n_i$  and  $n_j$ . This study deals with fixed transit demand problems.  $\mathbf{O}$  is assumed to be constant, representing transit demand for a given period of time of day, and does not change with transit supply. It should be recognized that, in reality, transit demand may depend on transit supply, thus TRN optimization ideally should be carried out in an iterative manner in a cycle of demand estimation and route network design. A transit route may be represented by an integer vector  $\mathbf{r}(i_1, i_2, \dots, i_s)$  with its component set  $(i_1, i_2, \dots, i_s)$  representing the sequence of a transit route's stops. A transit route network consisting of  $l$  routes may be represented by a set of integer vectors,

$$\mathbf{T}^{(l)} = \mathbf{T}^{(l)}\{ \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_l \}, \mathbf{r}_j = \mathbf{r}(n_{j1}, n_{j2}, \dots, n_{js(j)}), (j = 1, 2, \dots, l) \quad (1)$$

where  $s(j)$  is the number of transit stops on transit route  $\mathbf{r}_j$ . A transit route vector is a member of the combinatorial space  $\mathbf{P}_N$ , and a transit route network is a subset of  $\mathbf{P}_N$ . Based on the above definitions and notations, a fixed demand TRN design optimization problem may be stated as follows:

Maximize/minimize:

$$f(\mathbf{x}, \mathbf{T}^{(l)}, \mathbf{O}) \quad \forall \mathbf{x} \in \mathbf{X} \text{ and } \mathbf{T}^{(l)} \in \mathbf{P}_N \quad (2)$$

Subject to:

$$p_i(\mathbf{x}, \mathbf{T}^{(l)}) = 0, (i = 1, 2, \dots, i_p) \text{ and } q_i(\mathbf{x}, \mathbf{T}^{(l)}) \leq 0, (i = 1, 2, \dots, i_q) \quad (3)$$

where the real vector  $\mathbf{x}$  represents any continuous variables in the optimization process,  $\mathbf{O}$  is the OD matrix, and expressions in (3) represent various constraints in a TRN design process. Solving the TRN optimization problem, defined above, involves the search for an optimal set of feasible transit routes with unknown topology/geometry. It is difficult to solve problems with a large number of integer variables, since the associated solution procedure involves discrete optimization, which usually requires the search for optimal solutions from an intractable search space (Garey and Johnson 1979).

### Literature on TRN Optimization

A great deal of research has been conducted in the area of transit network optimization. The methods in the literature may be roughly grouped into two categories: mathematical approaches and heuristic approaches. However, there are no clear boundaries between these approaches. We consider an approach to be *mathematical* if the problem is formulated as an optimization problem over a relatively complete solution search space. Generic solution search methods are then employed to obtain solutions. Examples of such algorithms include various greedy type algorithms, hill climbing algorithms, simulated annealing approaches, etc. References and descriptions of various mathematical search algorithms may be found (e.g., Bertsekas 1998). We consider an approach to be *heuristic* if domain specific heuristics, guidelines, or criteria are first introduced to establish a solution strategy framework. Mathematical programming or other techniques are then employed to obtain the best results. The main difference between these two approaches is that the mathematical approach formulates a problem on a solution space with certain completeness that, theoretically, should include optimal solu-

tions. In contrast, the heuristic approach formulates a problem directly on solution sub-spaces defined based on domain specified heuristic guidelines.

Table 1 provides the main features of some of the approaches reported in the literature, where MATH represents mathematical optimization, and H&M (heuristic and mathematical) means that the author(s) established a solution based on a heuristic framework, but employed certain mathematical optimization methods at some solution stages. Most of the studies introduced some heuristics or certain simplifying assumptions to limit the solution search space or to reduce optimization objectives to a particular network structure or a few design parameters, e.g., route spacing, route length, stop spacing, bus size, or service frequency. (Detailed information and reviews of various mathematical optimization approaches may be found in Zhao 2003, among others.)

**Table 1. Main Features of Some Approaches Used in Transit Network Design**

Year	Author	Optimization Objectives	Design Variables	Solution Approaches	Constraints
1979	Mandl	Generalized time	Route	H&M	Coverage & directness
1991	Baaj & Mahmassani	Multi-objects	Route & frequency	AI/heuristic	Multi-constraints (heuristic guidelines)
1992	Bookbinder et al.	Disutility function-transfer inconvenience	Timetable/headway (offset time)	Math	Heuristic guidelines
1994	Shih & Mahmassani	Multi-objects	Route & frequency	H&M	Multi-constraints (heuristic guidelines)
1998	Bruno et al.	Generalized access cost to transit line	Location of a rapid transit line	Math	Route connectivity, demand coverage, etc.
1999	Sochodo & Koshi	Generalized social cost	Route & frequency	H&M	-
2003	Chien et al.	Total operator and user cost	Route shape & headway	Math	Route length, waiting time, load factors, etc.

