Optimizing station locations along transit rail lines with geographic information systems and artificial intelligence

M. K. Jha & C. Oluokun
Department of Civil Engineering, Morgan State University, USA

Abstract

Locating stations along transit rail lines is a challenging and interesting scientific problem. It requires careful investigation of the demand to be served, station construction cost, parking availability at the stations, right-of-way acquisition cost, accessibility to the stations, and many other factors. It also requires the development of an efficient algorithm to search for the optimal location and number of stations along a rail line. The progress in this area has been slow primarily because it turns out to be a relatively complex problem when one attempts to consider factors that are difficult to model. For example, how to formulate the problem as an optimization problem by itself is quite challenging. Developing an efficient algorithm is another challenge coupled with the formulation of key factors in the decision making. In this paper a model is developed for optimizing station locations along a given transit line. The model uses genetic algorithms, one of the artificial intelligence-based optimization techniques. It also uses geographic information systems. The approach seems quite promising. Some factors which will be modeled in future work, include: (1) simultaneous optimization of rail lines and station locations; (2) development of a system of rail lines to serve a metro area with efficient transfer of passengers on connecting rails; and (3) effects of future land use changes in estimating the demand.
Keywords: transit rail lines, optimization, genetic algorithms, artificial intelligence, geographic information systems.

1 Introduction

Locating stations along transit rail lines in urban areas is a challenging and interesting scientific problem. It was extensively studied by Vuchic and Newell

Computers in Railways IX, J. Allan, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Editors)
 ISBN 1-85312-715-9

in 1968 [1]. The approach was primarily analytical and considered optimally locating stations along transit rail lines based on travel time minimization. Other similar studies [2-5] that followed considered minor improvements to Vuchic and Newell's [1] original approach. Some studied extended the approach to optimize bus stop spacings [6, 7]. A number of studies [8-10] were found to obtain the rail line length and optimize various associated parameters, such as headway, number of trains to satisfy the demand, and station spacings. Table 1 provides a summary of existing literature on transit rail optimization models.

Vuchic and Newell's [1] original approach to the station location optimization problem, however, was oversimplified in many respects. For example, only travel times were minimized in optimizing interstation spacings, and effects of station construction cost, parking availability at the stations, and right-of-way availability for station construction were ignored. Moreover, variability in population density was not considered; it was approximated by developing a functional form that could be used in developing an analytical solution.

With the emergence of artificial intelligence-based optimization techniques, such as genetic algorithms [11], and ant-colony system [12-15] it is possible to develop a numerical solution to the station location optimization problem along transit rail lines while incorporating many significant factors in the formulation not considered previously. It is also possible to use a geographic information system in conjunction with the above mentioned optimization approaches [16-18], which can provide information on population density, spatial location of residential properties and existing highway networks, and other useful parameters necessary to realistically solve the station location optimization problem.

2 The research problem

The research problem is to optimally locate stations along a given transit rail line (Figure 1) which serves many-to-one demand pattern extending radially outwards from the heart of a central business district (CBD) to low-density suburban areas. Thus, it is assumed that the end points of the rail line are specified and we have to decide on the number and location of the intermediate stations which can minimize the sum of user, operator, and capital costs. The primary components of user costs are: cost associated with (1) traveling to the station and (2) in-vehicle travel time. The primary components of the operator cost are (1) cost of operating the train and (2) train frequency. The capital costs are: (1) land-acquisition cost (also commonly referred to as right-of-way cost) and (2) construction of stations and parking facilities.

In the future works after the station location model is developed, simultaneous optimization of rail line and stations along it will be performed. Finally, the methodology will be extended for developing a transit rail system with multiple rail lines serving an urban area. Developing such a model will be quite complex since many factors, such as transfer points and passenger origin-destination pattern (to name a few) will have to be considered.

Computers in Railways IX, J. Allan, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Editors)

Table 1: Existing literature on rail optimization models (source [5]).

