



A Transportation planning by designing a bus network using genetic algorithm

Rakesh Rosan Patra

Regional Institute of Education, Bhubaneswar

Abstract : The bus network design problem is an important problem in transportation planning. It is the problem of determining a network of bus lines which best achieves a predetermined objective. This may be done with or without the presence of rapid transit lines.

This study is devoted to solving this problem using genetic algorithm. The fitness function is defined as the benefit to the users of the bus network less the cost of the operator of the network, which is to be maximized subject to constraints that properly distribute bus routes over the study area. Objective function calculation depends on the basic data of the city and its bus lines and does not need traffic assignment results. So, it is calculated quickly and it makes the genetic algorithm operation faster. Several good solutions were generated through a sensitivity analysis by changing the parameters of the problem affecting bus route geographical distribution.

This basic requirement of efficient mobility of goods and passengers gives rise to, among many other things, the subject areas of optimal routing and scheduling. In the following sections the problems of optimal routing and optimal scheduling are explained. Finally, how these optimization problems, which are often difficult to solve using traditional optimization tools, have been solved using genetic algorithms are explained.

A network assignment problem was solved for each of the alternative bus networks and several measures of effectiveness were evaluated for them. A multi-objective analysis (concordance analysis) was performed based on 10 measures of effectiveness and 14 weighting systems. As a result, a bus network was proposed for the city of Mashad, Iran.

I. INTRODUCTION

Bus network design which generates bus lines with respect to some limitations such as fleet size and budget is one of the most important problems of transportation planning. In general, users are faced with many problems in transit operation such as long waiting time at stations, inaccuracy in bus arriving time, incompatibility between the bus routes and paths of most passengers, and insufficient capacity. These facts encourage some users to shift to a private transportation. Some unfavorable effects of using a private vehicle are congested network and increasing in fuel consumption, wear and tear of car parts, and number of accidents. One of the main reasons results in inefficiency in transit operation is that bus lines do not cover the network very

well. A bus network designer aims to remove mentioned occurrences by choosing the best routes associated with proper frequencies. It is not optimal to consider every route in a bus network due to deficiency in fleet size and the budget. Moreover, it is not practical to cover the entire network with bus lines. Therefore, the designer must choose the best bus lines to have the maximum network coverage taking fleet size and other constraints into consideration.

The bus network design is a very difficult problem to solve for several reasons. First, it is multi-criteria problem. Several factors should be considered in bus network design such as cost, frequency of buses, bus lines' coverage, transit travel time, and fleet size. Some of these objectives are conflicting. For instance, increasing the coverage of bus lines increases the cost. Each decision variable is associated with a specific bus line. As a result, designing a bus network for a large city has a lot of integer decision variables that make the model complicated. As a result, the well-known integer program solution methods such as branch and bound are not able to solve this problem for a large city optimally. Furthermore, they may not find a good feasible solution in reasonable amount of computational time.

Genetic Algorithm (GA) which is one of the most widely used heuristic methods, is a search technique used to find solutions for optimization problems. These solutions are often close to optimal or in some cases optimal solutions. GA which belongs to the global search heuristics is a particular evolutionary algorithm that uses techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover. The evolution usually starts from a population of randomly generated chromosomes. In each generation, the value of every chromosome in the population is calculated according to their objective function which is called fitness function. So, multiple chromosomes are stochastically selected from the current population (based on their fitness function), and modified. In other words, they are recombined and possibly randomly mutated to form a new population.

In this paper, a new method to design a bus network using GA is developed. In this method, each chromosome represents a set of bus lines or a bus network and each gene stands for one bus line. Each

gene can have the value of either zero or one. The value of one for gene i in chromosome j , means that bus line i has been selected in bus network j ; otherwise, it has not been selected. In addition, the objective function contains benefit and cost of the bus network and the GA tries to maximize benefit minus cost.

The results were analyzed using the network assignment procedures embedded in EMM/2 software to estimate bus networks' performance measures. The Concordance Analysis which is a multi-criteria decision making approach was used to identify the best alternatives. The results indicated that solutions found by GA perform better than solution provided by other methods.

The benefit of bus network is defined as coverage of bus stops and its cost is calculated as a function of bus network length. Objective function calculation in this method is so easy and it takes short time. This is the main advantage of this approach.

