APPLICATION OF GENETIC ALGORITHMS TO AN AIRLINE-NETWORK SCHEDULING

Koji URATA,
Manager,
Hokkaido Engineering Consultants
Sapporo, Hokkaido, 062 JAPAN
Fax: +81-11-852-1783
E-mail: ku432@mb.docon.co.jp

Keiiti SASAKI, M. Eng.
Doctor Course Student
Department of Civil Engineering
Muroran Institute of Technology
Muroran, Hokkaido, 050 JAPAN
Fax: +81-143-47-3411

Toru TAMURA, Dr. Eng.
Associate Professor
Department of Civil Engineering
Muroran Institute of Technology
Muroran, Hokkaido, 050 JAPAN
Fax: +81-143-47-3411
E-mail: tamura@oyna.cc.muroran-it.ac.jp

Kazuo SAITO, Dr. Eng.
Professor
Department of Civil Engineering
Muroran Institute of Technology
Muroran, Hokkaido, 050 JAPAN
Fax: +81-143-47-3279
E-mail: k-saito@muroran-it.ac.jp

Akira KAWAMUR, Dr. Eng.
Associate Professor
Department of Civil Engineering
Kitami Institute of Technology
Kitami, Hokkaido, 090 JAPAN
Fax: +81-157-23-9408
E-mail:kawamura@stce2.civil.kitami-it.ac.jp

Hussein S. LIDASAN, Dr. Eng.
Associate Professor
National Center for Transportation Studies
University of the Philippines
Diliman, 1101 Quezon City
Philippines
Fax: +63-2-929-5664

abstract: Genetic Algorithms(GA) includes generally three genetic operators, selection, crossover and mutation. The lack of dependence on function gradients makes it more suitable to such problems, like as discrete optimization design problem and optimization design problems with non-convexities or disjointing in design space. GA is to be suitable as a new approach for such problems. In this study, it has attempted to apply GA for an optimization of airline-network scheduling and to compare it with enumeration method. It is suggested that GA has an applicability for large calculation inherent in airline network optimization problems and is more effective method for large size networks.

1. INTRODUCTION

Because of the internationalization of regional airports and the development of double or triple tracking, the structure of airline networks has become complicated these days. This complicated situation is compounded by the increase in the number of new airports. In addition, airlines have to consider the most suitable network scheduling. Such schedules are based on their previous experiences, because it is difficult to forecasts the movement of demands and to plan schedules for aircraft and crews. Although there have been many studies on scheduling, there have not been any practical ones concerning the difficulty of finding the best model.

At present, much interest is being focused on East Asia and South East Asia and their need for adequate network systems. How these systems should be developed or restructured remains a point for debate. Similarly, interest has also been focused to American's CRS (Computer Reservation System). The focus is on how to make the most of this network for the future. Europe's express Railroad Planning has also attracted interest, especially in the area of rail and air linkage. Added to these concerns, Japan's problems with Narita (New Tokyo International) and Haneda (Tokyo International) Airports which are operated at full capacity and as the placement of an Asia Hub in Japan have sparked interest and further debates.

As a result, research into airline networks has been paid an attention in Japan. The research has expanded to other areas as well, such as the demands of international sightseeing tours, airline company marketing, and air traffic controllers. However, little of the research has focused on the development of airline network optimization. Of the research on an optimization, U.S.'s NASA and MIT are leading in this field, although their research has yet much to do for production of good results.

Robert W.Simpson(1969) developed the computer-based model for the mathematical method or scheduling. He was followed by such researchers as Y.Chan(1972), S.E.Eriksen(1978) and others, who produced the model for the dynamic planning method, after regulations were relaxed in 1978 in America. However, little new wits produced there after 1978. In Japan researchers produced models which included frequency or displacement time[Tamura, T. and Inano, S.(1987)] or relation of frequency and demand[Sugai, Y. and Hirata, H.(1990); Tokunaga, K. and Inamura, H.(1992)]. During these periods, European companies involved their researchers and managers in producing models for the decision support system.

Today, research on the development of airline optimization is faced with the following problems:1) Large airline network calculation, 2) Market competition and calculation, and 3) Multi transit network calculation. It is found that the mathematical method of scheduling involving large calculations is inefficient usage [Etschmaier, M. M. and Mthaisel, D. F. X.(1985); Tokunaga, K. and Inamura, H.(1992)]. Therefor, it is needed to search a new optimization method. To optimize the airline-networks coping with these problems, it seems to be suitable for application of Genetic Algorithms (GA) as a new optimization method because it is said to be suitable for discrete optimum design, and recently has gained much attention.

In this study it has attempt to apply GA for an optimization of airline-network scheduling. The objective is to examine an applicability and an efficiency of GA by comparison with the enumeration method through a case study. The results suggest that GA has an applicability for the large calculation inherent in airline network optimization problems and is more effective method for large size networks.