Decision Variables	Objective Function	Transit Mode	Street Network Geometry	Passenger Demand	Reference
Route Length	Min. operator and user cost	Rail	Rectangular grid	General. Inelastic. Many-to-one	[26]
Zone Length, Headway	Min. operator and user cost	bus	Rectangular grid	Piecewise uniform, inelastic	[27]
Route Spacing, Lengths and Headway	Min. operator and user cost	Bus and rail	Rectangular grid	Uniform, inelastic, many-to-one	[28]
Route Spacing	Min. operator and user cost	Bus	Rectangular grid	Uniform, inelastic, many-to-one	[29]
Route spacing and Headway	Min. operator and user cost	Bus	Rectangular grid	Uniform, inelastic, many-to-one	[30]
Route Density and Frequency	Min. operator and user cost	Bus	Rectangular grid	General Linear, inelastic, many-to-one	[31]
Route Spacing, Headway and Fare	Max. operator profit, Max. user benefit etc	Bus	Rectangular grid	Uniform, Elastic, many-to-one	[32])
Route Spacing, Headway and Stop Spacing	Min. operator and user cost	Feeder bus to rail	Rectangular grid	General Inelastic, many-to-one	[7]
Route Spacing, Headway and Fare	Max. profit, max. welfare, min. cost	Bus	Rectangular grid	Irregular, elastic, many-to-many, time dependent	[33]
Route Spacing, Zone Length, Headway	Min. operator and user cost	Bus	Rectangular grid	Uniform, inelastic, many-to-one	[34]
Station Location and Spacing	Min. total user travel time	Rail	Linear	Uniform, inelastic, many-to-one	[1]
Stop Location and Spacing	Min. operator and user cost	Rail	Rectangular grid	Uniform, inelastic, many-to-one	[4]
Stop Spacing	Min. operator and user cost	Bus	Radial	General, inelastic, many-to-many	[35]

2.1 The challenges

The research problem offers a number of challenges requiring extensive scientific investigation. First, how do we determine the demand of the effective

Computers in Railways IX, J. Allan, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Editors)

region to be served by the rail line. Previous study [1] assumed the demand as a function of the rail-line length (passengers/length). Several weaknesses were noted in Vuchic and Newell's [1] approach: (1) the rail-line was assumed to be straight whereas in reality rail-lines are a combination of tangent and curved (circular for horizontal alignment, parabolic for vertical alignment) sections; (2) passengers originate from their homes and are unevenly spread out in the catchment area (a catchment area is defined as the effective geographic region from which people will be attracted to ride the train); their choice of a boarding station along a rail line primarily depends on travel time from home to the station plus the in-vehicle travel time; (3) demand cannot be assumed to be uniformly distributed; it depends on the population density and attractiveness towards the train ride based on travel time; (4) in addition to the travel-time costs, land acquisition cost, and capital costs for station construction must also be considered in determining optimal station locations. (5) effects of future induced demand should be considered.

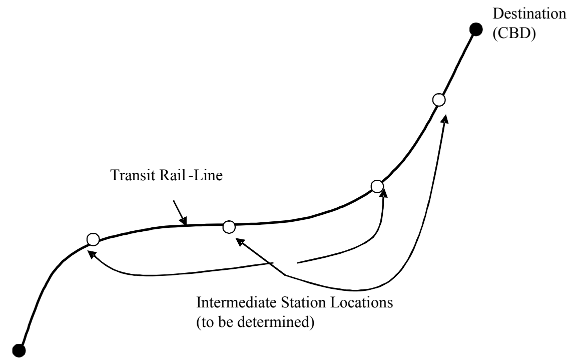

Figure 1: The transit rail station location optimization problem.

2.2 Our approach

For a moment, let's assume that we had to develop an optimization problem to the research problem mentioned above. The total cost is considered to be the objective function and station spacings are considered to be the decision variables. Since it is practically impossible to develop a straightforward relationship for the objective function (total cost) as a function of the decision variables (station spacings) while considering all important factors mentioned in the preceding section, traditional search algorithms (such as dynamic programming, linear programming, calculus of variations, gradient-based search) cannot be employed unless the problem is significantly oversimplified. Therefore, we turn our attention to artificial-intelligence based heuristics, such as genetic algorithms, artificial neural networks, simulated annealing, tabu search, or ant colony system. We present a GIS and genetic algorithms-based approach (discussed later) to solve the problem. GIS allows working directly with maps of the proposed rail-line, existing road networks, and land and property boundaries, and can be integrated to one of the artificial intelligence-based optimization techniques through specialized dynamic link libraries [16].