In the following sections, first the literature regarding the bus network design will be reviewed. Then, the GA for bus network design is described including details regarding formulation such as objective function elements, definition of chromosome and gene, and estimation of different parameters. Finally, conclusions and suggestions for future studies are discussed.

II. LITERATURE REVIEW

In this section, previous studies in the field of bus network design will be reviewed. The first subsection covers some important previous attempts in this field and their goals and assumptions. The second subsection includes limitations of previous studies that address researchers to use heuristic methods to solve this problem.

Previous Studies

Many researchers have worked on bus network design using different methods. The proposed approaches for bus network design can be classified into two groups, theoretical and practical approaches.

Theoretical Methods The theoretical approaches to bus network design focus only on mathematical problems. Route length, distance between stops, and time headway are decision variables seen in these models. The objective functions in these models are to minimize the operation costs or to maximize user benefits.

Some of these models are based on constant demand, however Hurdle (1973), Newell (1979), Kocur and Hendrichson (1982) are examples of these approaches that have used flexible demand. For instance, Guan et al. (2003) tried to minimize a function of number and total length of bus lines and total distance traveled by passengers subject to route length bounds, capacity and limits on number of transferred. Although there are a lot of decision variables exist in these models, their solutions which most of them are calculated approximately are not practical and efficient in real

world bus network design. These models can be solved optimally for small-size problems.

Practical Methods Another group of bus network designers have used real bus lines. Like the theoretical designers, practical designers aim to minimize the operation costs and maximize user benefits.

The pioneers in this group are Lampkin and Saalmans (1978), Silman et al. (1974), Mandle (1979), and Dobois et al. (1979). In all of the above efforts except Dobois et al. (1979), demand is constant. In addition, they divided the main problem into two different problems: first, designing the bus lines and second, calculating time headway for each line. Then, they calculated the operation cost or system efficiency.

Lines et al. (1966) designed a bus network for a city in England. The goal was minimizing total time including waiting time in bus stops, travel time in bus, and walking time from origin to bus stop and from bus stop to destination. They found the best bus lines then they calculated time headway and determined fleet size for each bus line. A simple algorithm has been used to find the best bus lines followed by determining time headway for each bus line while they minimized total time. This algorithm can be used for small cities that do not have a complex bus network, but it is not efficient for large cities.

Dufourd, Gendreau, and Laporte (1996) maximized population coverage of bus lines in their model which is used in many location models and transit planning models. Although trip coverage of bus lines is more important than population covered by lines, there is a relation between population coverage of one bus line and the number of trips covered by it. In fact, number of trips is increased as the population grows. The disadvantage of this research is that the operation cost was not covered in the objective function.

Heuristic methods for bus network design

There are many decision variables in bus network design problem, which considering all of them in planning increases the size of problem as well as the computational running time. Moreover, routine methods are not able to solve this problem optimally.

Path finding for a transit line is a difficult and complicated problem, because it is multi-criteria problem needs a lot of input data which most of them are not available and it is very expensive to collect them. All of the above mentioned methods either are useful for small cities or give only feasible solution for large scale problems which is far from the optimal.

Tabu and Greedy search algorithm. However, Zhao and Ubaka (2004) implemented a basic greedy search and fast hill climb search to minimize the number of transfers and maximize network coverage. Xiong and Schneider (1993) used GA to add some bus lines to an existing bus network to improve service quality. A neural network has been used instead of transit

passenger trip assignment to evaluate fitness functions which includes waiting time and cost. Yu et al. (2005) developed an Ant Colony method to design bus network. The objective function is minimizing number of transfers and maximizing passenger flow with respect to bus line length constraint.

Although their results may not be the best, depending on the complexity of the problem, they are often close to the optimal solutions and found quickly. Since in most cases the optimal solution can never be found, the main advantage of using heuristic methods is finding good results in a short time.

III. BUS NETWORK DESIGN USING GA

GA is one of the most powerful metaheuristic methods, which is used in this paper to solve a complicated and large problem. Simplifying the problem is the main advantage of metaheuristic methods. For instance, the simplified objective function that is easy to calculate is considered in GA and saves the running time significantly. As a result, GA can search in feasible area faster to find better solutions. Following subsections explain how GA is used in this paper to design a bus network.