The advantage of heuristic approaches is that they are always able to provide feasible solutions to problems of any size while the main disadvantage is that their results are almost certainly do not provide global or even local optimal solutions. This may be because heuristic search schemes are usually ad hoc procedures based on computer simulations of human design processes guided by heuristic rules. The corresponding search spaces are usually not clearly defined and search results are likely to be biased toward existing systems or any systems on which the set of design heuristics are based.

Compared with other methods in transit network design, mathematical approaches usually have more rigorous problem statements. A major disadvantage of mathematical approaches in TRN design is the computational intractability due to the need to search for optimal solutions in a large search space made up of all possible solutions. The resultant mathematical optimization systems derived from realistic combinatorial TRN problems are usually NP-hard, which refers to problems for which the number of elementary numerical operations is not likely to be expressed or bounded by a function of polynomial form (Garey and Johnson 1979). For this reason, existing mathematical optimization solution approaches to TRN problems are usually applied to relatively small and idealized networks for small urban areas or medium-sized urban areas with coarse networks. The route network structures may also be limited to certain particular configurations.

### **Solution Methodology**

Methodology was developed based on the following considerations: (a) the method should be generally applicable to the design and optimization of a wide range of TRN problems in practice; (b) the solution method should be as generic as possible and should not favor particular transit network configurations; and (c) solutions obtained from this method should give fairly good results in a reasonable amount of time, as permitted by the current computer power affordable to most transit agencies. Reliability of results should improve as the computer resource or power increases, and should approach the global optimum when there is no computer resource limitation.

### **Representation of Transit Service Area, Routes, and Route Network**

A transit service area is represented by a street network, which consists of a set of street nodes that are connected to each other by a set of street segments. A street segment,  $\mathbf{a}(n_1, n_2)$ , may be defined by its two end nodes  $n_1$  and  $n_2$ . In a directed network, segments  $\mathbf{a}(n_1, n_2)$  and  $\mathbf{a}(n_2, n_1)$  may be different as in the case of one-way streets or when travel impedance on the same link is different in the two opposite directions. In this study, only undirected network is considered (i.e.,  $\mathbf{a}(n_1, n_2)$  and  $\mathbf{a}(n_2, n_1)$  are considered the same), but the methodology can be easily extended to directed networks. It is also assumed that the street network is connected; thus, any two nodes in the street network are connected by at least one path.

The following is the mathematical representation of a street network. Denote  $\mathbf{N}^{(n)} = \mathbf{N}^{(n)}\{n_1, n_2, \dots, n_n\}$  as the set of  $n$  street nodes in a transit service area, then a street network consisting of  $m$  street segments may be written as  $\mathbf{A}^{(m)} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ ,

where  $\mathbf{a}_i = \mathbf{a}_i(n_{i1}, n_{i2})$  and  $n_{i1}, n_{i2} \in \mathbf{N}^n$  ( $i = 1, 2, \dots, m$ ). A path/route between any two nodes is defined as a sequence of non-reoccurring nodes, or  $\mathbf{p} = \mathbf{p}(n_1, n_2, \dots, n_k)$ , and there is one street segment, i.e.,  $\mathbf{a}(n_j, n_{j+1}) \in \mathbf{A}^{(m)}$  ( $j = 1, 2, \dots, k-1$ ), that connects any two neighboring nodes. A street network may also be represented through an adjacency list of street nodes. For a given node, called the *master* node of the list, its associated nodal adjacency list consists of all the neighboring nodes that can be connected to the master node with one street segment. The set of all nodal adjacency lists of a street network may be expressed as

$$\mathbf{L}(k) = \mathbf{L}(k)\{k_1, k_2, \dots, k_{m(k)}\}, k=1, 2, \dots, n \quad (4)$$

where:

- $\mathbf{L}(k)$  is the nodal adjacency list of the street node  $k$
- $k_j$  is the street node number of the  $j^{\text{th}}$  neighboring node in the list
- $m(k)$  is the number of nodes in the list

The TRN  $\mathbf{T}^{(l)}$  in (2) may also be expressed as a TRN matrix.

$$\mathbf{T} = \mathbf{T}[t_{ij}], t_{ij} = \begin{cases} 1, & \text{if node } j \text{ is on route } i, & i=1,2,\dots,l \\ \emptyset, & \text{if node } j \text{ is not on route } i, & j=1,2,\dots,n \end{cases} \quad (5)$$

In this study, for the purpose of representation uniqueness, it is assumed that the transit route stop set and the corresponding street node subset are the same.