3 Problem formulation

Our formulation is based on identifying and including the most important factors in the optimization model for locating stations along a transit rail line. From the user's perspective it is the travel time to the station from home that a transit rider worries the most since the freeways and arterial streets are heavily congested during rush-hours. The in-vehicle travel time is less significant compared to the time taken to arrive at the station (which is always worse and often unpredictable, especially during rush hours) since additional stations may only add up a few minutes to the in-vehicle travel time. From the agency's perspective it is the right-of-way (ROW) (or land acquisition) cost which is the most significant. The ROW cost will generally increase as one travels from low-density suburban areas towards high-density urban areas although it is very likely that an expensive property might exist just next to a cheaper one; therefore, it is difficult to model the right-of-way cost and it must be calculated using actual property and land values [19]. The station construction cost will be significantly higher if a tunnel construction (closer to the CBD) is necessary.

3.1 Assumptions

We assume that people within the catchment area make a decision of either riding a train at one of the available stations or driving directly to the CBD, based strictly on the total travel time (time to go to the station plus the in-vehicle travel time). The walk-time and wait-time while not mentioned here, are all assumed to be part of the total travel time. Detailed formulation for these quantities exists in the literature [1, 3, 5]. Therefore, we have omitted their discussion for the sake of brevity. We also assume that the demand pattern is many to one, i.e., everyone is destined to a single location (CBD). A sufficiently developed digital GIS map containing street network for the catchment area encompassing the rail line is assumed to be available. The GIS map of residential locations is also assumed to be available (see, [20] for a detailed discussion of how necessary information can be extracted from a GIS map).

3.2 Formulation methodology

Let there be N possible intermediate stations (Figure 2) that can be accommodated between the starting station (S) and destination (CBD). The total rail line length is L. Let the minimum distance to be maintained between station pairs for acceleration and deceleration (see [36] to learn how acceleration and deceleration can be of concern when the grade changes; in our analysis effects of grade has been ignored, but, can be easily considered later once the model is sufficiently developed) be $\Delta S_{\min}$. Therefore, $N \leq \dfrac{L}{\Delta S_{\min}}$, where N is an

Computers in Railways IX, J. Allan, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-715-9

integer. Further, let's assume that once the stations $Z = \{S, S_1, S_2, ..., S_k, ..., S_N\}$ are specified the approximate travel times from a residential location $P_i (i = 1,2,...., R)$ to a station (or if driving directly to the CBD is attractive, the user can so choose) can be obtained from a GIS database (see, Figure 2).

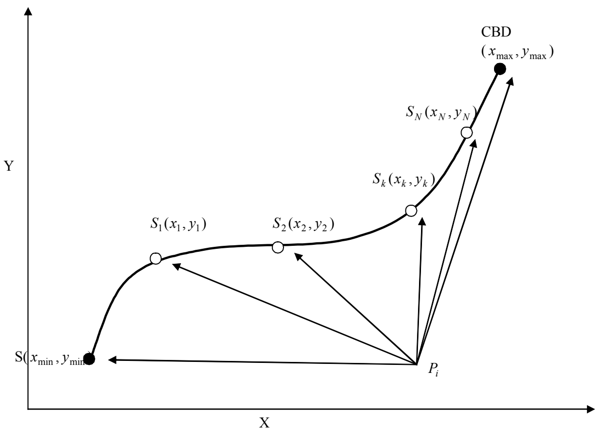

Figure 2: Station locations and travel times.

Theorem 1: A potential train rider located at P_i opts for riding the train at S_k iff $TT(S_k) < TT(S'_k)$; where TT represents the travel time and $Z' = \{S, S_1, S_2, ..., S_{k-1}, S_k, S_{k+1}, ..., S_N, CBD\} - \{S_k\}$.

