CHAPTER II

LITERATURE REVIEW

Mass Rapid Transit (MRT) systems are major transportation that serving millions of population in most of countries around the world. Wong, K.K & Ho, T.K.[3] explained to maximise the capacity of the rail line and provide a eliable service for passengers throughout the day, regulation of train service to maintain steady service headway is essential. On their research they presents an application of classical measures to search for the appropriate coasting point to meet a specified inter-station run time and they can be integrated in the on-board Automatic Train Operation (ATO) system and have the potential for on-line implementation in making a set of coasting command decisions. The main idea is to identify the coasting point for real time train scheduling control. Two bi-section methods, Golden Section and Fibonacci search, are highlighted, and the idea of how to fix the necessary coasting point with the gradient method is also presented.

The Golden section search algorithm on coasting control to regulate the train schedule, the fitness of two initial coasting points are determined in advance. These two values will then be useful for further search of new coasting point. The basic idea of the Golden section search is that the solution space is divided into two unequal parts, the ratio of the larger of two segments to the total length of the interval should be the same as the ratio of the smaller to the larger segment. If the coasting points are placed with this fractional spacing from either end on the solution space, the solution space will then be reduced to a length of 0.618 times the previous of uncertainty of interval. It is obvious that one of the two evaluations of coasting point is available for the next step by the virtue of the golden ratio. Thus only one additional “golden-spaced” evaluation is required to reduce the solution space by 0.618 fraction. The process is repeated again and the uncertainly of interval is further reduced by the golden ratio until the obtained coasting point satisfies the expected train operational requirement in an inter-station run. To achieve the maximum reduction in the subsequent solution space, the evaluations should be placed symmetrically about the centerline of the solution space. When this is done, a new evalutaion provides an additional reduction in the solution space.

The Fibonacci search is very similar to the Golden section search. The main different os the reduction ration on the solution space in each iteration is fixed at 0.618 with the Golden search, whilst the reduction ratio varies with the previous uncertainty of interval in Fibonacci search. Fibonacci search provides a better reduction ratio on the solution space in each iteration if the maximum number of iterations is predetermined in advance. The gradient method uses the derivative as illustated in equation II-1 to locate the necessary coasting point. Similarly, the fitness of two initial coasting point are predetermined and the search direction of the updated coasting point depends on the polarity and the magnitude of the gradient.

Gradient $=\frac{\Delta Run\ Time}{\Delta\ Coasting\ Location}$ ....(II-1)

Then the step length can be calculated by,

Step length = $\frac{Runtime_{flat-out} - Runtime_{Expected}}{Gradient}$ ....(II-2)

And the new coasting point = Old coasting point + Step length……………………………...(II-3)

In general, the step length becomes larger when the run-time of the current coasting location is far away from the expected one, or vice versa.

Feng, Xuesong explained the minimalization of the energy saving by optimization of target speeds. In this paper[6] explained according to Hay (1982), Andrews (1986), Mao,et al. (2008), when the maximum speed is raised, higher traction effort is required which implies different patterns of energy consumption. It is then necessary to investigate how the traction enegy cost of a High-speed railway (HSR) train changes under a range of taget speeds, taking into account the effect of inter-stop transport distances, traction characteristics of HSR trains and gradients, curvatures, etc. of the rail lines. In particular time-saving may be achieved by various target speeds and the cost to attain such an improvement through high-speed operation should be evaluated. Furthermore, when the target speed of a HSR train is connected to both energy saving and transport efficiency, the setting of its traget speed to improve both of them is essential to the service quality and operation cost. The difference in this research is the ability to provide quantitative evaluations of both energy and time saving with respect to target speeds of trains under different inter-stop transport distances, traction equipment characteristic and rail lines gradients, curvature,etc. By analyzing the effect of these factores on the Traction Energy Cost (TEC) and Technical Operation Time (TOT) per 10,000 passenger-kilometres (p-km), Feng propose to optimize the target speeds of HSR trains in a quantificational manner from the perspectives of both traction energy saving and transport efficiency improvement.

The whole trip of the train from one stop to the next is simulated in successive intervals. The lengths of the calculation intervals are equal to 1 s in the simulation work of this research. The traction force, speed and operating condition (i.e. motoring, coasting or braking) of the train are considered to be unchanged in one calculation interval. The train at a station is started up with its full traction power towards the target speed. With the first achievement of the target speed by the sustaining acceleration of the train from its startup, the train commences to coast till the difference between its speed and the target speed reaches a pre-set value, which is 10.00 km/h in this work, thereafter accelerate with its full traction power to the target speed alternately. In order to ensure accurate and safe stopping at stations, the train begins to decide whether brakes are necessary or not in a calculation interval when the train arrives at a rail site where there is a certain distance away from the next stop. This is decided according to the speed (v1) of the train and the permitted speed (v2), which is determined based on the braking performance of the train and the transport distance from the site of the train at the beginning of this calculation interval to the next stop.There are 2 condition :

1. If v1 ≥ v2, the train brakes to decrease its speed as soon as possible to a comparatively very small value, which is able to ensure absolute safety of its stop in the next station.
2. If v1 < v2, the train coasts.

Such decisions are made for each of the latter calculation interval still the train accurately stops in safety in the next station according to the v1 of the train in each latter calculation intervaland the v2, which is determined based on the location of the train at the beginning of each latter calculation interval and the braking performance of the train.

The traction force of a train utilizing a certain ratio of its full traction power in a calculation interval is first determined by the speed of this train, as expressed by equation II-4.

$$f_k^r = \frac{P_k^r}{V_k^r} \qquad \text{(II-4)}$$

Where :

1. f_k^r is the traction force of the train utilizing r% of its full traction power in the kth calculation interval,unit is N.
2. P_k^r is the traction power of the train utilizing r% of its full traction power in the kth calculation interval,unit is W.
3. V_k^r is the speed of the train utilizing r% of its full traction power in the kth calculation interval,unit is m/s.

Besides the speed, the traction force of a train is also affected by its operating condition. When a train is coasting or braking, its traction force is 0N. As a result, the traction force of a train utilizing r% of its full traction power in a calculation interval is able to be generally interpreted by equation II-5.

$$f_k^r = \begin{cases} \frac{P_k^r}{V_k^r} \text{ if } \left(v^{tm} - v_{k-1}^{pr}\right) > C^{tm} \text{ } or \text{ } \left(v_k^{ul} - v_{k-1}^{pr}\right) > C^{ul} \\ 0 \text{ if } \left(v^{tm} - v_{k-1}^{pr}\right) \leq C^{tm} \text{ } or \text{ } \left(v_k^{ul} - v_{k-1}^{pr}\right) \leq C^{ul} \end{cases} \quad \text{...(II-5)}$$

Where :

1. $v^{tm}$is the target speed of the train, unit is m/s.
2. v_k^{ul} is the upper speed limit in the k th calculation interval, which is equal to v^{tm} when there is no requirement by the track geometry of the rail line, unit is m/s.
3. $v_{k-1}^{pr}$is the speed of the train utilizing pr% of its full traction power in the (k-1)th calculation interval, unit is m/s.
4. $C^{tm}$is the permitted maximum difference between speed of the train and the target speed, unit is m/s.
5. C^{ul} is the permitted maximum difference between speed of the train and the upper speed limit, which is equal to C^{tm} when there is no requirement by the track geometry of the railline, unit is m/s.

As illuminated by equation II-4 and II-5, the speed of a train in a calculation interval is decided by the traction force of the train utilizing some proportion of its full traction power, according to the target speed of the train, the upper speed limit required by the track geometry of the rail line in this calculation interval, the speed of the train in the previous calculation interval, the mass of the train and the resistance force from the rail line can be seen in equation II-6.

$$V_k^r = v_{k-1}^{pr} + \frac{f_k^r - f_k^L}{M} \Delta t \quad \text{...(II-6)}$$

Where :

1. $f_k^L$is the resistance force from the rail line in the kth calculation interval, unit is N, which is explained by equation II-7.
2. M is the mass of the train, unit is kg.
3. Δt is the equivalent length of the calculation intervals ,i.e.1.00s.

