Railroad Yard Location Optimization Using A Genetic Algorithm

AVIJIT MAJI
Office of Traffic and Safety
Maryland State Highway Administration
7491 Connelley Drive, Hanover, MD 21076
USA
avijit.maji@gmail.com

MANOJ K. JHA
Department of Civil Engineering
Morgan State University
1700 East Cold Spring Lane
Baltimore, Maryland 21251
USA
manoj.jha@morgan.edu

Abstract: **-** Freight trains use railroad network to transport goods from an origin to a destination. The small consignments are accumulated and arranged in a nearby marshal yard and transported to the destination marshal yard for delivery. Strategy to choose these marshal yards plays an important role in freight traffic control, management and economy. A shortest path algorithm combined with a genetic algorithm optimization technique is developed in this paper to identify the potential locations of the marshal yards in a given railroad network based on minimization of transportation cost.

Key-Words: **-** Yard location, genetic algorithm, Floyd's algorithm, railroad network, shortest path algorithm, optimization.

1 Introduction

Freight trains use existing railroad network to haul freight cars from an origin to a destination. The hauling incurs cost to the service provider. This cost is dependent on the distance covered by the freight train or the travel time, number of freight trains and the number of freight cars being used for the transportation. The number of freight trains is again dependent on the maximum number of freight cars that can be hauled by a locomotive. Sometimes the number of freight cars shipped from the origin to the destination is not enough to run the freight train with maximum or full capacity. This leads to under capacity movement of the freight train and increase in the operation cost. To reduce the operation cost, the under capacity freight trains from nearby origins are gathered at one location, accumulated and rearranged to form a train of capacity and hauled to another nearby common location to their destinations. These points of common locations are known as yards. In this study, the yard close to the origins is termed as origin yard for those set of origin locations and the yard close to the destinations is termed as destination yard for those set of destination locations. Construction and maintenance of these yards also incur cost to the service provider. As the number of yards increases, total hauling cost may decrease up to a certain extent but the yard construction cost will increase proportionately. Therefore, the selection of the yard location is a crucial decision. This can be done by optimizing the cost of hauling and construction of yards together.

The decision of selecting yard locations raise following questions:

- Which locations can be considered as the potential origin and destination yards?
- Which origins can be grouped for a particular origin yard?
- Which destinations can be grouped for a particular destination yard?
- How should the routes be oriented?
- How many yards are necessary to minimize the overall cost of operation?

The answers to these questions are interrelated. They need to be addressed together not separately. Therefore, we need to understand the railroad network, the shortest path algorithm and the optimization process involved in the solution methodology.

2 Railroad Network

Railroads run width and breadth of a country and form a network. The railroad network as any other network can be defined as a combination of set of nodes of branches [1]. The nodes signify the point where two or more rail-line merges or diverges. These nodes are interconnected by branches denoted by block-sections or links. In general, the railroad network can be represented as a graph with relationship between the connecting nodes. The impedance between the nodes defines the relationship. It can be in terms of travel cost or travel time and so forth. In this study symbolically the network is represented as G(N, A), where N denotes the nodes and A denotes the links and the travel time as the impedance to the links. Every node in this is a potential yard location. It depends on the minimization result which nodes will be ultimately considered as the yard location. An example railroad network is shown in Fig. 1. For convenience the network is shown in square grid but in practical it will not be in a square grid. The freight cars starting from their origin are hauled through the nodes to their destination. Among these intermediate nodes any pair of nodes can be the origin and destination yards of the freight cars.