Using GA for Bus Network Design

In this research the population coverage is defined as a benefit of bus network. In this case, the best bus network is the one that covers the entire city. This, however, is not acceptable and practical because financial resources limitations do not allow for providing such a bus network. Therefore, the cost of bus network is considered as a function of the bus network length and the objective function is equal to the user benefits minus cost of the bus network which is attempted to be maximized. As mentioned, the user benefits of the bus network are defined as a function of population coverage of all bus lines. Therefore, if two bus lines overlap each other, the common population coverage is calculated in bus network benefit only once while, the cost of the bus network is calculated for the length of both of them which results in decreasing the objective function. The GA chooses the bus lines among candidate bus lines that maximize the objective function. Therefore, the algorithm selects the routes which have the following three characteristics:

- 1- Cover crowded regions of city.
- 2- Minimize overlap between different bus lines.
- 3- Minimize bus network length or bus network cost.

Candidate Bus Lines 130 bus lines were considered for the city of Mashad as candidate routes. These bus lines were identified in a preliminary analysis as potentially good routes for inclusion in the network. The GA was used to select the best subset of them as a designed bus network. Since all candidate bus lines are not selected in final solution, different routes for important origin-destination pairs were considered. In other words, by

using GA to choose best bus lines, it is beneficial to consider every route as a candidate route. Candidate bus lines were generated according to three policies: First, they cover all regions and main streets of the city. Second, at least one end point of each route is a park and ride. Third, bus lines service to suburban are not considered in the candidate routes because each suburban area must have one route to connect to the closest external parks and ride. In fact, there are no other meaningful choices for suburban areas.

Population coverage Population coverage for each station is defined as the number of people who are in an area around the station within a walkable distance to the station. Thus, this area can be considered as a circle which its center is on the station and its radius is equal to 1500 feet. It is possible that the coverage area of two stations have a common area depending on the distance between two stations. The city of Mashad has 141 zones with designated areas and populations. Therefore, the coverage area of a station may be located in more than one zone. The coverage area is calculated as follows:

1- If the coverage area of one station does not have common area with other stations' areas but is located in multiple zones, the common area between the station and each zone is calculated separately.

2- If the coverage area of two stations overlap each other and also they are located in multiple zones, not only the common area between each pair of station and zone is calculated, but also the common area of two stations with each of the zones is calculated.

Both cases are shown in figure 1. In figure 1-a, the circle is the coverage area of station A that has been located in zones one and two. A_1 is the common area between coverage area of station A and zone one. Similarly, A_2 is the common area between coverage area of station A and zone two.

a. One station b. Two stations FIGURE 1 Coverage area of one and two station(s) located in two zones.

In figure 1-b, stations A and B have a common coverage area and both of them are located in two zones. In this case, A_1 and A_2 are the common areas between coverage areas of station A and zone one and zone two, respectively. Similarly, B_1 and B_2 are the common areas between coverage area of station B and zone one and zone two. C_1 is the common area of stations A and B coverage area located on zone one. Likewise, C_2 is the common area of the same stations' on zone two.

The population coverage for each station as well as common population coverage between each pair of stations can be calculated by knowing the areas mentioned above and population distribution of each zone. As a result, the population coverage of each bus line is calculated according to equation (1):

$$Y_j = \sum_{i=1}^{n_i} X_i - \sum_{i=1}^{n_i} \sum_{\substack{k=1 \\ k \neq i}}^{n_j} Z_{ik}$$

Where

Y_j =Population coverage of bus line

X_i = Population coverage of station

Z_{ik} =Common Population coverage of station i and k

n_i =Number of stations of bus line

Then, the common population coverage for each pair of bus lines can be calculated as follows(2)

$$R_{ij} = \sum_{m=1}^{n_i} \sum_{n=1}^{n_j} Z_{mn}$$

Where

R_{ij} : The common population coverage of bus lines i and j

As a result, the total population coverage for a bus network can be calculated as shown in equation (3):

$$P_i = \sum_{j=1}^{n_i} Y_j - \sum_{j=1}^{n_i} \sum_{k=1}^{n_j} R_{jk}$$

Where

P_i : the total population coverage of bus network i ;

N_i : Number of bus lines exist in the bus network i ; and

S_i : The set of bus lines exist in bus network i

Definition of Chromosome and Gene Bielli, Caramia, and Carotenuto (2002) used GA for bus network design. They considered two genes for each bus line. The first gene represented frequency of the route and the second gene represented an on/off switch that enables or disables the use of that route in the corresponding network. They used outputs of transit assignment in the objective function; therefore, they should assign all bus networks separately in every iteration.