### **Representation of Search Spaces for Transit Routes and Route Network**

The solution search spaces in this study are locally and iteratively defined, and the size of a local search space may be flexible based on available computing resources. A local path space consists of three components: (1) a master path; (2) a key-node representation of the master path; and (3) a set of paths that are in the neighborhood of the master path. A master path is a path from which a local path space will be generated. Key nodes are a set of nodes on a master path selected to defined paths in the local path space. A local path space is derived from the local node spaces of the key nodes on the master path. An  $i^{\text{th}}$  order local node space, denoted as  $\mathbf{N}_{(i)}(k)$ , of a master node  $k$  is defined as the set of nodes that can be connected to the master node with  $i$  or fewer street segments. The order of a local node space provides a measurement of the degree of localization. Figure 1 illustrates a three-key-node (nodes  $n_1, n_2,$  and  $n_3$ ) representation of a master path (solid line) and the

three first order local node spaces,  $\mathbf{N}_{(1)}(n_1) = \{n_1, n_{11}, n_{12}, n_{13}, n_{14}\}$ ,  $\mathbf{N}_{(1)}(n_2) = \{n_2, n_{21}, n_{22}, n_{23}, n_{24}\}$ , and  $\mathbf{N}_{(1)}(n_3) = \{n_3, n_{31}, n_{32}, n_{33}, n_{34}\}$ .

Denote the  $(i-1)^{\text{th}}$  order local node space of a master node  $k$  as  $\mathbf{N}_{(i-1)}(k) = \{k_1, k_2, \dots, k_{q(k)}\}$ , where  $q(k)$  is the number of nodes in this local node space, then

$$\mathbf{N}_{(i)}(k) = \{k_1, k_2, \dots, k_{q(k)}\} \subset \mathbf{L}(k_1) \subset \mathbf{L}(k_2) \subset \dots \subset \mathbf{L}(k_{q(k)}) \quad (6)$$

where  $\mathbf{L}(k_j)$  is the nodal adjacency list of node  $k_j$ . A local node space is a subspace of the street node space  $\mathbf{N}_{(i)}(k) \in \mathbf{N}^{(n)}$ . As the order  $i$  increases, it will approach to the original street node space  $\mathbf{N}^{(n)}$ . The procedure to generate a local path space from a master path has three steps: (1) Select  $s$  key-nodes from the node set of the master path  $\mathbf{p} = \mathbf{p}(n_1, n_2, \dots, n_r)$ , i.e.,  $\{m_1, m_2, \dots, m_s\} \in \{n_1, n_2, \dots, n_r\}$ ; (2) Generate a sequence of local node spaces from these key-nodes,  $(\mathbf{N}_{(i)}(m_1), \mathbf{N}_{(i)}(m_2), \dots, \mathbf{N}_{(i)}(m_s))$ ; and (3) Define the local path space as the set of paths consisting of piecewise shortest path segments that start from nodes in the first local node space  $\mathbf{N}_{(i)}(m_1)$ , sequentially pass the nodes in each of the intermediate local node space  $\mathbf{N}_{(i)}(m_j)$  ( $j = 2, 3, \dots, s-1$ ), and end at nodes in the last local node space  $\mathbf{N}_{(i)}(m_s)$ . The shortest path segments used to connect nodes in neighboring local node spaces are from a  $k$ -level shortest path space  $\mathbf{P}_s^{(k)}$  that consists of all the first  $k$  shortest paths between any two nodes in the street node space  $\mathbf{N}^{(n)}$ . (References on algorithms of finding a  $k$ -level shortest path space may be found in Zhao 2003.) The resultant path space, denoted as,  $\mathbf{P}_{(i)}^{(k)}(\mathbf{p}^{(s)})$ , will be referred to as the local path space based on the  $s$ -nodes representation of the master path  $\mathbf{p}$ , or simply the local path space of path  $\mathbf{p}$ .

The local network search spaces of a transit network  $T^{(l)} = \{r_1, r_2, \dots, r_j\}$  is defined as

$$T_{(i)}^{(k)}(T^{(l)}) = T_{(i)}^{(k)} \left\{ P_{(i)}^{(k)}(r_1^{(s_1)}), P_{(i)}^{(k)}(r_2^{(s_2)}), \dots, P_{(i)}^{(k)}(r_j^{(s_j)}) \right\} \quad (7)$$

where  $P_{(i)}^{(k)}(r_j^{(s_j)})$  is a local path space of  $s_j$ -node representation for master path  $r_j$ . It may be seen that as the two numbers  $i$  and  $k$  increase, a local path space of any master path will approach to the combinatorial path space  $P_N$

$P_{(i)}^{(k)}(r_j^{(s_j)})$  will be the path search space of the corresponding transit route  $r_j$ . In general, routes derived from smaller numbers of key nodes will result in better route directness and smaller local path search space, but their flexibility is also limited. Routes with larger numbers of key nodes are relatively more flexible to reach more neighboring nodes, thus may cover more trips. However, this will also result in larger local path search spaces, requiring more computing resources.