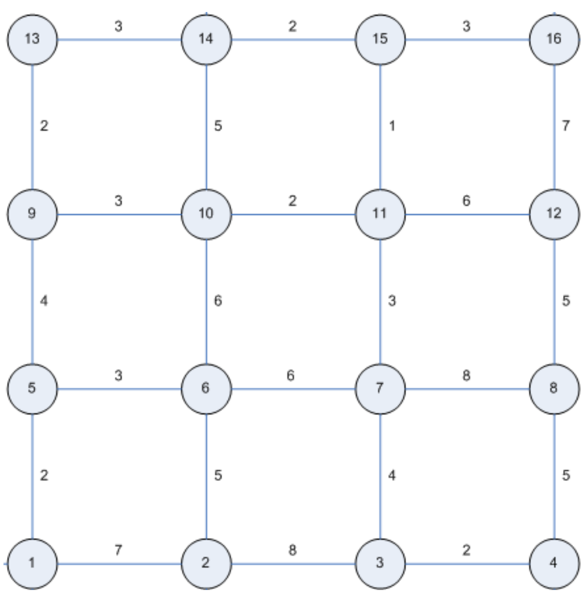

Fig. 1: A railroad network

In this figure the circles are the nodes and the number inside them are the unique number to identify them. The lines connecting the nodes are the links of the network and the numeric value on these links are the impedance for the connecting nodes. This impedance is in the form of travel time. The links in this figure do not have any direction. This type of network is known as non-oriented network and it resembles two-way railroads.

3 Problem Formulation

Let N_{ij} be the number of cars need to be hauled from origin 'i' to destination 'j'. If the maximum number of cars that can be hauled by a locomotive is N_{max} then the number of trains required to haul Nij cars is given by:

$$T_{ij} = \text{Round up to higher integer of } (N_{ij}/ N_{max})$$

If the fixed cost to haul a train is F and the variable cost per cars is V to haul for unit time then the total cost to haul N_{ij} is given by:

$$\text{Total cost} = (F\, T_{ij} + V\, N_{ij}) \text{ x Travel time}$$

Calculation of travel time is another optimization process where we have to consider either the train will move through the nearest yards or directly from origin to destination. If the cost to haul the train directly from origin to destination is cheaper compared to hauling it through the nearest yards then the trains should be hauled directly from origin to destination. If the train is hauled through the yards then the train is first hauled to nearest yard. There, other trains are accumulated and new trains are formed of same destination or destination yard. Then, those trains are hauled to their respective destination locations. Therefore, the yard locations are the critical parameters in this problem which need to be found out. The travel time can be considered as the function of the yard location. If the location can be presented by 1 or 0, where 1 represents the location is selected as a yard location and 0 for not selecting as a yard. The objective of this problem is to minimize the total cost and it can be presented as follows:

Min

$$\sum_{j=1}^{n} \sum_{i=1}^{n} (F\,T_{ij} + V\,N_{ij}) \text{ x Travel time} \qquad i \neq j$$

Where,

$$\text{Travel time} = f(\mathit{Yard\ Location})$$

$$\mathit{Yard\ Location} = \begin{cases} 1 \text{ if location is selected} \\ 0 \text{ if location is not selected} \end{cases}$$

ISSN: 1790-5095

ISBN: 978-960-474-136-6

Freight trains routes are chosen in such a way that they costs least. The path which yields least cost for a pair of origin and destination is known as shortest path. Several algorithms are available to calculate and identify the shortest path in the road network. One of the most efficient algorithms to find the shortest path from one node to all other nodes in a network was given by E.W. Dijkstra in 1959 [2]. But in practical transportation problem it is often required to calculate the shortest path between all pairs of nodes in the network. So, a special type of algorithm developed by R.W. Floyd in 1962 [3] is used to find the shortest path between all pairs of nodes. This algorithm is discussed in detail in the following section.

4 Floyd's Algorithm [3]

The Floyd's algorithm starts by denoting all the nodes with unique integers 1, 2, 3…, n. Then the initial matrix for shortest path costs Do and predecessor matrix Qo are formed. If there exists a link between a pair of nodes then the initial cost is the minimum cost of the links connecting those two nodes. If d_{ij}^{0} denotes the cost of the shortest path from node i to node j at iteration 0 then the values of the initial matrix will be as follows:

$$d_{ij}^{0} = \min\left[l_1(i,j), l_2(i,j), \ldots\ldots, l_m(i,j)\right] \qquad (1)$$

where,

$l_m(i,j)$ = length of m^th^ link between node i and node j

So the value of d_{ij}^{0} = 0 when i = j and if there are no branches between the nodes then mathematically the cost is considered to be infinity ($d_{ij}^{0} = \infty$).