$f_k^L = f_k^B + f_k^S$ ....(II-7)

Where :

1. $f_k^B$is the basic resistance force from the railline in the kth calculation interval, unit is N, which is given by equation II-8.
2. f_k^S is the special resistance force from the gradient, curvature, etc. of the railline in the kth calculation interval,unit is N.

The special resistance force from e.g .the gradient of the rail line in the kth calculation interval is explained by equation II-8.

$f_k^B = \alpha_0 + \alpha_1(V_k^r) + \alpha_2(V_k^r)^2$ ....(II-8)

Where $\alpha_0, \alpha_1, \alpha_2$ are the coefficients determinded by the traction characteristics of the train. Different types of trains have their own sets of such coefficients :

$f_k^{Gradient} = Mgsin$□ ....(II-9)

Where :

1. $f_k^{Gradient}$ is the special resistance force from the gradient of the rail line in the kth calculation interval, unit is N.
2. g is the acceleration of gravity, unit is N/kg.
3. θ is the angle between the level and the slope, unit is degree.

The TEC of a train for a calculation interval is computed by equation II-10, based on the train's energy cost curves presenting detailed numerical values of the energy cost intensities of the train with different traction forces, speeds and operating conditions. Various types of trains have their respective sets of energy cost curves revealing distinct energy cost intensities of these trains with the same traction force, speed and operating condition. The TECs of all the calculation intervals of the simulation work from the startup of a train at a station to its stop at another station are summed into the TEC of the trip between these two stops. The TOT of the train between two stops is calculated by summation of the calculation intervals.o_k

$TEC_k = E_k^{f_k^r . V_k^r . o_k} \Delta t$ ....(II-10)

Where :

1. TEC_k is the TEC of the train for the kth calculation interval, unit is kWh.

2. $E_k^{f_k^r.V_k^r.o_k}$is the energy cost intensity of the train utilizing r% of its full traction power for the traction force of f_k^r, the speed of V_k^r and the operating condition of o_k (i.e.motoring,coasting or braking) in the kth calculation interval, unit is kWh/s.

In order to evaluate the TEC in the light of the work load of passenger transport, the TEC per 10,000 p-km of a HSR train with a target speed of v between stops is defined by equation II-11. E_{ij}^v

$$e_{ij}^v = \frac{E_{ij}^v}{P_{ij}^v R_{ij}^v D_{ij}^v} \quad \text{......(II-11)}$$

Where :

1. e_{ij}^v is the TEC per 10,000 p-km of the train with the target speed of v from station i to station j, unit is kWh/10,000 p-km.
2. E_{ij}^v is the TEC of the train with the target speed of v from station i to station j, unit is kWh.
3. P_{ij}^v is the number of the passenger seats of the train with the target speed of v from station i to station j.
4. R_{ij}^v is the utilization ratio of the passenger seats of the train with the target speed of v from station i to station j.
5. $D_{ij}^v$is the transport distance covered by the train with the target speed of v from station i to station j, unit is10,000km.

The TEC per 10,000 p-km of a HSR train increases with raising the target speed of the train. It seems that the argument favors the lower target speed but the importance of transport efficiency should be considered. For the purpose of estimating the passenger transport efficiency of a HSR train, the TOT per 10,000 p-km of the passenger transport with a target speed of v between stops is defined by equation II-12:

$$t_{ij}^v = \frac{T_{ij}^v}{P_{ij}^v R_{ij}^v D_{ij}^v} \quad \text{......(II-12)}$$

Where :

1. t_{ij}^v is the TOT per 10,000 p-km of the train with the target speed of v from station i to station j, unit is hours(h)/10,000 p-km.
2. T_{ij}^v is the TOT of the train with the target speed of v from station i to station j, unit is h.

In consideration of both traction energy saving and transport efficiency improvement, the TOC per 10,000 p-km of the passenger transport service by a HSR train with a target speed of v between stops is defined by equation II-13.

$$c_{ij}^{v} = \alpha e_{ij}^{v} + \beta t_{ij}^{v} \quad \text{(II-13)}$$

Where :

1. c_{ij}^{v} is theTOC per 10,000 p-km of the train with the target speed of v from station i to station j, unit is ex : Yuan RMB/10,000 p-km.
2. α , β is the unit prices of the TEC and TOT of the train for the HSR enterprise.

The unit price of the TOT could be explained by equation II-14 in view of the expense and desired benefit of the HSR enterprise.

$$\beta = \frac{F_{od}^{v} P_{od}^{v} \bar{R}_{od}^{v}}{T_{od}^{v}} \quad \text{(II-14)}$$

Where :

1. F_{od}^{v} is the fare of the whole trip from the station of origin o to the station of destination d (including the travel from station i to station j) by riding the train with the target speed of v, unit is ex : Yuan RMB.
2. P_{od}^{v} is the number of the passenger seats of the train with the target speed of v from the station of origin o to the station of destination d.
3. $\bar{R}_{od}^{v}$ is the expected utilization ratio of the passenger seats of the train with the target speed of v from the station of origin o to the station of destination d.
4. T_{od}^{v} is theTOT of the train with the target speed of v from the station of origin o to the station of destination d, unit is h.

The fare of the whole trip from the station of origin o to the station of destination d by riding the HSR train with the target speed of v is explained by equation II-15 to reflect the difference of the fares for different target speeds:

$$F_{od}^{v} = \frac{v}{v^{D}} F_{od}^{v^{D}} \quad \text{(II-15)}$$

Where :

1. v^{D} is the designed top-speed ,i.e. the upper limit of the reachable maximum speed,of the HSR train, unit is km/hand.

2. $F_{od}^{v^D}$ is the fare of the whole trip from the station of origin o to the station of destination d by riding the HSR train with its designed top-speed, unit i.e. Yuan RMB.

Kuo, C.C. & Nichollds, G.M. explained a mathematical modeling approach toimproving locomotive utilization at a freight railroad. In this paper[6] they explained moving freight by rail remains one of the major transportation modes in today's business world. Although railcars compare unfavorably with trucks and airplanes with respect to mobility, flexibility, and speed, the shipping costs are lower and the energy-efficiency is higher. In order to become more competitive in the logistics industry, railroads have taken a number of new initiatives to improve their operations in recent years. One of such efforts made by Consolidated Rail Corporation (Conrail) is described in this paper. The main focus of the present study is on helping Conrail increase the utilization of its locomotive fleet by developing a mixed integer linear program (MILP) to determine the least-cost plan of allocating locomotives to yards and moving light engines between yards. The MILP is tested on a set of real data gathered at Conrail and it is proven to be superior to the existing method. A simple sensitivity analysis is also performed to gain insight into the trade-off between investment in additional locomotives and cost of light engine moves.

Effati, S. and Roohparvar, H. explained the minimization of the fuel costs in the train transportation. In this paper[8] they presenting case studies in modern large scale constrained optimization, the purpose of which is to illustrate how recent advances in algorithms and modelling languages have made it easy to solve difficult optimization problems using of software. They use measure theory technique and iterative dynamic programming algorithm for solving of a trajectory optimization problem: how to drive a train so as to minimize fuel costs. There are 2 methods that they use which is :

1. They optimal control problem to using of atomic measures change this one to an infinite dimensional linear programming problem and then they approximate the latter one to a finite dimensional linear programming problem, then by the optimal solution of the final problem they can obtained sub-optimal controls and then by these controls they obtain the approximate solution of the original problem.
2. They will using dynamic programming as iteration for solving optimal control problem.

An important problem in transportation is to minimize fuel costs in the operation of a train. To keep things simple, assumsption segment of track that is straight although it may contain hills and valleys. Let x denote position along the track measured from some fixed reference point. Letting v denote the derivative of position with respect to time and a the time-derivative of v, we arrive at the equation II-16 and equation II-17 describing the motion of the train.

v(t) = x(t)dt,

a(t) = v(t)dt,

a(t) = h(x(t)) – (e+b|$v(t)$|+cv^2(t)) + $\mu_a(t) - \mu_b(t)$ ....(II-16)

h(x(t)) = $\sum_j^{m-1}(s_{j+1} - s_j)\frac{1}{\pi}\arctan(\frac{x(t)-z_j}{\varepsilon})$...(II-17)

where :

1. m represents the number of hill sections
2. s_j is the slope along the jth section
3. z_j is the breakpoint between the jth and the j + 1st section, and Ɛ gives a spread which is related to the length of the train itself. This involves an initial uphill climb followed by a level section and then a final downhill run.
4. h(x(t)) represents the acceleration/deceleration caused by going down/up hills
5. e, b, and c are constants so that the three terms e + bjv(t)j + cv2(t) represent friction (both from the track and from the surrounding air)
6. μ_a represents the acceleration provided by the engines
7. μ_b represents the deceleration from applying the brakes.