### **Integer Constraints for Transit Route Network**

Integer constraints in this study include the following: (a) fixed route constraints prescribing fixed guideway lines or bus routes that are specified by transit planners to meet certain planning goals, which will remain unchanged during the optimization process; (b) constraints prescribing starting, ending, or in-between areas through which transit routes must pass, which may include major activity centers or transfer points; (c) route length constraints for individual transit lines or for the entire system; and (d) constraints on the number of transit stops on individual routes.

### **Route Directness Constraints**

Route directness used in this study is defined as follows:

$$d(r) = \sum_{i=1}^{s-1} \sum_{j=i+1}^s w_{ij} (d_{ij}^{(r)} / d_{ij}^{(s)}), \quad (8)$$

where:

$s$  is the number of nodes on route  $r = r(n_1, n_2, \dots, n_s)$

$d_{ij}^{(r)}$  is the distance between nodes  $n_i$  and  $n_j$  measured along the transit route

$d_{ij}^{(s)}$  is the shortest network distance between nodes  $n_i$  and  $n_j$   
 $w_{ij}$  are weighting factors

For geometry based route directness,  $w_{ij} = w_{ij}^G \equiv 2/(s^2 - s)$ , and for ridership based route directness,

$$w_{ij} = w_{ij}^R \equiv (o_{ij} + o_{ji}) / \sum_{i=1}^{s-1} \sum_{j=i+1}^s (o_{ij} + o_{ji})$$

where  $o_{ij}$  and  $o_{ji}$  are coefficients of the OD matrix. The geometry based route directness, denoted as  $d^G(\mathbf{r})$ , reflects the average ratio of the two travel distances,  $d_{ij}^{(r)}$  and  $d_{ij}^{(s)}$ , between each node pair on route  $\mathbf{r}$ . A value of  $d^G(\mathbf{r}) = 1$  indicates that, on average, transit vehicles on route  $\mathbf{r}$  travel along the shortest paths between route stops. The ridership based route directness,  $d^R(\mathbf{r})$ , represents the average ratio of the distance a person travels between OD points along transit route  $\mathbf{r}$  to the distance traveled along the shortest path. A value of  $d^R(\mathbf{r}) = 1$  indicates that, on average, passengers on transit route  $\mathbf{r}$  travel along the shortest paths between OD points. Route directness constraints used in this study may be expressed as  $d^G(\mathbf{r}_i) \leq d_r^G$  or  $d^R(\mathbf{r}_i) \leq d_r^R$  ( $i = 1, 2, \dots, I$ ), where  $d_r^G$  and  $d_r^R$  are the two travel directness constraint parameters. In general, smaller  $d_r^G$  and  $d_r^R$  imply better services, but may result in higher transit operating cost. Large  $d_r^G$  and  $d_r^R$  mean that some potential transit riders may be turned away and that existing transit riders may be forced to look for other alternatives, thus leading to loss of ridership and, eventually, higher operation cost.

### **Network Directness Constraints**

Transit network directness has a physical meaning similar to that of the route directness, except that the directness measurement is based on geometry or ridership characteristics of the entire route network, instead of individual transit routes.

### **Out-of-Direction (OOD) Constraints**

The OOD constraint used in this study is derived from the formulation given by Welch et al. (1991). Denote  $d_{ij}^{(O)}(\mathbf{r})$  as the OOD impact index for travel between nodes  $i$  and  $j$  on transit route  $\mathbf{r}$ , then

$$d_{ij}^{(O)}(\mathbf{r}) = r_{ij}^{(1)}(\mathbf{r})[l_{ij}(\mathbf{r}) - d_{ij}]/r_{ij}^{(2)}(\mathbf{r})$$

where:

$r_{ij}^{(1)}(\mathbf{r})$  is the *through* ridership, or the number of trips on route  $\mathbf{r}$  that pass through nodes  $i$  and  $j$  without boarding or alighting in between

$r_{ij}^{(2)}(\mathbf{r})$  is the OOD ridership, which is the number of trips on route  $\mathbf{r}$  that involve either boarding or alighting or both at nodes between nodes  $i$  and  $j$

$l_{ij}(\mathbf{r})$  is the distance between nodes  $i$  and  $j$  along route  $\mathbf{r}$

$d_{ij}$  is the distance along the shortest path between these two nodes in the street network

$d_{ij}^{(0)}(\mathbf{r})$  represents the extra travel distance that incurs to each *through* passenger in order to serve an OOD passenger.