Objective Function in GA As mentioned before, the objective function defined in GA is equal to bus network benefit minus its cost. Bus network benefit is a function of its population coverage and bus network cost is a function of its length. In the following sections, calculations of bus network benefit and cost will be explained.

Bus Network Benefit Bus network benefit is defined as a function of decrease in transit travel time. The lower transit travel time, the better bus network. Decrease in the travel times for auto users was not considered in the objective function because demand of transit is independent of auto demand and changing in the bus network does not affect auto travel time significantly. Demand of transit contains two groups. The first group includes the trips completed by taxi if no bus line exists.

It can be assumed that for the first group, there is a balance between the cost of taxi and utility of travel time decreasing. In other words, users in the first group accept to pay more to get their destinations earlier. . Based on data of Mashad city, the average of trip length by bus is equal to 5 mile, the average of bus speed is equal to 10 mile/hour, and the average of pedestrian speed is equal to 3 mile/hour (The Suggested Transit

System for Mashad 2003). As a result, decrease in travel time by using bus is equal to:

$$qd(5/3 - 5/10) = 1.16qd(\text{Passenger-hour})$$

In section 3-1-2 the bus network population coverage was calculated while the benefit of the bus network is the function of transit demand. Therefore, the bus network population coverage should be converted to the transit demand. According to data of Mashad city, each person does 1.6 trips per day, the share of bus mode from daily trips is equal to 20%, and the share of peak hour from daily trips is equal to 10% (The Suggested Transit System for Mashad 2003). Therefore, if represents the population coverage of bus network , the demand of bus in peak hour is equal to:

$$1.6 \times 0.2 \times 0.1 \times P_i = 0.032 P_i (\text{Passenger})$$

Moreover, the value of time in Iran in 2001 was equal to 4 dollar per passenger-hour and according to transportation studies in Mashad, is equal to 0.45 (The Suggested Transit System for Mashad 2003). As a result, the benefit of the bus network is equal to:

$$4 \times 0.45 \times 0.032 \times 1.16 \times P_i = 0.0668 P_i (\text{Dollar})$$

Bus Network Cost To calculate bus network cost, the data of Mashad bus network in 2003 were referred. According to these data, the price of a new bus is equal to 60 thousand dollars. Moreover, the operational cost is roughly 60 percent of the fixed cost which is considered as a vehicle fixed cost. Based on fleet database, the bus lifetime is about 15 years. According to

one of the suggested bus networks for Mashad, the network needs 780 buses and it is 2736 mile long (The Suggested Transit System for Mashad 2003). Therefore, the operation cost in the first year is equal to $0.6 \times 780 \times \$60000 = \28080000 and the total cost is equal to $\$28080000 \times 780 \times \$60000 = \$74880000$. Therefore, by assuming that interest rate is equal to 10 percent, which is acceptable for the economy of Iran, the uniform annual total cost per 15 years is equal to $\$74880000 \times (A/P, 10\%, 15) = \9844756 .

Crossover and Mutation Probability The probability of crossover is set as 0.7 and the probability of mutation is set as 0.3. In addition, in elitism operation the five best bus networks in each population transfer to the next generation directly. Therefore, the objective function of best solution in each generation is not less than that of the previous generation.

Transit Assignment of the Results with EMME/2 Since the benefit and cost of bus network considered in the GA objective function are calculated approximately and several criteria have not been included in the objective function, the results need to be evaluated more carefully. The EMME/2 software is used to assign the transit demand on the bus networks resulted from the application of the GA and find the performance measures associated with each bus network. Ten performance measures have been selected for evaluation. These ten performance measures are

categorized into three groups. The first group is the system operator cost which has three measures: fleet size that is the representative of capital cost, bus cumulative travel distance (bus-km) which is the representative of maintenance cost, and bus network length which is the representative of administrative cost.