Note : The control variables are the functions μ_a and μ_b.

The objective is to take the train from one place given by initial condition.

x(o) = x_0 $\boxed{\text{To another given by}}$ x(T) = x_f

v(0) = v_0 $\Longrightarrow$ v(T) = v_f

in such a way as to minimize fuel costs, which take to be proportional to the total amount of work done in equation II-18.

$\int_0^T \mu_a(t)v(t)dt$. ....(II-18)

Model adapted from a paper by Kautsky and Nichols (1983) in paper[8]. In face, the following optimal control pronlem :

Minimize : $\int_0^T \mu_a(t)v(t)dt$. (equation II-18)

Subject to : v(t) = x(t)dt

v(t)dt = h(x(t)) – (e+b|$v(t)$|+cv^2(t)) + $\mu_a(t) - \mu_b(t)$ (equation II-16)

x(o) = x_0 ; v(o) = v_0 ; x(T) = x_f ; v(T) = v_f

Jutte, S. & Thoneman, U.W try to explained on the other way to minimalize the cost of train operation by scheduling the crews. In this paper[9] the explained railway operation uses tens of thousands of train movements (trips) and requires thousands of crew members to be assigned to these trips. Despite the large size of the problem, crew schedules need to be generated in short time, because large parts of the train schedule are not finalized until few days before operation. They present a column generation based decomposition algorithm which achieves high-quality solutions at reasonable runtimes. Their divide-and-price algorithm decomposes the problem into overlapping regions which are optimized in parallel. A trip belonging to several regions is initially assigned to one region (‘‘divide’’). The corresponding dual information from optimization is then used as a bonus to offer the trip to other regions (‘‘price’’). Pricing and assignment of trips are dynamically updated in the course of the optimization.

Chung, et al. explained a hybrid genetic algorithm for train sequencing in the Korean railway. In this paper[10] they explained addresses the train-sequencing problem encountered in the Korean railway. It first presents a mixed integer programming model for the problem, in which the mileage must be balanced for each train route, while various field constraints must be satisfied, including overnight stay capacity and maintenance allocation restrictions. Then, it proposes a hybrid genetic algorithm as a solution approach to the problem. The proposed algorithm utilizes a modified elite group technique along with two heuristic procedures based on the mixed integer programming model. Finally, the proposed solution approach is tested with real-world data from the Korean railway.

The train-sequencing problem is to determine a set of train routes covering all trains for a given timetable. A train route identifies the train sequence, operated by rolling stock in the route. Due to the complex nature of the problem, a huge number of solutions may exist for a given timetable. Therefore, a proper performance measure must be identified
to find a practical solution amongst them. The Korean railway considers the balancing of mileage over routes as an important feature in a train-sequencing solution because it can yield useful by-products, such as ease in periodic rolling stock preventive maintenance, and evening out the wear on available rolling stock. There are many restrictions on facility capacity and the operational requirements for real-world train sequencing processes:

1. Each route has lower and upper bounds on totalmileage.
2. All turnarounds must keep a minimum terminal turnaround time and deadheading is not allowed.
3. The number of trains to be covered is limited to a maximum value for each route.
4. Each depot has finite overnight stay and maintenance capacity.

Maintenance capacity is the average daily number of rolling stock sent for periodic preventive maintenance. The maintenance processing time in each depot is defined using the start and end times and duration; therefore, it is the average daily number of routes that must visit a depot during their sequence.

They presents a mathematical model for the train-sequencing problem. The model is based on rolling stock operations in the Korean railway and it includes several logical constraints, as well as facility and operational requirements. They assume that all trains are indexed numerically in descending order of departure time and that dummy trains are included to identify the first and last trains in each route.

Index:

- $i, j \in I \cup Io \cup Id$, train indices $I = \{1, 2, . . . , n\}$. Io and Id are the subsets of dummy trains; $Io =\{o1, o2, . . . , o2\times s \}$, $Id = \{d1, d2, . . . , d2\times s \}$;
- $k \in K$, a route index, $K = \{1, 2, . . . , m\}$
- $p \in P$, a station index, $P = \{1, 2, . . . , s\}$
- $r \in R$, a maintenance-type index, $R =\{1, 2, . . . , q\}$

Input and parameters:

- SS_i , the first departure station (terminal) of train i
- AS_i , the last arrival station (terminal) of train i
- ST_i , departure time of train i at SSi
- AT_i , arrival time of train i at ASi
- IT_r , maintenance time of type r
- BT_r , possible start time of maintenance type r
- ET_r , possible end time of maintenance type r
- D_i , mileage of train i
- MST_{ij} , minimum turnaround time, connecting from train i to train j
- H_{pr} , number of maintenance allocations of type r in station (depot) p
- $C_{pr} = \{<i, j>|AS_i = SS_j = p, i \in I \cup Io, j \in I \cup I_d , \min(ST_j , ET_r) - \max(AT_i , BT_r) \geq IT_r \}$, a set of pairs of consecutive trains that can accommodate maintenance type r at station p
- SL_p, overnight stay capacity of station (depot) p
- MaxSche, an upper bound on total mileage of a route
- MinSche, a lower bound on total mileage of a route
- MaxTr, maximum number of trains that can be assigned to a route.

Decision variables:

- $x_{ijk} = \begin{cases} 1 \text{ if trains i and j are assigned to route k in sequence} \\ 0 \text{ otherwise}, \forall i \in I \cup Io, \forall j \in I \cup Id, \forall k \in K \end{cases}$
- $y_{kpr} = \begin{cases} 1 \text{ if maintenance type r is allocated to route k in station p} \\ 0 \text{ otherwise}, \forall k \in K, \forall p \in P, \forall r \in R \end{cases}$
- TDk, total mileage of route k
- $\overline{TD}$, the average mileage for all routes.

Train-sequencing model:

Min $\sum_{k\in K}|\overline{TD} - TDk|/m$ ......(II-19)

The objective function is formulated to minimize the mean absolute deviation (MAD) of the mileage for all routes to balance mileages. MAD can be substituted by the standard deviation because they are proportional to each other within a constant multiple under a mild assumption.

s.t.

1. $\sum_{k\in K}\sum_{i\in I\cup Io} x_{ijk} = 1, \quad \forall j \in I$......(II-20)
2. $\sum_{k\in K}\sum_{i\in I\cup Id} x_{ijk} = 1, \quad \forall j \in I$......(II-21)
3. $\sum_{j\in I\cup Io} x_{ijk}$ - $\sum_{j\in I\cup Id} x_{ijk}$ = 0, ∀k ∈ K, ∀i ∈ I......(II-22)
4. $\sum_{i\in I}\sum_{j\in I} x_{ijk} \geq 1, \quad \forall k \in K$......(II-23)
5. $\sum_{i\in Io}\sum_{j\in I} x_{ijk} = 1, \quad \forall k \in K$......(II-24)
6. $\sum_{i\in I}\sum_{j\in Id} x_{ijk} = 1, \quad \forall k \in K$......(II-25)
7. $\sum_{i\in I\cup Io}\sum_{j\in I\cup Id} D_i x_{ijk} - TD_k = 0, \quad \forall k \in K$......(II-26)
8. $TD_k \geq$ MinSche, ∀k ∈ K ......(II-27)
9. $TD_k \leq$ MaxSche, ∀k ∈ K ......(II-28)
10. $ST_j - AT_i + (1 - x_{ijk}) M \geq MST_{ij}$; ∀i, j ∈ I ; ∀k ∈ K ......(II-29)
11. $SS_j - AS_i + (1 - x_{ijk})M \geq 0$; ∀i, j ∈ I ; ∀k ∈ K ......(II-30)
12. $AS_i - SS_j + (1 - x_{ijk})M \geq 0$; ∀i, j ∈ I ; ∀k ∈ K......(II-31)
13. $\sum_{i\in I}\sum_{j\in I\cup Id} x_{ijk} \leq MaxT_r, \quad \forall k \in K$......(II-32)
14. $\sum_{k\in K}\sum_{i\in I:AS_i=p}\sum_{j\in Id:SS_j=p} X_{ijk} = SL_p \quad \forall j \in I$......(II-33)
15. $\sum_{k\in K} y_{kpr} = H_{pr}$; ∀p ∈ P, ∀r ∈ R......(II-34)
16. $\sum_{p\in P}\sum_{r\in R} y_{kpr} \leq 1, \quad \forall k \in K$......(II-36)
17. $\sum_{<i,j>\in C_{pr}} x_{ijk} - y_{kpr} \geq 0$,∀k ∈ K, ∀p ∈ P, ∀r ∈ R......(II-37)

Where :

- Constraints on equation II-20 and II-21 guarantee the unique assignment of a train to a route.
- Constraints on equation II-22 represents flow conservationfor each route.
- Constraint II-23 prohibits emptyroutes, requiring at least one train to be included in each route.
- Constraints II-24 and II-25 indicate that the first and last trains on each route are dummy trains.
- Constraint II-26 computes the total mileage for each route.