### Optimization Objective Functions

Objective functions considered in this study are various trip coverage functions or their combinations. The goal is to obtain a TRN structure with minimum transfers, while optimizing service coverage. If a trip between an OD pair requires no transfers, the trip is called a zero-transfer trip, while a trip between an OD pair that requires  $k$  or fewer transfers will be called a  $k$ -or-less transfer trip. A  $k$ -or-less transfer trip coverage function, or simply a  $k$ -or-less transfer function, is defined as the total number of OD trips that can be accomplished with  $k$  or fewer transfers in a transit network service area. The following is a description of various transfer coverage functions used in this study. Denote  $f_k$  as a  $k$ -or-less transfer function, then

$$f_k = f_k(\mathbf{T}, \mathbf{O}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [(o_{ij} + o_{ji})h(a_{ij}^{(k)})], \quad k = 0, 1, 2, \quad (9)$$

where

$\mathbf{T}$  is the TRN matrix defined in (5)

$\mathbf{O}$  is the OD matrix

$h$  is a step function that has the property:

$$h(x) = 1 \text{ for } x > 0, \text{ and } h(x) = 0 \text{ for } x \leq 0$$

Coefficients  $\alpha_{ij}^{(k)}$  in (9) are defined as  $\alpha_{ij}^{(0)} = \sum_{k=1}^l t_{ki} t_{kj}$ ,  $\alpha_{ij}^{(1)} = \sum_{k=1}^l \sum_{m=k}^l (t_{ki} t_{mj} \alpha_{km})$ ,

$\alpha_{ij}^{(2)} = \sum_{k=1}^l \sum_{m=k}^l (t_{ki} t_{mj} \beta_{km})$ ,  $\alpha_{km} = \sum_{j=1}^n t_{kj} t_{mj}$ , and  $\beta_{km} = \sum_{i=1}^l \alpha_{ki} \alpha_{mi}$ , where  $t_{ki}$ ,  $t_{kj}$  and  $t_{mj}$

are coefficients of matrix  $T$ . It may be seen that calculation of transfer objective function,  $f_2$ , is computational intensive, compared with functions  $f_0$  and  $f_1$ , due to the great number of arithmetic operations involved to obtain all the required coefficients.

The use of any of the transfer functions alone as the objective function may result in the optimization of one TRN parameter at the cost of others. The following are two objective functions that combine multiple coverage functions, thus giving more balanced results.

$$t_1(T, \mathbf{O}) = [f_0 + 2(f_1 - f_0) + \alpha(f_T - f_1)] / f_T \quad (10)$$

$$t_2(T, \mathbf{O}) = [f_0 + 2(f_1 - f_0) + 3(f_2 - f_1) + \alpha(f_T - f_2)] / f_T, \quad (11)$$

where

$\alpha$  is a weighting coefficient to penalize uncovered trips during the optimization process of the TRN system

$f_T$  is the total number of trips in the transit network service area

The physical meaning of the objective function  $t_2$  is the average number of vehicle boardings that a transit rider has to make to accomplish an OD trip. The optimal value of  $t_2$  is 1.0, indicating that all trips are zero-transfer trips. Uncovered trips ( $f_T - f_2$ ) are penalized by  $\alpha$ . The value of  $\alpha$  needs to be determined by transit planners. For example, by setting  $\alpha = 4$ , each of the uncovered trips is considered as four vehicle boardings. In general, the larger the value of  $\alpha$ , the greater relative importance is given to service coverage. The physical meaning of  $t_1$  is similar to that of  $t_2$ .

### **Algorithm 1—Basic Greedy Search Method**

The basic assumption of Basic Greedy Search (BGS) is that the demand distribution in a TRN service area has certain continuity. In other words, nodes with certain transit demands are probably close to nodes with similar demands. In such cases, it will be more effective in searching for a better solution by evaluating paths that are near nodes or areas with higher trip distributions. (Detailed description of various search algorithms used in this study can be found in Zhao 2003.) Assume

that during a solution search process, an intermediate TRN result  $T^{(l)} = \{r_1, r_2, \dots, r_j\}$  has been obtained. The solution search procedure for the next stage of BGS method involves the following steps:

1. Select key nodes from route  $r_j$ . For illustration, assume the three-node representation of route  $r_j$  is used (see Figure 1), which is denoted as  $r_j^{(3)}$ .
2. From the three key nodes  $n_1, n_2$ , and  $n_3$ , generate three first order local node spaces

$$N_{(1)}(n_1) = \{n_1, n_{11}, n_{12}, n_{13}, n_{14}\}$$

$$N_{(1)}(n_2) = \{n_2, n_{21}, n_{22}, n_{23}, n_{24}\}$$

$$N_{(1)}(n_3) = \{n_3, n_{31}, n_{32}, n_{33}, n_{34}\}$$

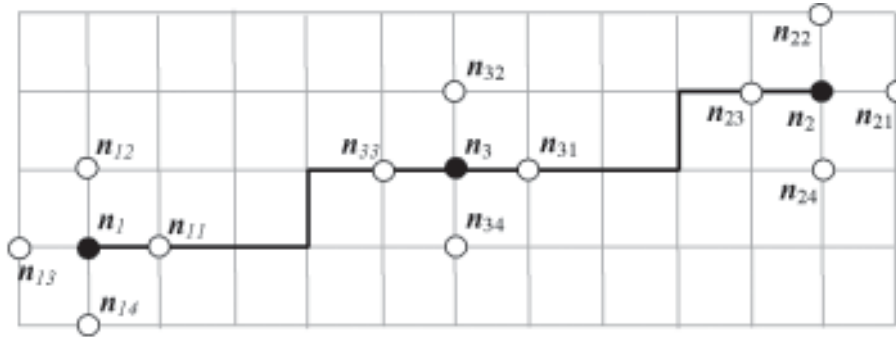
There are five nodes in each of the three local node spaces.