For the initial predecessor matrix the values are considered as follows:

$$q_{ij}^{0} = \begin{cases} i, & \text{for } i \neq j \\ -, & \text{otherwise} \end{cases} \qquad (2)$$

After defining the initial matrices followings steps need to be followed to obtain the shortest path between all pair of nodes:

Step 1: Consider k = 1

Step 2: Each element d_{ij}^{k} of the shortest path matrix is obtained using the following equation:

$$d_{ij}^{k} = \min\left[d_{ij}^{k-1}, d_{ik}^{k-i} + d_{kj}^{k-1}\right] \qquad (3)$$

Step 3: The elements of the predecessor matrix is calculated as follows:

$$q_{ij}^{k} = \begin{cases} q_{kj}^{k-1}, & \textit{for } d_{ij}^{k} \neq d_{ij}^{k-1} \\ q_{ij}^{k-1}, & \textit{otherwise} \end{cases} \qquad (4)$$

Step 4: If k = n, then the algorithm is completed, else if k<n, then k is increased by 1, i.e. k = k + 1 and the whole process is repeated from step 2.

Sample shortest path information for a pair of origin and destination obtained by Floyd's algorithm is shown in Fig. 2. In the figure the links with double line arrow shows the shortest path to reach the destination node 19, 20 and 23 from the origin node 22. Any other path joining node 22 to node 19 or 20 or 23 will cost more to the user.

Along with this algorithm optimization tool, Genetic Algorithm (GA), is integrated to find the optimum number and location of the yards. The GA used for this study is a modified from the conventional process. Details of GA are discussed in the following section.

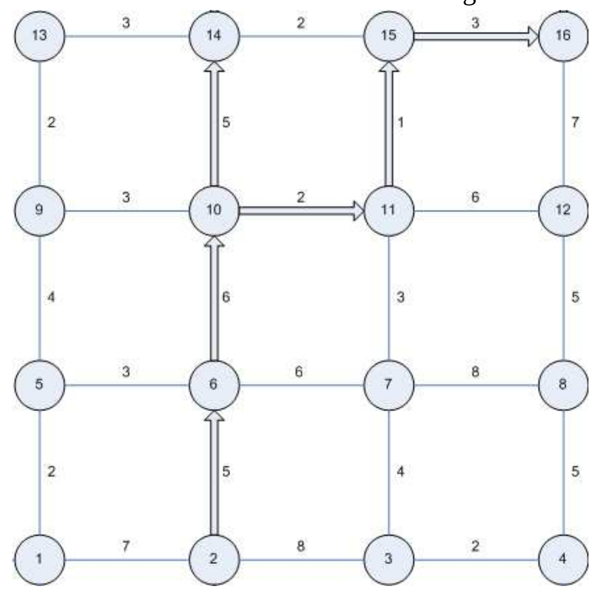

Fig. 2: Shortest path from node 2 to node 6, 10, 11, 14, 15 and 16

5 Genetic Algorithm

GA is one of the nontraditional evolutionary optimization algorithms envisaged by Professor John Holland of University of Michigan, Ann Arbor in the mid sixties [4]. This algorithm mimics the natural evolution of genetics to constitute the search for the optimum value of a mathematical function [4]. Similar

to gene where genetic properties are stored, the probable candidate solution of the variables is coded in a string structures. This particular problem is about finding the optimal location of the yards and the variables which govern the location have values either 1 or 0. The value 1 stands for the location selected as a yard and the value 0 otherwise. Therefore, the variable values can be directly used in the gene and their position can correspond to the unique node identification number.
The yard locations in the railroad network don't have any regular pattern. Therefore, only mutation is used as the evolution technique in the solution process. To expedite the solution process, the variable value of the corresponding locations which don't qualify as a yard in a particular solution is changed to 0. Following this in the calculation process will eventually mimic the Darwin's principle of "survival of the fittest". Details of application of the algorithms in the solution process is discussed the following section, methodology.