The second group is bus network benefit and contains only one measure which is the real demand of transit traveled with bus lines. The third group is user cost and it is divided into two sub groups. The first sub group is transit utilization cost and contains four measures calculated only for transit users: the cumulative passenger travel distance (passenger-km), the

cumulative passenger walking distance (passenger-km), the cumulative passenger travel time (passenger-hour), and the cumulative passenger waiting time (passenger-hour).

Fourteen weighting systems for different performance measures to be used in the Concordance Analysis have been considered which are shown in Table 1. The Concordance Analysis is a multi-criteria decision making method and recognizes rival alternatives. In other words, it gets dominant alternatives, which are either one or more than one, for each weighting system. Therefore, the alternatives that are selected as dominant choices in more weighting systems are rival alternatives.

Index No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1-1	0.11	0.083	0.083	0.066	0.066	0.055	0.055	0.11	0.083	0.083	0.066	0.066	0.055	0.055
1-2	0.11	0.083	0.083	0.066	0.066	0.055	0.055	0.11	0.083	0.083	0.066	0.066	0.055	0.055
1-3	0.11	0.083	0.083	0.066	0.066	0.055	0.055	0.11	0.083	0.083	0.066	0.066	0.055	0.055
2-1	0.33	0.5	0.25	0.6	0.4	0.5	0.33	0.33	0.5	0.25	0.6	0.4	0.5	0.33
3-1-1	0.055	0.041	0.083	0.033	0.066	0.055	0.083	0.036	0.027	0.056	0.022	0.044	0.037	0.056
3-1-2	0.055	0.041	0.083	0.033	0.066	0.055	0.083	0.07	0.055	0.011	0.044	0.088	0.073	0.011
3-1-3	0.055	0.041	0.083	0.033	0.066	0.055	0.083	0.036	0.027	0.056	0.022	0.044	0.037	0.056
3-1-4	0.055	0.041	0.083	0.033	0.066	0.055	0.083	0.074	0.055	0.011	0.044	0.088	0.073	0.011
3-2-1	0.06	0.043	0.084	0.035	0.069	0.057	0.086	0.06	0.043	0.084	0.035	0.069	0.057	0.086
3-2-2	0.06	0.043	0.084	0.035	0.069	0.057	0.086	0.06	0.043	0.084	0.035	0.069	0.057	0.086

IV. RESULT:

Table 2 shows the results of applying this method on the city of Mashad. Fifteen feasible and good solutions with the same number of GA iterations were obtained. One of the most important specifics of this code is that the seed of random number generation depends on the time of the computer. Therefore, many useful results can be got not only by running the code on different computers, but also by running the code on the same machine at different times. The 15 solutions shown in the table 2, are dominant results among many runs on several computers. In the table 2, the number of interior bus routes, bus network length for interior routes, and GA objective function are shown.

15	66	564	920.44
----	----	-----	--------

Solution Number	Interior Route Number	Interior Bus Network Length (Mile)	Objective Function (Dollar)
1	72	611	991.85
2	65	551	1133.25
3	51	429	1191.70
4	72	643	829.18
5	65	574	1011.77
6	69	613	765.48
7	66	554	1151.55
8	72	599	1058.22
9	68	616	848.81
10	59	514	1309.03
11	51	444	1266.96
12	72	634	799.03
13	66	561	952.00
14	55	449	1463.25

The GA objective functions for these eight solutions do not have the maximum values, because it represents the profit of the bus network approximately and does not include all criteria. In fact, a simplified objective function was used in GA in order to search for more solutions in shorter time, and this is a key point of using metaheuristic methods such as GA. As a result, solution number two, six, and thirteen, selected fourteen times, are the best solutions, where solution number sixteen was selected twelve times. In other words, solutions obtained by GA are better than that proposed by the Sharif University of Technology Transportation Center which has considered all of the criteria. In addition, solutions number four and nine have been selected eleven times and solutions number eleven and twelve have been selected ten times.