3. Connect nodes in node spaces  $N_{(1)}(n_1)$  and  $N_{(1)}(n_3)$  with the shortest paths in space  $P_S^{(1)}$ , to obtain  $5 \times 5 = 25$  shortest path segments. These shortest path segments are then extended with shortest paths to nodes in node space  $N_{(1)}(n_2)$  to obtain  $25 \times 5 = 125$  paths. These 125 paths form the local path space of route  $r_j$ , based on three-node representation  $P_{(1)}^{(1)}(r_j^{(3)})$ .
4. Replace route  $r_j$  in the existing TRN  $T^{(l)}$  with a path  $r_{jk} \circ P_{(1)}^{(1)}(r_j^{(3)})$  to obtain  $T^{(l)} = \{r_1, \dots, r_{j-1}, r_{jk}, r_{j+1}, \dots, r_j\}$ , and perform function evaluation for  $k = 1, 2, \dots, 125$ . If a better result is obtained, replace  $r_j$  with  $r_{jk}$  and go to Step 1 to start a new search. If no better result is found from all the 125 paths  $r_{jk} \circ P_{(1)}^{(1)}(r_j^{(3)})$ , go to Step 5.
5. Select the next route from the transit route network, e.g., route  $r_{(j+1)}$  and go to Step 1 to start a new local search for route  $r_{(j+1)}$ .
6. The search process will be considered converged if no better results can be found from the local path search spaces of all the individual routes.

### **Algorithm 2—Fast Hill Climb Search Method**

Conceptually, the Fast Hill Climb (FHC) method is similar to the deepest decent method in continuous research fields. First,  $l$  new solutions are formed by replacing one route at a time in the network, with the best route from its local search space. These  $l$  best routes from the local search spaces also make up a new solu-

**Figure 1. Three-Key-Node Representation of Transit Route**



tion. These  $l + 1$  solutions are compared and the best one is chosen as the current solution. Note that the computation process to obtain the  $l$  best routes is independent to each other, making it suitable for parallel computing.

### Numerical Experiments

The first test experiment was based on a real network in Switzerland (Mandl 1979). This problem was also used by Shih and Mahmassani (1994) and Baaj and Mahmassani (1991) as a benchmark problem to test their approaches to TRN and TNS design optimization. Mandl problem consisted of a street network of 15 nodes with a total demand of 15,570 trips per day. For this particular problem, the length of a street segment was defined in terms of in-vehicle travel time in minutes. In Table 2, the first row identifies the source of the solutions to the benchmark problem. The second row indicates solutions to the benchmark problem with different numbers of routes, total route length, and/or search methods. The methods used to obtain the results are indicated in the third row. For each solution, the unshaded column provides the statistics for the layout produced in the original studies, and the shaded column gives the statistics for the results produced from the FHC method developed in this study.

It may seem that the percentages of zero transfer trips were higher for all solutions produced in this study. Except for Mandl's original results, all solutions provided 100% trip coverage with zero or one (one-or-less) transfer involved in each trip.

The second experiment involved a large-scale TRN optimization problem based on the service area of the Miami-Dade Transit Agency (MDTA), encompassing a

**Table 2. Comparison of Results from Different Methods**

Problem Source	Mandl <sup>1</sup>		Baaj & Mahmassani						Shih & Mahmassani			
	1		1		2		3		1		2	
Search Method	Mandl	FHC <sup>2</sup>	B&M <sup>3</sup>	FHC	B&M	FHC	B&M	FHC	S&M <sup>4</sup>	FHC	S&M	FHC
0-transfer trips (%)	69.94	76.43	78.61	82.34	79.96	86.64	80.99	82.98	82.59	84.84	87.73	91.78
1-transfer trips (%)	29.93	23.57	21.39	17.66	20.04	13.36	19.01	17.02	17.41	15.16	12.27	8.22
2-transfer trips (%)	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total route length	82	82	126	125	144	144	106	105	124	124	151	152
Number of routes	4	4	6	6	8	8	7	7	6	6	8	8
Average Transfers	1.30	1.24	1.21	1.18	1.20	1.13	1.19	1.17	1.17	1.15	1.12	1.08