Lemma 1: A train rider is indifferent in choosing S_k over S'_k if $TT(S_k) = TT(S'_k)$

3.2.1 The worst case solution

Let TC be the total cost when all possible stations are chosen and is defined as:

$$TC(\Delta S_1, \Delta S_2, ..., \Delta S_{N+1}) = \sum_{k=1}^{N+1} TC(\Delta S_k) = \sum_{k=1}^{N+1} \left[\sum_{i=1}^{R} UC(\Delta S_k)(P_i) + OC(\Delta S_k) + CC(\Delta S_k) \right] \tag{1}$$

where UC, OC, and CC are user cost, operator cost, and capital cost, respectively. $P_i, i = 1,2,....., R$ represent number of residential properties (trip generators) location of which is known from a GIS. It is further assumed that approximate travel times from $P_i\ \forall i = 1,2,...., R$ to $S_k\ \forall k = 0,2,....., N+1$ are also available from a GIS. User cost is a function of travel time (unit travel time cost times total travel time), Operator cost is a function of number of train trips between the origin and CBD (please note that developing a train schedule to serve the demand, by itself is a complex research problem; here we assume that number of train trips necessary to serve the demand is given from which cost of train operation can be obtained. Cost of train operation is only assumed to be a function of number of stops, i.e., higher the number of stops, larger the operating cost), and capital cost is composed of station construction cost and cost of right-of-way acquisition. The station construction cost is also assumed to be available once the location of a station is specified. The right-of-way cost is determined based on the impact to existing land and structures on the properties due to the station and parking lot constructions. It is based on a sophisticated analysis that uses not only the value of the fraction of the properties taken, but also, the after values of the property remnants and extent of impact (see [19]).

The future induced demand can also be considered if sufficient information were available about how the land use pattern might change over time. The above formulation represents the worst case solution, which we call $TC_{\max}$

3.3 The optimization problem

The optimization problem can now be expressed as:

$$\begin{aligned} & Min \quad \sum_{j=1}^{k} TC(\Delta S_j) \\ & s.t. \\ & L - \Delta S_{\min} \geq \Delta S_k \geq \Delta S_{\min} \\ & 1 \leq k \leq N,\ k \text{ is an integer} \end{aligned} \tag{2}$$

Theorem 2: If Z^* represents the optimal set of stations and $\left\{\Delta S_1^*, \Delta S_2^*, \ldots\ldots, \Delta S_k^*\right\}$ corresponding Interstation spacings, then $TC^*(\Delta S_1^*, \Delta S_2^*, \ldots, \Delta S_k^*) \leq TC_{\max}$.

Theorem 3: If the orientation of S_k is given by (x_k, y_k) in the (X,Y) plane (see, Figure 2) then it is sufficient to find x_k to locate S_k.

Lemma 1. For every x_k there is one and only one y_k

4 Solution methodology using genetic algorithms

Genetic Algorithm (GA) [11] is a technique for solving an optimization problem, but not all problems can be solved in the default format of GAs. For different systems, one may have to develop different solution procedures based on the philosophies and principles of GAs, and the nature of the problem.

The application of GAs to a specific problem includes several steps. A suitable encoding for the solution must be devised first. We also require a fitness function through which the individuals are selected to reproduce offspring by undergoing genetic operators. The key steps of how genetic algorithms can be developed for the proposed research problem can be described as follows:

Computers in Railways IX, J. Allan, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Editors)

Step I. Generate a random number (integer) between 1 and N: $r_d(1,N)=k$; k is an integer . This is the number of intermediate stations and the initial solution.
Step II. Generate k random numbers $(r_{c1},r_{c2},....,r_{ck})=[\lambda_1,\lambda_2,.....,\lambda_k]$ in the interval $(\Delta S_{\min}, L-\Delta S_{\min})$, which represent the interstation spacings. These represent the encoded solution to the problem called chromosomes (see [23]). This also represents the initial population.
Step III. Calculate fitness (total cost function) of the population members.
Step IV. Develop appropriate genetic operators (see [21], [23]).
Step V. Develop a selection-replacement scheme.

5 Conclusions and future work

Simultaneous optimization of a transit rail line and stations along it can be performed by developing a two-stage genetic algorithm. In the first stage during the optimal search, a rail-line can be obtained following procedures developed by Jha and Schonfeld [25]. In the second stage, stations can be located along that rail-line using the proposed genetic approach.