- Constraints II-27 and II-28 enforce the lower and upper bounds of mileage for each route, respectively.
- Constraint II-29 guarantees the minimum turnaround time between consecutive trains in a route, where *M* is a large constant value.
- Constraints II-30 and II-31 prohibit deadhead operations in route turn around.
- Constraint II-32 limits the number of trains that can be allocated in each route.
- Constraint II-33 represents the overnight stay capacity for each depot.
- Constraints II-34 and II-35 guarantee that the required number of routes are directed to maintenance stations for the corresponding type of operations.
- Constraint II-36 restricts the number of maintenance allocations for each route.
- Constraint II-37 ensures the assigned type of maintenance operations in a route.

Min,Y.H et al explained an appraisal of a column-generation-based algorithm for centralized train-conflict resolution on a metropolitan railway network. They explained the train-conflict resolution is decentralized around dispatchers each of whom controls a few segments in a global railway network with her rule-of-thumb to operational data. Conceptually, the global sub-optimality or infeasibility of the decentralized system is resolved by a network controller who coordinates the dispatchers and train operators at the lower layers on a real-time basis. However, such notion of a multi-layer system cannot be effectual unless the top layer is able to provide a global solution soon enough for the dynamic lower layers to adapt in a seamless manner. Unfortunately, a train-conflict resolution problem is NP-hard as formally established in this paper[11] and an effective solution method traded off between computation time and solution quality has been lacking in literature. They proposed a column-generation-based algorithm that exploits the separability of the problem. A key ingredient of the algorithm is an efficient heuristic for the pricing subproblem for column generation.

Firstly they gave some terminologies about railway networks, train timetables, train-conflicts, and train-conflict resolution problem.

a) Railway Networks

   An undirected graph consisting of the nodes of stations and the edges of segments. A segment is the set of tracks connecting two adjacent stations. This definition reflects that the model proposed in the paper does not consider the routing of trains within a station.

b) Train Timetables

   The set of quadruples (station, arrival time, depature time, track assignment) for all trains and the stations which they visit over a predetermined time horizon.

c) Train-conflicts

A situation in which two or more trains claim resources in an infeasible manner, namely, in the way that violates a safety regulation. Conflicts arise when a set of trains fail to respect the safe headway times due to delays.

d) Train-conflict Resolution Problem

Input : A railway network, a timetable, and a set of train-conflicts.

Output: A conflict-free timetable with the minimum total weighted deviation w.r.t. the original timetable which satisfies the constraints on the earliest departure times, the headway times, the transit times, and the dwelling times.

Secondly is the make the assumptions which are :

a) Each station has enough number of platforms.

This is motivated by the empirical data from KORAIL (Korea Railroad) which shows that the stations of Seoul metropolitan area have enough number of platforms for train-conflict resolution. Indeed, in the experiment they observe that the number of platforms required by our algorithm does not exceed the number of platforms in any station. If this is not the case, our model does not guarantee a feasible platform assignment.

b) A railway network is a tree.

c) Each track is unidirectional.

Assumptions b and c reflects the Seoul metropolitan railway network. Based on these assumptions, they duplicate each segment between two adjacent stations. Then, two adjacent stations have two parallel segments to which they assign opposite directions. And then view a segment as a set of tracks with the same direction. See Figure II-1. Now choose a station in the railway network as the root of a tree.

A segment is called inbound if it is oriented towards the root, and outbound otherwise.

Then every train is one of the following types:

i. An inbound train which only traverses inbound segments.

ii. An outbound train which only traverses outbound segments.

An in-outbound train which traverses inbound segments first and outbound segments afterwards. An in-outbound train can be thought of as two trains. One of them traverses inbound segments from the origin to a certain station from which the other traverses outbound segments to the destination. For example, in Figure II-1, an in-outbound train from station 2 to station 4 is the concatenation of an inbound train from station 2 to the root and an outbound train from the root to station 4. Then, we can obtain a feasible timetable of all trains by first computing the timetable of inbound trains and then that of outbound trains. This method does not guarantee an optimal solution. But, at the same time, it does not violate the optimality by much since there are few in-outbound trains in practice. For instance, there are no in-outbound trains in the Seoul metropolitan area. Thus the method guarantees the optimality. In addition, it is pragmatic since the problem size can be halved, corresponding to our purpose. Therefore, we will solve the conflict resolution problem by devoting a segment to two opposite directions.

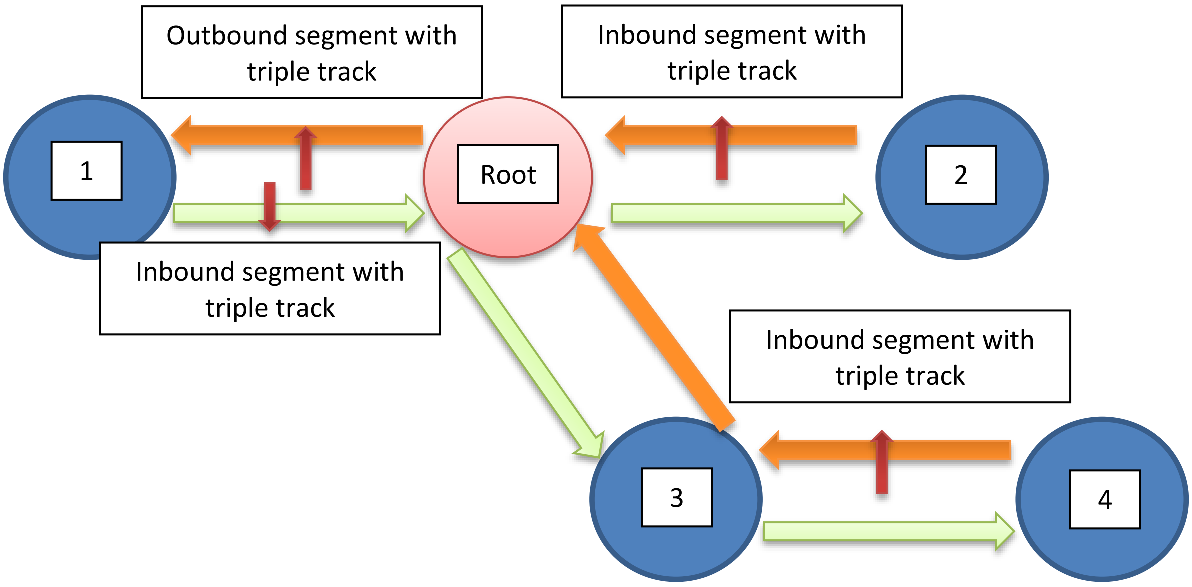

Figure II.1 A railway network

The train-conflict resolution problem can be formulated as an MIP. The second column indicates the equation numbers of the corresponding features of the MIP-model II-38 – II-37 presented below. Note that intra-track headway times, dispatch sequencing, and dwelling time requirements at stations are the most common features of the MIP-models for the train-conflict resolution problem or timetabling problem. Regarding the dwelling modes of trains, notice that even when a station is originally not a stopping station for a train, the total deviation of a new timetable can be smaller when the train is allowed to dwell at the station. If so, the train will have a longer transit time on the incident segments due to deceleration and acceleration. Such difference may be significant for a metropolitan network where segments are relatively short and trains are more frequent. Zhou and Zhong (2005) in this paper[11], to the best of the author's knowledge, were first to explicitly incorporate the dwelling modes of trains as decision variables into their timetabling model for a high speed train line where the difference is also significant. However, it would be too excessive for our goal of achieving minimal computation to design an algorithm searching over all possible combinations of the dwelling modes of trains over segments. Thus, in our algorithm, we initially fix the dwelling modes of trains and search a better combination in the postprocessing. The feature of our model is the intra-track safe headway time: on the same segment, the departure or arrival times of any pair of consecutive trains should not be too close even if they use different tracks of a segment. This is to enforce the safe headway time at the points where tracks cross.