V. CONCLUSIONS

In this paper, a new approach to design a bus network for City of Mashad was introduced. This approach uses GA to select good solutions among candidate solutions. In fact, it selects dominant subsets from a set that has n elements. It does not have any limitations on the number of bus lines. So, every suitable bus line from different standpoints can be considered as a candidate route and GA selects the best subset of them. GA objective function has two criteria: population coverage and length of a bus network. Also, capital cost and maintenance cost for the whole period of bus life have been defined as a function of bus network length. GA tries to maximize the objective function and therefore, tries to find the bus network which has the maximum

population coverage and the minimum length. To provide a good spread across the city, constraints on the number of bus lines which enter or exit each park and ride location have been added.

Suggestions for Future Studies

Some avenues for future studies are as follows:

1- Path for each route was designed manually. Using GA to design the paths increases the chromosome length and increases runtime. Using other methods to design paths and then using GA to find the best bus network may increase the efficiency of this method.

2- Other properties of zones (such as zone demand) can be used in the objective function instead of population and this may increase the accuracy of results.

3- A uniform distribution of population over the zones was assumed. This helps to simplify the calculations. Using accurate forms of population distribution in each zone will increase the accuracy of results.

REFERENCES

- [1] Bansal, A N. "Optimization of Bus Route Network for a Fixed Spatial Distribution." *Scientific Management of Transportation Systems*. Amsterdam: North Holland Publishing Company, 1981. 346-355.
- [2] Bielli, M, M Caramia, and P Carotenuto. "Genetic Algorithms in Bus Network Optimization." *Transportation Research Part C: Emerging Technologies*, Vol. 10, Issue 1, 2002: 19-34.
- [3] Ceder, A, and H M Wilson. "Bus network Design." *Transportation Research*, Vol. 20B, No. 4, 1986: 331-334.
- [4] Dubois, D, G Bell, and M Libre. "A Set of Methods in Transportation Network Synthesis and Analysis." *Operational Research*, Vol. 30, 1979: 797-808.
- [5] Dufourd, H, M Gendreau, and G Laporte. "Locating a Transit Line Using Tabu Search." *Transportation Science*, Vol. 4, 1996: 1-19.
- [6] Guan, J F, H Yang, and S C Wirasinghe. "Simultaneous optimization of transit line configuration and passenger line assignment." *Transportation Research Part B*, Vol. 40(10), 2003: 885-902.
- [7] Hurdle, G. "Minimum Cost Location for Parallel Public Transportation Lines." *Transportation Science*, Vol. 7, 1973: 341-350.
- [8] Kocure, G, and C Hendrickson. "Design of Local Bus Service with Demand Equilibrium." *Transportation Science*, Vol. 16, 1982: 149-170.
- [9] Lampkin, W, and P D Saalman. "The Design of Routes Service Frequencies and Schedules for Municipal Bus Undertaking." *Operational Research Quarterly*, Vol. 18, 1978: 375-397.
- [10] Lines, A H, W Lampkin, and P D Saalman. "The design of Routes and Service Frequency for a Municipal Bus Company." *Business Operation Research LTD*, London, U.K., 1966.
- [11] Mandle, C E. "Evaluation and Optimization of Urban Public Transportation Network." *European Journal of Operational Research*, Vol. 5, 1979: 396-404.
- [12] Newell, G. "Some Issues Relating to the Optimal Design of Bus Routes." *Transportation Science*, Vol. 13, 1979: 20-35.
- [13] Silman, L A, Z Barazily, and U Passy. "Planning the Route System for Urban Buses." *Operational Research*, Vol. 1, 201-211.
- [14] The Suggested Transit System for Mashad. Tehran: Sharif University of Technology Transportation Center, 2003.
- [15] Yu, B, Z Yang, C Cheng, and C Liu. "Optimizing bus transit network with parallel ant colony algorithm." *Eastern Asia Society for Transportation Studies*, Vol. 5. 2005. 374-389.
- [16] Zhao, F, and A Gan. Optimization of transit network to minimize transfers. Tech. Rep. BD015-02, Florida Department of Transportation, Center for Transportation Research: Florida International University, 2003.
- [17] Zhao, F, and I Ubaka. "Transit network optimization – minimizing transfers and optimizing route directness." *Journal of Public Transportation* 7 (1), 2004: 67-82.
- [18] Zhao, F, and X Zeng. "Simulated annealing–genetic algorithm for transit network optimization." *Journal of Computing in Civil Engineering* 20 (1), 2006: 57-68.