<sup>1</sup>Mandl's method

<sup>2</sup>Fast Hill Climb method

<sup>3</sup>Baaj and Mahmassani's methods

<sup>4</sup>Shih and Mahmassani's methods

region of about 300 square miles with a population of about 2.3 million. MDTA is ranked the 16<sup>th</sup> largest transit agency in the United States. At the time of this research, MDTA operated 83 transit routes, including a rail rapid transit system of 22.5 route miles (Metrorail), a 4.5-mile downtown automated circulation system (Metromover), and 81 bus routes with about 4,500 transit stops. The street network used in this experiment consisted of 4,300 street segments and 2,804 street nodes. In the optimization process, Metrorail and Metromover alignments were fixed and the longest and shortest bus routes were about 32 miles and 4 miles, respectively. The total length of the transit system was about 1,300 route miles, omitting some small loops at the ends of some routes or in shopping centers. The OD matrix was generated from the 1999 validated Miami-Dade travel demand model, which provided the daily number of passenger trips between each pair of traffic analysis zone centroids. These were manually distributed to the surrounding street network nodes with considerations given to land use patterns and street network connectivity. The total demand was 161,944 daily transit trips. All the numerical results were obtained on a personal computer with a 2.8GHz CPU and 1GB RAM memory. Table 3 presents the results from the BGS and FHC methods. There were two sets of results produced by each method, one based on an initial guess network that was the existing route network and the other based on a program generated initial guess network. The constraints were that the total

route length of the network should not exceed that of the existing system by more than 10%, and that the total number of transit lines remained the same as the existing system. Two objective functions were used, one maximizing zero-transfer trips ( $f_0$ ) and the other maximizing one-or-less transfer trips ( $f_1$ ). The values of the objective functions are given in the shaded cells.

Compared to the existing network, the FHC method with objective function  $f_0$  gave the best zero-transfer trip coverage, with an improvement of 85% (from 14.28% to 26.41%), while the BGS search method yielded an improvement of 84%. For objective function  $f_1$ , the FHC method again gave the best one-or-less transfer trip coverage, with a 48% improvement (from 55.13% to 81.57%). These improvements were achieved with a small increase of 5% in total network route mileage. Assuming most transit riders may be only willing to transfer once per trip (Stern 1996), the one-or-less trip coverage shown in the fourth row would be the actual total trip coverage of the corresponding route networks. The remaining trip demand either required two or more transfers or were not satisfied.

**Table 3. Comparison of Results with the Existing Network**

Network Parameters	Existing Results	Zero transfer objective function $f_0$				One-or-less transfer objective function $f_1$			
		BGS		FHC		BGS		FHC	
Solution Method	-	Exist <sup>1</sup>	Prog <sup>2</sup>	Exist	Prog	Exist	Prog	Exist	Prog
Initial Guess Network	-	Exist <sup>1</sup>	Prog <sup>2</sup>	Exist	Prog	Exist	Prog	Exist	Prog
0-transfer trips (%)	14.28	22.25	26.29	23.43	26.41	18.01	20.14	18.9	19.47
1-or-less transfer trips (%)	55.13	69.71	72.71	71.37	73.77	77.86	80.74	80.45	81.57
2-or-less transfer trips (%)	65.2	77.69	74.91	77.75	76.25	86.59	83.58	87.36	84.23
Total covered trips (%)	65.66	77.86	74.92	77.89	76.25	86.72	83.58	87.48	84.24
Total route mileage	1,278	1,330	1,340	1,340	1,348	1,304	1,341	1,345	1,346
Trips per route mile	83	95	91	94	92	108	101	105	101
Average Transfers	1.94	1.82	1.68	1.78	1.69	1.89	1.79	1.86	1.78
CPU Time (hours)	-	0.39	0.39	0.79	0.65	3.79	4.68	15.76	12.93

<sup>1</sup>Existing network as initial guess network

<sup>2</sup>Program generated initial guess network

The number of covered trips per route mile shown in Table 3 was defined as

$$R_{T/L}^{(2)} = f_2 / l_T$$

where

$f_2$  was the number of trips accomplished with two-or-less transfers

$l_T$  was the total length of the TRN

As a network efficiency indicator, the best  $R_{T/L}^{(2)}$  value was given by the BGS method with  $f_1$  as the objective function. The average transfers were defined as  $[f_0 + 2(f_1 - f_0) + 3(f_2 - f_1)] / f_2$ , which was the average number of boardings per transit rider who could complete a trip with two or fewer transfers. For the same objective function, the FHC method produced slightly better results than the BGS method. It may be seen that the differences in results produced by the BGS and FHC search methods were insignificant, but the BGS method was significantly faster than the FHC method.

Table 4 presents results obtained from composite trip coverage functions  $t_1$  and  $t_2$  described in (10) and (11), with the shaded cells indicating the objective function values. The penalty  $a$  was set at 4 in both functions  $t_1$  and  $t_2$ . It may be seen that improvements in various trip coverage functions were consistent instead of being achieved at the cost of each other, as in the case of single trip coverage function shown in Table 2. Overall, FHC produced slightly better results than those from method BGS, but at a higher computational cost.