6 Methodology

The problem of finding the appropriate number and location of yards is solved using shortest path algorithm and GA. The shortest path algorithm is used to find the route from the origin to destination through yards if any. The GA is used to find the number and location of the yards. Both of these methods are used simultaneously here. The chromosome of GA is used to define the location of the possible yards. Each chromosome's allele of the GA used here represents the nodes in the network and they are coded with 1 or 0. The positions of allele in the chromosomes are considered as the unique node number of the network. On the other hand, the value 1 of allele represents the node as a yard location and the value 0 represents the node not considered as a yard location. The positions of allele and their values are determined randomly using uniformly distributed random numbers.

Initially a set of parent chromosomes are created as the initial parent solution pool and cost to haul the freight cars to their destination through the origin yard and destination if any is calculated. The calculation process starts by finding the nearest origin and destination yard location for the pair of origin and destination of freight cars. The shortest path information obtained from Floyd's algorithm is used to select the route from origin to nearest origin yard, nearest origin yard to a possible destination yard and from there to the final destination. This process is repeated for all sets of origin and destination pairs. Now the freight cars from its origin are assumed to haul to its origin yard. There all cars from all other nearby origins destined to same location are grouped together to form a freight train of capacity. If the number is not sufficient to form a train of capacity then the train is made as close as to the capacity. The allele corresponding to yard locations which didn't qualify or not considered in the process are changed to value 0. Also the pair of origin and destination yard locations which connects only one pair of origin and destination is also eliminated from the consideration list. The freight cars in this case are assumed to take the shortest route of the origin and destination pair. After this, the cost to haul the freight cars from origin to origin yard, origin yard to destination yard and destination yard to destination or directly from origin to destination is calculated depending on the situation. The cost is calculated based on the shortest path information between the pair of origin and destination through the origin destination yards if any. The duplicate sets of solutions are removed from the parent pool of solution. After that the solutions are sorted and they are ready to create child alternative solutions.

In this study child set of alternatives are created by mutating the parent set of solution. The parent set of solutions are considered one by one. The present yard location from the solution is swapped with another randomly chosen location not considered as yard in the solution. This is done for all sets of parent solution and in turns it creates a complete set of possible child solution. Then the cost to haul the freight cars from the origin to their respective destination either through the origin, destination yards or directly is calculated. This calculation is similar to that of the parent set of solution. Here also the shortest path information is used for the calculation. The obtained child solution set is compared with the parent set as well as among itself and the duplicate solutions are removed from the child set of solution. After this the present parent solution pool is updated with best solutions from the parent and child solution pool. With the new set of parent solution, the whole process of creating child set of alternatives is repeated till the parent solution is no longer updated in the process.

7 Problem Solution

A simple network of sixteen nodes (Fig. 1) is considered as a case study for this problem. This link travel time information is used to obtain the shortest path information using Floyd's algorithm as described before. In this study the maximum capacity of a freight train is considered to be thirty freight cars. The variable cost to haul a freight car is one unit per hour and the fixed cost for a train is hundred units per hour. The construction and maintenance cost of a yard is considered to be five hundred units and details of all considerations are shown in table 2. The number of freight cars need to be hauled from the origin to their respective destination is shown in table 1.