Thirdly is make an MIP-formulation, they denote Q, P, and I to denote the set of stations, segments, and trains, respectively. P and I here are either all inbound or all outbound. Also, Ip I denotes the set of trains traversing segment p. The parameters and decision variables of the proposed MIP-formulation are as follows:

1. Parameters
   - c_p^i = cost of the unit departure delay of train i at segment p.
   - w_p^i = cost of the unit deviation (delay or early arrival) from original arrival time of train i at segment p.
   - d_p^i = original departure time of train i at segment p.
   - a_p^i = original arrival time of train i at segment p.
   - h_q = safe headway time between two consecutive trains arriving or departing at station q via the same track.
   - h_m = safe headway time between two consecutive trains arriving or departing at a station via different tracks.
   - s_q^i = minimum dwelling time of train i at station q (defined as 0 if q is not a stopping station for train i).
   - $\underline{t}_p^i$(oo) = minimum transit time over segment p of train i when dwelling mode of i at the two end stations are oo; oo= ss, sp, ps, or pp meaning, in their order, stop–stop, stop–pass, pass–stop, or pass–pass.
   - $\bar{t}_p^i$(oo) = maximum transit time over segment p of train i with dwelling mode oo.
   - b_p^i = earliest possible (nonnegative) departure time of train i at segment p.
   - n_p = number of tracks of segment p.
   - M = a sufficiently large positive number.
2. Decision variables
   - D_p^i = departure time of train i at end station q of segment p = (q, r).
   - A_p^i = arrival time of train i at end station r of segment p = (q, r).
   - E_p^i = difference between original arrival time, a_p^i and actual arrival time, A_p^i.
   - $Z_{p,k}^i$ = 1 if train i is assigned to the kth track of segment p and 0 otherwise.

- X_p^{ij} = 1 if train i departs earlier than train j at end station q of segment p = (q, r) and 0 otherwise.
- Y_p^{ij} = 1 if train i arrives earlier than train j at end station r of segment p = (q, r) and 0 otherwise.
- oo_p^i = 1 if train has a dwelling mode oo at segment p and 0 otherwise, where oo = ss, sp, ps, or pp meaning in their order, stop–stop, stop–pass, pass–stop, or pass–pass.

After all, they are conducting formula for the objective function and constraints. The objective is to minimize the total weighted deviation (namely, both earliness and tardiness) between original and rescheduled timetables of trains in equation II-38.

$$min \sum_{p\in \mathrm{P}} \sum_{i\in I_p} \{c_p^i(D_p^i - d_p^i) + w_p^i E_p^i\} \text{ ....(II-38)}$$

There are several constraints that can be see from equation II-39 to II-47. Deviation from the original timetable. These constraints are two linear constraints that are equivalent to

$$E_p^i \geq |A_p^i - a_p^i| : \forall p \in \mathrm{P},\ \forall i \in I_p,\ A_p^i - a_p^i \leq E_p^i \ and\ -(A_p^i - a_p^i) \leq E_p^i \text{ .... (II-39)}$$

Track assignment. If a segment p has multiple tracks, a train i traversing segment p has to choose exactly one track: ∀p ∈ P, ∀i ∈ I_p,

$$\sum_{k=1}^{n_p} Z_{p,k}^i = 1 \text{ ....(II-40)}$$

Train-dispatch sequencing. A departing sequence of trains should be equal to their arrival sequence over the same track: ∀p ∈ P, ∀i,j ∈ I_p with i≠j, k =1,....., n_p

$$2 - Z_{p,k}^i - Z_{p,k}^j + (X_p^{ij} - Y_p^{ij}) \geq 0 \text{ and } 2 - Z_{p,k}^i - Z_{p,k}^j - (X_p^{ij} - Y_p^{ij}) \geq 0 \text{ ....(II-41)}$$

Earliest departure times of trains. Each train i should observe its earliest possible departure time: ∀p ∈ P, ∀i ∈ I_p,

$$D_p^i \geq b_p^i \text{ ....(II-42)}$$

Intra-track headway safety. Two trains i and j traversing segment p via track k have to observe the minimum headway time: ∀p ∈ P, ∀i,j ∈ I_p with i≠j, k =1,....., n_p,

$$M(1 - Z_{p,k}^i) + M(1 - Z_{p,k}^j) + \begin{cases} M(1 - X_p^{ij}) + D_p^j - D_p^i \geq h_q \\ M(1 - X_p^{ij}) + A_p^j - A_p^i \geq h_r \\ MX_p^{ij} + D_p^i - D_p^j \geq h_q, and \\ MX_p^{ij} + A_p^i - A_p^j \geq h_r \end{cases} \text{ ....(II-43)}$$

Inter-track headway safety. Two trains i and j traversing segment p have to observe the minimum headway time: ∀i,j ∈ I_p with i≠j,

$$M(1 - X_p^{ij}) + D_p^j - D_p^i \geq h_m$$
$$M(1 - Y_p^{ij}) + A_p^j - A_p^i \geq h_m$$
$$MX_p^{ij} + D_p^i - D_p^j \geq h_m, and \quad \text{.......................................(II-44)}$$
$$MY_p^{ij} + A_p^i - A_p^j \geq h_r$$

Dwelling modes and corresponding transit times at segments. For each segment p = (q, r), each train chooses a dwelling mode. (However, if station q is originally a dwelling station, then a train has no other choice than to stop at q.) Accordingly, the transit time of each train is determined between its lower- and upper-bounds: ∀p = $(q, r) \in \mathrm{P}$ and ∀i ∈ I_p.

$SS_p^i + SP_p^i + PS_p^i + PP_p^i = 1$ ....(II-45)

($SS_p^i + SP_p^i = 1$, added if q is originally a dwelling station of i,)

$\underline{t}_p^i$(ss) SS_p^i + $\underline{t}_p^i$(sp) SP_p^i+ $\underline{t}_p^i$(ps) PS_p^i+$\underline{t}_p^i$(pp) PP_p^i ≤ $A_p^i - D_p^i$ and

$\bar{t}_p^i$(ss) SS_p^i + $\bar{t}_p^i$(sp) SP_p^i+ $\bar{t}_p^i$((ps) PS_p^i+$\bar{t}_p^i$((pp) PP_p^i ≥ $A_p^i - D_p^i$.......................................(II-46)

Dwelling time requirements. Each train has to dwell for a positive predetermined time at station q if it stops at station q. Otherwise the dwelling time should be zero: ∀q ∈ Q, for each pair of consecutive segments incident to q, p' = (s,q) and p = (q, r), and for each train traversing p' and p, namely ∀i ∈ Ip'∩Ip, we have

$s_q^i(SS_p^i + SP_p^i) \leq D_p^i - A_{p'}^i$ ≤ M($SS_p^i + SP_p^i$) and

$s_q^i(SS_{p'}^i + SP_{p'}^i) \leq D_p^i - A_{p'}^i \leq$ M($SS_{p'}^i + PS_{p'}^i$) ....(II-47)

Notice that each pair of equation II-47 $(SS_{p'}^i + PS_{p'}^i) = SS_p^i + SP_p^i$ also implies so that the dwelling modes over p' and p are consistent.

Notice that the constraints from equation II-47are the only coupling constraints that control the movement of a train over adjacent segments sharing the same station. Hence, if equation II-47 is relaxed, the MIP-model is separable into the independent subproblems with respect to the segments p ∈ P. More specifically, each independent subproblem is a train-conflict resolution problem on a single segment. Conversely, a feasible solution of the original problem, namely a global timetable, is the combination of the timetables from the subproblems that satisfy the coupling constraints of equation II-47.