The proposed research is just one step in solving the transit rail line problem in developing a system of rail lines with due consideration to transfer points and their impact on selecting appropriate rail-line lengths and location stations while serving the necessary demand. Thus, this research may be considered as the beginning of a long-term project. Please note that isolated research studies dealing with efficient transfer of passengers along rail routes are available. Therefore, what may be necessary is to carefully integrate the already developed methodology in the proposed research. Since we consider this issue beyond the scope of this paper, elaborate discussion has been skipped. In future developments we shall also consider many-to-many travel demand patterns for more realistic applications.

References

[1] Vuchic, V.R. & Newell, G.F., Rapid Transit Interstation Spacing for Minimum Travel Time, *Transportation Science,* **2(4),** pp. 359-374, 1968.
[2] Vuchic, V.R., Rapid Transit Interstation Spacing for Maximum Number of Passengers, *Transportation Science,* **3(4),** 214-232, 1969.
[3] Wirasinghe, S.C., Nearly Optimal Parameters for a Rail/Feeder-Bus System on a Rectangular Grid. *Transportation Research-A,* **14**, pp. 33-40, 1980.
[4] Hurdle, V.F. & Wirasinghe, S.C., Location of Rail Stations for Many-to-One Travel Demand and Several Feeder Modes, *Journal of Advanced Transportation,* **14(1),** pp. 24-45, 1980.
[5] Spasovic, L. & Schonfeld, P., Method for Optimizing Transit Service Coverage. *Transportation Research Record* **1402**, pp. 28-39, 1997.
[6] Wirasinghe, S.C. & Ghoneim, N.C., Spacing of Bus-Stops for Many to Many Travel Demand, Transportation *Science*, **15(3),** pp. 210-221, 1981.

Computers in Railways IX, J. Allan, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Editors)

[7] Kuah, G.K. & Perl, J., Optimization of Feeder Bus Routes and Bus Stop Spacings. *Journal of Transportation Engineering,* **114(3),** pp. 341-354, 1988.

[8] Wirasinghe, S.C. & Seneviratne, P., Rail Line Length in an Urban Transportation Corridor. *Transportation Science,* **20(4),** pp. 237-245, 1986.

[9] Chien, S. & Schonfeld, P., Optimization of Grid Transit System in Heterogeneous Urban Environments, *Journal of Transportation Engineering,* **123(1),** pp. 28-35, 1997.

[10] Chien, S. & Schonfeld, P., Joint Optimization of a Rail Transit Line and its Feeder Bus System, *Journal of Advanced Transportation,* **32 (3),** pp. 253-284, 1998.

[11] Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning,* Addison-Wesley Publishing Company, Inc.: Massachusetts, 1989.

[12] Dorigo, M., Caro, G. D., & Gambardella, L.M., Ant Algorithms for Discrete Optimization, *Artificial Life,* **5(3),** pp. 137-172, 1999.

[13] Dorigo, M., Maniezzo, V., & Colorni, A., The Ant System: Optimization by a Colony of Cooperating Agents, *IEEE Transactions on Systems, Man, and Cybernetics-Part B,* **26 (1),** pp. 1-13, 1996.

[14] Dorigo, M. & Gambardella, L.M., Ant Colonies for the Traveling Salesman Problem, *Biosystems,* **43,** pp. 73-81, 1997.

[15] Jha, M.K., Optimizing Highway Networks: A Genetic Algorithm and Swarm Intelligence Based Approach. In *Computing in Civil Engineering,* pp. 76-89, A. Songer & J. Miles (eds). American Society of Civil Engineers (ASCE), Reston, VA, 2002.

[16] Jha, M.K. & Schonfeld, P., Integrating Genetic Algorithms and GIS to Optimize Highway Alignments. *Transportation Research Record* **1719**, pp. 233-240, 2000.

[17] Jha, M. K., Geographic Information Systems and Artificial Intelligence Integration in Transportation: An Overview, *Proceedings of the World Multiconference of Systemics, Cybernetics, and Informatics,* Volume VII, Part I, pp. 13-17, Orlando, FL, 2001.