Table 1: Freight car demand

Origin	*Destination*	*No. of cars*
1	11	80
1	16	100
2	12	50
2	14	50
2	16	100
3	6	100
5	9	70
5	16	25
9	4	10
9	5	13
10	1	40
12	9	20
15	3	18
15	7	10

Table 2: Costs considered for the study

Type of cost	*Amount (Unit)*
Hauling cost of a car	1 per unit time
Fixed hauling cost of a train	100 per unit time
Construction and maintenance cost of yards	250 each location
Hauling capacity of a locomotive	30 freight cars

Table 3: Obtained original destination schedule

Origin	*Destination*	*No. of cars*
1	10	180
2	7	200
3	6	100
5	9	70
5	10	25
7	3	18
7	4	10
7	11	200
9	5	13
9	10	10
10	1	40
10	7	10
10	9	20
10	11	205
11	7	28
11	10	20
11	12	50
11	14	50
11	16	225
12	11	20
15	11	28

Analyzing this information with the methodology developed, node 7, 10 and 11 of the network is obtained as the optimum yard location. So, the optimum number of yards required for this hypothetical problem is three. If no yard has been considered then the total hauling cost would have been 43891 units but the freight cars redirected through origin and destination yard reduced the hauling cost to 42341 units. This reduced cost also includes the cost incurred for the yards. The optimized original destination schedules of the freight cars are given in table 3.

The schedule in table 3 describes how the freight cars will be hauled from their origin to destination through the intermediate yards if any. For example freight cars from node 1 forms six full capacity train of 180 freight cars and hauled to node 10. From there along with other freight cars seven trains are formed. Six are of full capacity and one with 25 freight cars, 5 less than full capacity making a total of 205 cars. They are then hauled to node 11. From there along with other freight cars originated from other nodes, 8 trains of total 225 freight cars are formed and hauled to node 16. One train with 15 cars and rest are of full capacity trains. In this process 80 cars are hauled from node 1 to node 11 and 100 cars from node 1 to node 16. In the same process cars from other origin are hauled to their destination. The freight car movement schedule in the network is shown in Fig. 3.

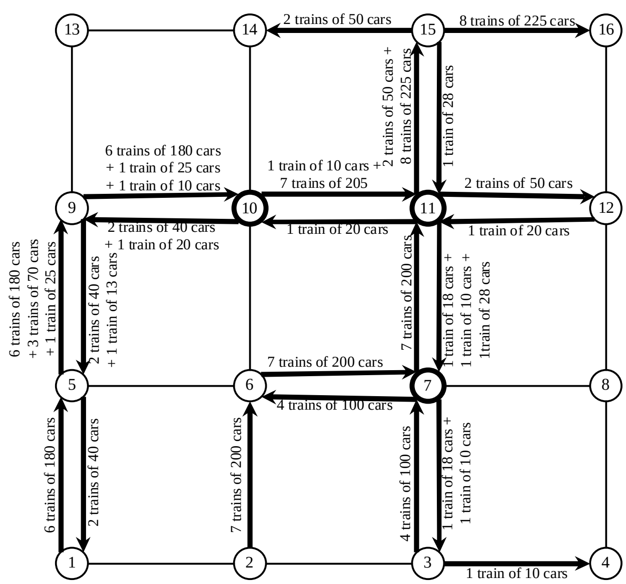

Fig. 3: Freight car movement schedule

8 Conclusion

The methodology proposed here is capable of finding yard locations for any size of rail-road network. The

ISSN: 1790-5095

ISBN: 978-960-474-136-6

solution depends mostly on the different costs considered in the study, the origin destination locations, number of freight cars and of course on the network considered. Therefore the locations may change with the variation of those factors. At present the process discussed here to find the yard location has certain limitations in the calculation process. They are discussed as follows:

- The model is not tested with real data, all information were hypothetical.
- Capacities of the yards are not considered in the study.
- Yard time is not considered in the study.

These limitations can be incorporated in the model in future works, which will make the calculation more robust.

References:

[1]. Teodorovic, D. (1986). Transportation networks, Gordon and Breach science publishers, New York, US.

[2]. Dijkstra, E.W. (1959). "A Note on Two Problems in Connexion with Graphs," Numerische Mathematik, 1(1), 269-271.

[3]. Floyd, R.W. (1962). "Algorithm 97: Shortest path," Communications of ACM. 5(6), 345.

[4]. Deb, K. (1995). Optimization for Engineering Design: Algorithms and Examples, Prentice-Hall, New Delhi, India.