Xu, et. Al explained Research and Simulation on the Regenerativve Braking Process of DC Railway Traction System. In this paper[12] they conducts research and simulation on the regenerative braking process of the DC railway traction system. The simulation model of the regenerative braking system's main circuit configuration and the slip frequency vector control system. The regenerative braking current's waveform confirms them model's accuracy. The simulation results will provide the important basis for the improvement of DC feeder line protection algorithm. By comparing the regenerative braking current waveform and the remote short-circuit current, the effective of DDDL protection will be elaborated.

Yang, L.X., et. al explained railway freight transportation planning with mixed uncertainty of randomness and fuzziness. In this paper[13] they investigated the railway freight transportation planning problem under the mixed uncertain environment of fuzziness and randomness in which the optimal paths, the amount of commodities passing through each path and the frequency of services need to be determined. Based on the chance measure and critical values of the random fuzzy variable, three chance-constrained programming models are constructed for the problem with respect to different criteria. Some equivalents of objectives and constraints are also discussed in order to investigate mathematical properties of the models. To solve the models, a potential path searching algorithm, simulation algorithms and a genetic algorithm are integrated as a hybrid algorithm to solve an optimal solution.

Yun, Bai, et.al explained Energy-Efficient Driving Strategy for Freight Trains Based on Power Consumption Analysis. In this paper[14] they examined the main forms of energy consumption, which include work of resistance and kinetic energy loss caused by braking. From the view of theoretical analysis and simulations, the paper discusses the energy consumption with different coasting distance before braking and lower speed restriction, and train speed uniformity. The former makes a great impact on the kinetic energy loss and the latter determines the work of resistance.

According to the boundary conditions, the train speed is zero at the start and end points. The running distance is unalterable value S. Therefore, if a freight train is considered as a mass belt, its motion equation may be expressed as follows:

$M.v\frac{dv}{dx} = \frac{p(x)}{v(x)}$ - R(v) – B + $\int_0^S \theta(s).g(x-s)ds$ ....(II-48)

Where :

1. M—total mass of train, M>0
2. p(x)—traction power of train, p(x) ≥ 0, x∈ (0, S)
3. v(x) —running speed of train, v(x) ≤ Res, x∈ (0, S)
4. R(v) —resistance of the train, R(v) = av^2 +bv+c ≥ 0
5. B—braking force of the train, B = b(r,v,v) ≥ 0
6. ϴ(s)—mass density function of the train at a point which is s meters away from the head.

7. $g(x–s)$—additional track resistance (including slopes and curves) applied on a point which is s meters away from the head.

The energy consumption depends on the work made by traction and its average energy efficiency. Under the condition of given train formation, the energy efficiency is unalterable; hence, how to reduce the work of traction is the focus of the energy saving operation of freight trains. According to equation II-48 the work of traction mainly used to overcome the work of resistance on the train, increase the kinetic energy of the train to compensate for the kinetic energy loss resulting from braking, and supply for the gravitational potential difference. The train movement transfer formulation between energy and work can be described as follows:

$$E = F.s/\eta = (\int_0^s r(v)ds + \sum_i \frac{1}{2} m \left(v_{i_t}^2 - v_{i_{t'}}^2 \right) + mg\Delta h + \mathcal{E}(T))/\ \eta$$

Where :

v_{i_t}—speed at the end of braking;

$v_{i_{t'}}$—speed at the end of braking section as calculated under the control of coasting

Δh —the difference between the elevations at the start andend points

η —gross transmission efficiency

$\mathcal{E}$ (T)—energy consumed by the self-running of locomotive.

With a given schedule time, the energy consumed by the self-running of locomotive is irrelevant to the driving method, and, according to statistics, only accounts for a fraction of the total energy consumption of the train. The gravitational potential difference of the train will be constant across the start and end points of the running route because the elevation difference is always the same between both points. Therefore, the key to minimize the energy consumption of the train is to minimize the work of the resistance and the loss of the kinetic energy resulting from braking, by optimizing the locomotive driving sequence, under the given hardware availability and running condition.

The analysis of key factors on energy-efficient operation. According to the statistics of simulations, 80–90% of the energy consumption is used to overcome the work of resistance on the train, and 10–20% of the energy consumption is used to increase the kinetic energy of the train to compensate for the kinetic energy loss resulting from braking. Hence, the research on energy-efficient operation of the train can be studied from two aspects. One is the relationship between speed uniformity and work of resistance because resistant formula indicates that the uniformity of velocity makes significant impact on the work of resistance, and the other is the loss of kinetic energy caused by braking.

Keeping a train at a uniform speed has two meanings. First, the train should coast as long as possible wherever down a long and steep slope. This is to avoid excessively high speed as the train is simultaneously accelerated by traction and gravity. Especially, it can avoid braking for deceleration when the speed is close to the restriction. Second, a high traction, if possible, should be applied to the train, which is at a certain distance ahead of a long and steep upslope, without braking later. This is to accelerate the train to a proper speed at the start of steep upslope, and guarantee the speed of train no slower than the lower target speed when the train is on the slope. It will not only minimize the duration of the high traction power, but also avoid idle running, which is the case that the traction force is higher than the adhesive traction of the train.

Train braking may be classified into deceleration braking and stop braking by its purpose. Generally, deceleration braking is typically unnecessary since it can be avoided by observing the track profile and speed restrictions ahead and selecting an appropriate coasting point at the general situation. Stop braking is referred to as braking and stopping the train as scheduled, or as arranged by the dispatcher, to satisfy, for example, the needs of marshalling yards or give way to oncoming train in the opposite direction on a single track. To avoid unnecessary braking, it should be noted that the gravitational potential of the train should be maximally utilized when the train is running down a relatively long slope (the velocity will be increased due to the positive additional resistance of the slope), by coasting or lowering the traction handle on a proper stage so that the speed is controlled in an appropriate range. In addition, where there is a lowered speed restriction ahead, the switchover point from traction to coasting should be calculated beforehand, thus the train can coast to the speed restricted section without braking.

Although the braking for stop is generally necessary, it may be avoided in special conditions if a freight train is to meet another train on a semi-automatic block in single-track section. For example, if the crew realizes another train will be met, and the oncoming train in the opposite direction will soon complete its station pull-in, the crew may intentionally make more use of lower traction or coasting control to extend the running time in the section and avoid the stop braking while pulling in the station if the oncoming train completed its station pull-in. In addition, the train may have a higher coasting ratio to minimize its speed just before it has to be braked, provided that the constraint of schedule time is satisfied. This helps reducing the kinetic energy loss of train.

To optimally drive a freight train by considering the effect of all the above controllable factors, one should first calculate the highest train-handle stage which can avoid deceleration braking and satisfy the constraint of handle enduring time. Next, when the train locates a section favorable for converting gravitational potential to kinetic energy, which is judged by the track grades and the difference between current velocity and restriction velocity, the train should lower handle to coast to minimize the work of traction while keeping the speed at a proper value. If the train runs in an undulate track section, the train may keep the speed to fluctuate as less as possible to reduce its work of resistance. If the train is approaching a long and steep upslope, its handle should be raised as early as possible. When the train reaches the foot of the slope, the train may be accelerate to the maximize speed under the restriction. When the train successively climbs up the slope, this traction power should also keep the speed over the lower target limit wherever possible. Thus the train can ride over the slope with its initial kinetic energy. Furthermore, the crew may increase the coasting rates before the stop braking if the schedule time can be satisfied, which is helpful to minimize the energy consumption.

According to the train motion transfer formulation between energy and work, the energy consumption of a train is primarily used to overcome its work of resistance, increase its kinetic energy to compensate for the kinetic energy loss resulting from braking, overcome gravitational potential difference and supply for the self-running of locomotives. Due to the fact that the gravitational potential differences and the energy consumed by the self-running of locomotives are unchanged in a given running condition, the key to minimize the energy consumption of train is to minimize the work of resistance and the loss of kinetic energy resulting from braking, by optimizing the locomotive driving sequence. As indicated by the simulation experiment, the energy consumption of a train will be saved up to 9% by avoiding unnecessary deceleration braking and increasing the coasting distance properly before braking, while the schedule time is only slightly affected. And via reasonably unifying the speed to minimize the work of resistance, it is possible to reduce the energy consumption by 6.8% without increasing the running time. By combing both energy-efficient driving strategies, it is possible to improve energy saving more significantly, while the running time is roughly kept unchanged. The results so far suggest that it is feasible and remarkable to reduce energy consumption by improving train operation.