[18] Jha, M. K., Using a GIS for Automated Decision-Making in Highway Cost Analysis. *Transportation Research Record* **1768**, pp. 260-267, 2001.

[19] Jha, M.K. & Schonfeld, P., Geographic Information Systems-Based Analysis of Right-of-Way Cost for Highway Optimization. *Transportation Research Record* **1719**, pp. 241-249, 2000.

[20] Jha, M. K., McCall, C., & Schonfeld, P., Using GIS, Genetic Algorithms, and Computer Visualization in Highway Development. *Computer-Aided Civil and Infrastructure Engineering,* **16(6),** pp. 399-414, 2001.

[21] Jong, J.-C., Jha, M. K., and Schonfeld, P., Preliminary Highway Design with Genetic Algorithms and Geographic Information Systems. *Computer Aided Civil and Infrastructure Engineering,* **15(4),** pp. 261-271, 2000.

Computers in Railways IX, J. Allan, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Editors)

[22] Jha, M.K., Criteria-Based Decision Support System for Selecting Highway Alignments, *Journal of Transportation Engineering,* **129(1),** pp. 33-41, 2003.

[23] Jong, J.-C. & Schonfeld, P., An Evolutionary Model for Simultaneously Optimizing 3-Dimensional Highway Alignments. *Transportation Research-B,* **37(2),** pp. 107-128, 2003.

[24] Kim, E., Jha, M.K., Lovell, D.J., & Schonfeld, P., Intersection Cost Modeling for Highway Alignment Optimization. *Computer-Aided Civil and Infrastructure Engineering,* **19(2),** pp. 136-146, 2004.

[25] Jha, M.K. & Schonfeld, P., Transit Rail Line Optimization with Genetic Algorithms and GIS, Paper Number 03-4383, *Proceedings of the 82nd Annual TRB Meeting,* Washington, D.C., 2003.

[26] Wirasinghe, S.C. & Seneviratne, P., Rail Line Length in an Urban Transportation Corridor. *Transportation Science,* **20(4),** pp. 237-245, 1986.

[27] Schonfeld, P. & Tsao, S.M., Branched Transit Service: An Analysis. *Journal of Transportation Engineering,* **109(2),** pp. 257-272, 1984.

[28] Byrne, B.F., Cost Minimizing Positions, Lengths, and Headways for Parallel Public Transit Lines Having Different Speeds, *Transportation Research-B,* **10**, pp. 209-214, 1976.

[29] Holroyd, E.M., The Optimal Bus Service: A Theoretical Model for a Large Uniform Urban Area. In *Vehicular Traffic Science* (L.C. Edie, ed.), American Elsevier: New York, pp. 309-328, 1967.

[30] Byrne, B.F. & Vuchic, V. R., Public Transportation Line Position and Headway for Minimum Cost. In *Traffic Flow and Transportation* (G.F. Newell, ed.), American Elsevier: New York, pp. 347-360, 1972.

[31] Hurdle, V.F. (1973). Minimum Cost Locations for Parallel Public Transit Lines. *Transportation Science,* **7(4),** pp. 340-350, 1973.

[32] Kocur, G. & Hendrickson, C., Design of Local Bus Service with Demand Equilibrium. *Transportation Science,* **16(2),** pp. 149-170, 1982.

[33] Chang, S.K. & Schonfeld, P., Analytical Optimization of Parallel Bus Routes with Time Dependent and Elastic Demand, *Working Paper 89-22, Transportation Studies Center, University of Maryland, College Park,* 1989.

[34] Chang, S.K. & Schonfeld, P., Optimization Models for Comparing Conventional and Subscription Bus Feeder Services. *Transportation Science,* **25(4),** pp. 281-298, 1991.

[35] Wirasinghe, S.C. & Ghoneim, N.C., Spacing of Bus-Stops for Many to Many Travel Demand. *Transportation Science,* **15(3),** pp. 210-221, 1981.

[36] Kim, D. N. & Schonfeld, P., Benefits of Dipped Vertical Alignments for Rail Transit Routes, *Journal of Transportation Engineering,* **123(1),** pp. 20-27, 1997.

Computers in Railways IX, J. Allan, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Editors)