## Conclusion

The methodology developed from this work has a systematic mathematical statement of TRN problems, including the definition of various objective functions, solution search spaces, and constraint conditions commonly used in transit planning fields, and a systematic scheme that flexibly defines solution search spaces based on available computing resources and/or optimization problem sizes. Two local search schemes have been developed to obtain results for large-scale practical problems in a reasonable amount of time.

The feasibility of the proposed method has been tested through practical TRN optimization problems of realistic sizes. Numerical results showed that the methodology developed in this work was capable of tackling large-scale transit network design optimization problems. Further improvements may include development

**Table 4. Comparison of Results with Existing Networks**

Network Parameters	Existing Networks	Average number of boardings function $t1$		Average number of boardings function $t2$	
		BGF	FHC	BGF	FHC
Objective Function $t1$	2.76	2.14	2.13	2.19	2.15
Objective Function $t2$	2.65	2.12	2.1	2.15	2.11
0-transfer trips (%)	14.28	24.22	24.55	22.4	24.88
1-or-less transfer trips (%)	55.13	80.7	81.31	79.45	80.23
2-or-less transfer trips (%)	65.2	83.54	84.37	83.67	83.61
Total covered trips (%)	65.66	83.54	84.38	83.68	83.62
Total route mileage	1,278	1,350	1,358	1,336	1,377
Trips per route mile	83	100	101	102	98
Average Transfers	1.94	1.74	1.76	1.79	1.74
CPU Time (hours)	-	7.62	9.85	90.41	114.59

of TRN optimization methods that consider dynamic transit demand, demand and travel time in different time period of a day, and waiting and transfer penalties.

### Acknowledgements

This research was sponsored by a grant from the Florida Department of Transportation (FDOT). The opinions, findings, and conclusions expressed in this paper are those of the authors and not necessarily those of the FDOT. The authors would like to thank the reviewers for their useful comments and suggestions.

### Endnote

<sup>1</sup>Depending on particular applications, length/distance may refer to either geometric length/distance or travel time.

## References

- Baaj, M. H and H. S. Mahmassani. 1991. An AI-based approach for transit route system planning and design. *Journal of Advanced Transportation* 25(2): 187-210.
- Bertsekas, D. P. 1998. *Network optimization: continuous and discrete models*. Belmont, MA: Athena Scientific.
- Bookbinder, H. J. and A. Désilets. 1992. Transfer optimization in a transit network. *Transportation Science* (26)2: 106-118.
- Chien, S., B., V. Dimitrijevic, and L. N. Spasovic. 2003. Optimization of bus route planning in urban commuter networks. *Journal of Public Transportation* 6(1): 53-80.
- Garey, M. R. and D. S. Johnson. 1979. *Computers and intractability: A guide to the theory of NP-completeness*. Freeman, NY.
- Mandl, C. E. 1979. Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research* (5): 396-404.
- Shih, M. C. and H. S. Mahmassani. 1994. *A design methodology for bus transit networks with coordinated operations*. SWUTC/94/60016-1. Austin, TX: Center for Transportation, University of Texas at Austin.
- Soehodo, S. and M. Koshi. 1999. Design of public transit network in urban area with elastic demand. *Journal of Advanced Transportation* 33(3): 335-369.
- Stern, R. 1996. Passenger transfer system review. *Synthesis of Transit Practice* 19, Transportation Research Board.
- Welch, W., R. Chisholm, D. Schumacher, and S. Mundle. 1991. Methodology for evaluating out-of-direction bus route segments. *Transportation Research Record* 1308: 43-50.
- Zhao, F. 2003. *Optimization of transit network to minimize transfers*. Draft technical report prepared for the Florida Department of Transportation. Miami: Lehman Center for Transportation Research, Florida International University.

## **About the Authors**

**FANG ZHAO** (*zhaof@fiu.edu*) received her Ph.D. in civil engineering from Carnegie Mellon University. She teaches in the civil and environmental engineering department at Florida International University, and is associate professor and deputy director of the Lehman Center for Transportation Research. Her main research interests include public transportation, travel demand modeling, geographic information systems, and transportation data modeling. Dr. Zhao is a registered professional engineer in the state of Florida and member of the Transportation Research Board's Transit Planning and Development Committee and New Technologies and Systems Committee.

**IKE UBAKA** (*ike.ubaka@dot.state.fl.us*) is program manager for transit systems planning at the Florida Department of Transportation, where he is responsible for statewide transit systems planning. He has over 18 years of experience in transit planning, including service with the Los Angeles County Metropolitan Transportation Authority and the Orange County (California) Transportation Authority. Mr. Ubaka holds a master's degree in urban and regional planning from UCLA and a master of administration degree from the University of California, Riverside. He is a member of the American Institute of Certified Planners.