Zhao, Y., et.al explained performance measurement of a transportation network with a downtown space reservation system. In this paper[15] they propose an evaluation approach for a novel travel demand management strategy known as the downtown space reservation system (DSRS). This approach takes into account three perspectives, i.e., transportation service provider's, the user's, and the community's and is based on network-Data Envelopment Analysis (DEA) where the perspectives are inter-related through intermediate inputs/outputs. Two types of network-DEA models (radial and slacks-based models) are considered. An example is provided using data propagated from a microscopic traffic simulation model of the DSRS. The results show that individual node performance can drive network DEA performance and that this information can inform future designs of the DSRS.

Blanco, V., et.al explained *the expanding the Spanish high-speed railway network.* In this paper[16] they presents a model for the expansion of transportation networks incorporating specific requirements about population coverage, budget constraints, intermediate goals and origin–destination flows, among others. The model is applicable to the current expansion project of the Spanish high-speed railway network that has been proposed by the Spanish Government under the program Strategic Planning of Infrastructure and Transport. Their approach looks for solutions that may be used as additional information in the decision-making process of any network expansion.

He, S., et. al explained fuzzy dispatching model and genetic algorithms for railyards operations. In this paper[17] they present a fuzzy dispatching model to assist the coordination among multi-objective decisions in railyards dispatching plan. The objectives considered here are to maximize railyards output and on-time service of the dispatching plan. Genetic algorithms are used to investigate and analyze the fuzzy dispatching model. Experimental results based on real-world problem instances show that the combined, genetic algorithms and fuzzy theory approaches could be a promising way for solving railyards dispatching problem.

Masoud, Y. ,et. al explained solving railroad blocking problem using ant colony optimization algorithm. The aim of this paper[18] to minimize the costs of delivering all commodities by deciding which inter-terminal blocks to build and specifying the assignment of commodities to these blocks, while observing limits on the number and aggregate volume of the blocks assembled at each terminal. This paper presents a metaheuristic algorithm based on ant colony optimization for solving this problem. To evaluate the efficiency of the proposed algorithm and the quality of solutions, experimental analysis is conducted, using several simulated test problems. The results on the test problems are compared with those of solution generated with CPLEX software. The results show high efficiency and effectiveness of the proposed algorithms.

Whitley, Darrell explained a genetic algorithm tutorial. In this paper[19] he examined the tutorial for the canonical genetic algorithm as well as more experimental forms of genetic algorithms, including parallel island models and parallel cellular genetic algorithms. The tutorial also illustrates genetic search by hyperplane sampling. The theoretical foundations of genetic algorithms are reviewed, include the schema theorem as well as recently developed exact models of the canonical genetic algorithm.

Genetic Algorithms are a family of computational models inspired by evolution. These algorithms encode a potential solution to a specifc problem on a simple chromosome-like data structure and apply recombination operators to these structures so as to preserve critical information. Genetic algorithms are often viewed as function optimizers, although the range of problems to which genetic algorithms have been applied is quite broad. An implementation of a genetic algorithm begins with a population of (typically random) chromosomes. One then evaluates these structures and allocates reproductive opportunities in such a way that those chromosomes which represent a better solution to the target problem are given more chances to "reproduce" than those chromosomes which are poorer solutions. The "goodness" of a solution is typically defined with respect to the current population. In a broader usage of the term, a genetic algorithm is any population-based model that uses selection and recombination operators to generate new sample points in a search space. Many genetic algorithm models have been introduced by researchers largely working from an experimental perspective. Many of these researchers are application oriented and are typically interested in genetic algorithms as optimization tools. The goal of this tutorial is to present genetic algorithms in such a way that students new to this field can grasp the basic concepts behind genetic algorithms as they work through the tutorial. It should allow the more sophisticated reader to absorb this material with relative ease. The tutorial also covers topics, such as inversion, which have sometimes been misunderstood and misused by researchers new to the field.

Firstly he explained the encodings and optimization problems. Usually there are only two main components of most genetic algorithms that are problem dependent: the problem encoding and the evaluation function. Consider a parameter optimization problem where we must optimize a set of variables either to maximize some target such as profit, or to minimize cost or some measure of error. We might view such a problem as a black box with a series of control dials representing different parameters; the only output of the black box is a value returned by an evaluation function indicating how well a particular combination of parameter settings solves the optimization problem. The goal is to set the various parameters so as to optimize some output. In more traditional terms, we wish to minimize (or maximize) some function F$(X_1, X_2, \dots\dots, X_M)$. Most users of genetic algorithms typically are concerned with problems that are nonlinear. This also often implies that it is not possible to treat each parameter as an independent variable which can be solved in isolation from the other variables. There are interactions such that the combined efects of the parameters must be considered in order to maximize or minimize the output of the black box. In the genetic algorithm community, the interaction between variables is sometimes referred to as epistasis.

The first assumption that is typically made is that the variables representing parameters can be represented by bit strings. This means that the variables are discretized in an a priori fashion, and that the range of the discretization corresponds to some power of 2. For example, with 10 bits per parameter, we obtain a range with 1024 discrete values. If the parameters are actually continuous then this discretization is not a particular problem. This assumes, of course, that the discretization provides enough resolution to make it possible to adjust the output with the desired level of precision. It also assumes that the discretization is in some sense representative of the underlying function. If some parameter can only take on an exact finite set of values then the coding issue becomes more difficult. For example, what if there are exactly 1200 discrete values which can be assigned to some variable X_i. We need at least 11 bits to cover this range, but this codes for a total of 2048 discrete values. The 848 unnecessary bit patterns may result in no evaluation, a default worst possible evaluation, or some parameter settings may be represented twice so that all binary strings result in a legal set of parameter values. Solving such coding problems is usually considered to be part of the design of the evaluation function.

Aside from the coding issue, the evaluation function is usually given as part of the problem description. On the other hand, developing an evaluation function can sometimes involve developing a simulation. In other cases, the evaluation may be performance based and may represent only an approximate or partial evaluation. For example, consider a control application where the system can be in any one of an exponentially large number of possible states. Assume a genetic algorithm is used to optimize some form of control strategy. In such cases, the state space must be sampled in a limited fashion and the resulting evaluation of control strategies is approximate and noisy (c.f., Fitzpatrick and Grefenstette,1988) in paper[19]. The evaluation function must also be relatively fast. This is typically true for any optimization method, but it may particularly pose an issue for genetic algorithms. Since a genetic algorithm works with a population of potential solutions_ it incurs the cost of evaluating this population. Furthermore, the population is replaced (all or in part) on a generational basis. The members of the population reproduce, and their offspring must then be evaluated. If it takes 1hour to do an evaluation, then it takes over 1year to do 10000 evaluations. This would be approximately 50 generations for a population of only 200 strings.

Assuming the interaction between parameters is nonlinear, the size of the search space is related to the number of bits used in the problem encoding. For a bit string encoding of length L, the size of the search space is 2^L and forms a hypercube. The genetic algorithm samples the corners of this L-dimensional hypercube. Generally, most test functions are at least 30 bits in length and most researchers would probably agree that larger test functions are needed. Anything much smaller represents a space which can be enumerated. Ofcourse, the expression 2^L grows exponentially with respect to L.

The point is that as long as the number of “good solutions” to a problem are sparse with respect to the size of the search space, then random search or search by enumeration of a large search space is not a practical form of problem solving. On the other hand, any search other than random search imposes some bias in terms of how it looks for better solutions and where it looks in the search space. Genetic algorithms indeed introduce a particular bias in terms of what new points in the space will be sampled. Nevertheless, a genetic algorithm belongs to the class of methods known as “weak methods” in the Artifcial Intelligence community because it makes relatively few assumptions about the problem that is being solved. Of course, there are many optimization methods that have been developed in mathematics and operations research. What role do genetic algorithms play as an optimization tool. Genetic algorithms are often described as a global search method that does not use gradient information. Thus, nondifferentiable functions as well as functions with multiple local optima represent classes of problems to which genetic algorithms might be applied. Genetic algorithms, as a weak method, are robust but very general. If there exists a good specialized optimization method for a specific problem, then genetic algorithm may not be the best optimization tool for that application.

Raj, K.A.A.D and Rajendran, C., explained . A genetic algorithm for solving the fixed-charge transportation model: Two-stage problem. In this paper[20] they propose genetic algorithms to solve a two-stage transportation problem with two different scenarios. The first scenatio considers the per-unit transportation cost and the fixed cost associated with a route, coupled with unlimited capacity at every DC. The second scenario considers the opening cost of a distribution center, per-unit transportation cost from a given plant to a given DC and the per-unit transportation cost from the DC to a customer. Subsequently, an attempt is made to represent the two-stage fixed-charge transportation problem (scenario 1) as a single-stage fixed-charge transportation problem and solve the resulting problem using the genetic algorithm. Many benchmark problem instances are solved using the proposed genetic algorithms and performances of these algorithms are compared with the respective best existing algorithms for two-scenarios. The results from computational experiments show that the proposed algorithms yield better solutions than the respective best existing algorithms for the two scenarios under consideration.

In their work on continous management of airlift and tanker resources (Smith, Becker and Kramer, 2004)[23]. They explained efficient allocation of aircraft and aircrews to transportation missions is an important priority at the USAF Air Mobility Command (AMC), where airlift demand must increasingly be met with less capacity and at lower cost. In addition to presenting a formidable optimization problem, the AMC resource management problem is complicated by the fact that it is situated in a continuously executing environment. Mission requests are received (and must be acted upon) incrementally, and, once allocation decisions have been communicated to the executing agents, subsequent opportunities for optimizing resource usage must be balanced against the cost of solution change. In this paper, they describe the technical approach taken to this problem in the AMC barrel allocator, a scheduling tool developed to address this problem and provide support for day-to-day allocation and management of AMC resources. The system utilizes incremental and configurable constraint-based search procedures to provide a range of automated and semi-automated scheduling capabilities. Most basically, the system provides an efficient solution to the fleet scheduling problem. More importantly to continuous operations, it also provides techniques for selectively reoptimizing to accommodate higher priority missions while minimizing disruption to most previously scheduled missions, and for selectively “merging” previously planned missions to minimize nonproductive flying time. In situations where all mission requirements cannot be met, the system can generate and compare alternative constraint relaxation options. The barrel allocator technology is currently transitioning into operational use within AMC's Tanker/Airlift Control Center (TACC).

In their work on solving the theater distribution vehicle routing and scheduling problem using group theoritic tabu search (Crino, Moore, Barnes and Nanry, 2004)[23]. They explained the military “theater distribution vehicle routing and scheduling problem” (TDVRSP) is associated with determining superior allocations of required flows of personnel and materiel within a defined geographic area of operation. A theater distribution system is comprised of facilities, installations, methods, and procedures designed to receive, store, maintain, distribute, and control the flow of material between exogenous inflows to that system and distribution to end-user activities and units within the theater. An automated system that can integrate multimodal transportation assets to improve logistics support at all levels has been characterized as a major priority and immediate need for the U.S. military services. This paper describes both the conceptual context, based in a flexible group theoretic tabu search (GTTS) framework, and the software implementation of a robust, efficient, and effective prescriptive generalized theater distribution methodology. This methodology evaluates and prescribes the routing and scheduling of multimodal theater transportation assets at the *individual asset operational level* to provide economically efficient time-definite delivery of cargo to customers.

In their work on Solving the aerial fleet refueling problem using group theoretic tabu search (Barnes, Wiley, Moore and Ryer, 2004)[23]. They explained the aerial fleet refueling problem (AFRP) is concerned with the efficient and effective use of a heterogeneous set of tanker (refueling) aircraft, located at diverse geographical locations, in the required operations associated with the deployment of a diverse fleet of military aircraft to a foreign theater of activity. Typically, the “receiving” aircraft must traverse great distances over large bodies of water and/or over other inhospitable environs where no ground based refueling resources exist. Lacking the ability to complete their flights without refueling, each receiving aircraft must be serviced one or more times during their deployment flights by means of in-flight refueling provided by one of the available tanker aircraft. The receiving aircraft, aggregated into receiver groups (RGs) that fly together, have stipulated departure and destination bases and each RG's arrival time is bounded by a stated desired earliest and latest time. The excellence of a suggested solution to this very challenging decision making problem is measured relative to a rigorously defined hierarchical multicriteria objective function. This paper describes how the AFRP for the Air Mobility Command (AMC), Scott Air Force Base, IL, is efficiently solved using group theoretic tabu search (GTTS). GTTS uses the symmetric group on n letters (S_n) and applies it to this problem using the Java™ language.

In his work on an algorithm for finding the shortest sailing distance from any maritime navigable point to a designated port (Beeker, 2004)[23]. He explained Many simulations and deployment analyses use networks to evaluate ship paths at sea. Ships are constrained to travel along links between specified nodes. When ships may need to change their destination while at sea, the network path can deviate significantly from an actual shortest path. Additionally, when it is desired to initialize simulation data from actual ship positions, ships must move to a node to use the network. By triangulating open navigable water areas, a shortest path from all points at sea to selected ports can be calculated. A Delaunay triangulation of the open areas can be calculated in $O(n \log n)$. Using the Delaunay triangulation, the shortest path triangulation from a point (port) can be calculated in $O(nk)$ where n is the number of points specifying the sea-land border and κ is the number of islands. The resulting data can provide distance from any point on the sea surface to the port in *O*(1) time when the triangle containing the point is known, or in $O(n/2)$if the triangle must be determined.

In their work on the algebra of airlift (Brigantic and Merrill, 2004)[23]. This paper describes fundamental algebraic relationships that characterize the movement of cargo and passengers via strategic aircraft that comprise an air mobility system. In particular, we present formulae for computing the number of airlift missions required to deliver a given force a given distance; relationships for estimating the throughput or average daily tons that can be delivered to a particular airfield based on that airfield's operating characteristics; and formulae for estimating the force total closure time based on airlift fleet parameters and the intended operational scenario. We also provide a methodology to compute million ton-miles per day (MTMs/D), a common metric for conducting air mobility analyses and force structure capacity assessments.

In their work on the Using low-dimensional patterns in optimizing simulators: An illustration for the military airlift problem (Powell, Wu, and Simao, 2004)[23]. They explained optimization models are sometimes promoted because they provide “optimal” solutions as defined by a cost model. Simulation models, by contrast, are guided by rules that are specified by experts in operations. While these may seem heuristic in nature, they often reflect issues that are difficult to capture in a cost-based objective function. “Optimizing simulators” combine the intelligence of optimization with the flexibility of simulation in the handling of system dynamics, but still suffer from the limitation that the behavior is entirely determined by a cost model. In this paper, they show how a cost-based model can be guided through a set of low-dimensional patterns which are essentially simple rules determined by a domain expert. Patterns are incorporated through a penalty term, scaled by a coefficient that controls that tradeoff between minimizing costs and minimizing the difference between model behavior and the exogenous patterns.

In their work on The covalidation of dissimilarly structured strategic airlift models (Wright, Bauer, and Oxley, 2004)[23]. They presented a methodology which allows comparison between models constructed under different modeling paradigms. Consider two models that exist to study different aspects of the same system, specifically Air Mobility Command's strategic airlift system. One model simulates a fleet of aircraft moving a given combination of cargo and passengers from an onload point to an offload point. The second model is a linear program that optimizes aircraft and route selection given cargo and passenger requirements in order to minimize late and nondeliveries. Further, the optimization model represents a more aggregated view of the airlift system than does the simulation. The two models do not have immediately comparable input or output structures, which complicates comparisons between the two models. They develop a methodology that structures this comparison and use it to compare the two large-scale models described above. Further, their technique has the fortunate byproduct of improving the fidelity of the models through a series of iterative refinements of the input to each model based on the output of the other. Models that compare favorably with regard to their methodology are deemed covalid models.