2. FORMULATION OF AIRLINE NETWORK OPTIMIZATION PROBLEMS BY GA

2.1 Characteristic of GA

GA, unlike many conventional search algorithms, can be viewed in the two points:

(1) In conventional optimization methods, for example, the step-wise method, a single point is considered based on some decision rules. Even GA is a step-wise method, in that it considers many points in the searched space simultaneously, and in that evolution of nature is applied to the optimum design[Sugai, Y. and Hirata, H..(1990)].

(2) The main problem in combinatorial optimization is the convergence of a local optimum solution. GA intends to avoid this problem by using the probability rules to guide its search, i.e. the individuals with low fitness are also allowed to survive the next generation in it probability. This is somewhat similar to the simulated annealing method[Kobayashi, S. (1993A)]. Although there have been some studies in which the optimum search can be cloned efficiently by GA, there has been less progress on research concerning the theory in the behavior of GA[Kobayashi, S.(1993B)].

2.2 Genetic Process of GA

There are three essential genetic processes in GA: namely, reproduction (selection), crossover and mutation. The basic principle and procedure are explained briefly as followings:

At the outset, there must be a code or scheme that allows for a bit string representation of possible solutions to the problem. This coding of the string is the equivalent or a 'chromosome' in nature, i.e. the string that has the information of optimization just as the chromosome or a living being carried by genetic information.

The optimization process starts by random selection of an initial population of strings, where a string means an individual, or a possible solution.

The next process is reproduction (selection). First, the fitness of each string in the initial population is calculated. In GA, the value of fitness is used to decide which strings will survive (reproduction) in the following generation, and which strings will die (selection). High fitness value means that the string is fitter to the environment, and thus produces more children in the following generation. Here, the evaluation of fitness for each string is accomplished by the value of fitness function of that string. The fitness function of the unconstrained problems can be obtained simply by transforming the objective function. The transformation methods are concerned with how to prevent the solution from being a local optimum which has been a main problem in combinatorial optimization. The transformation in this paper will be explained later. The point here is that the strings with the higher fitness have greater probability of leaving more children in the following generation, and at the same time, the strings with less fitness die away in it at high probability, because normally the population is kept unchanged. The crossover process allows the characteristics between two strings to be exchanged to create two new strings. It starts from selecting two parent strings at random from the mating pool, which is made by the reproduction process, and then a random location along the string is chosen. The strings are then separated at this point, and recombined by exchanging the parts behind this point to form two new strings.

Mutation is the third step in these genetic processes, and is one that safeguard the process from a complete premature loss of valuable genetic materials during the reproduction and crossover. Mutation is applied with a low probability to the strings, determined by a random location on the strings, and the switching of the 0 or 1 at that location.

The processes described above are repeated until a convergent condition is satisfied. There are normally three conditions to satisfy convergence. One is that the value of fitness function has not been updated after some generations. The second is that the percentage of the strings with the highest fitness in the population is getting large. Finally, the maximum iterative generation.

2.3 Optimization of airline network

The optimum is calculated by giving the simple airline network schedule optimization problem to the test data. The result by GA is compared with the result by enumeration. Next, the optimum is calculated for more complicated network.

A hub-and-spoke type of airline network (Figure 1) is considered as a subject. The design function for optimization is basically the maximization of the fare income of the airline. Here, the fare for each air route is assumed to be uniform and the maximization of the number of passengers per day as the design function is estimated.

The details of each condition is determined as follows:

1. For the number of aircraft, the calculation for two, three, and four aircraft is done.
2. The number of seats is assumed to be nineteen (19).
3. The time which the aircraft is operable is assumed to be nine hours between 8:00 a.m. and 5:00 p.m.. Landing after 5:00 p.m. is allowed but departing after that time is not.
4. The travel time between airports shown in the figure includes the time for maintenance.
5. The airport No.1 in Figure 1 is the airport from which all the aircraft depart at the beginning of the day and return to the end of the day.
6. It is assumed that an aircraft is in service with passengers on board.

In designing a flight schedule, the differences in departure times poise a large problem. Specifically, the problem is to find the aircraft departure times closest to the times when passengers wish to depart. In this study, data from a 1986 survey of passengers at Okadama airport was used to construct a graph showing the relation between time of departure and demand. The number of passengers was based on the following two assumptions:

1. Each unit of time equals 30 minutes and if departure approximates to the time which the passengers wish to leave, all passengers take the flight.
2. Otherwise, each unit of time farther from the desired flight produces a corresponding decrease in passenger number.

This concept is expressed in Figure 2. The block area shows the demands of departure time *k*. More detail explanations on the equation of time and demand, can be seen in the paper by Tamura and Inano (1987).

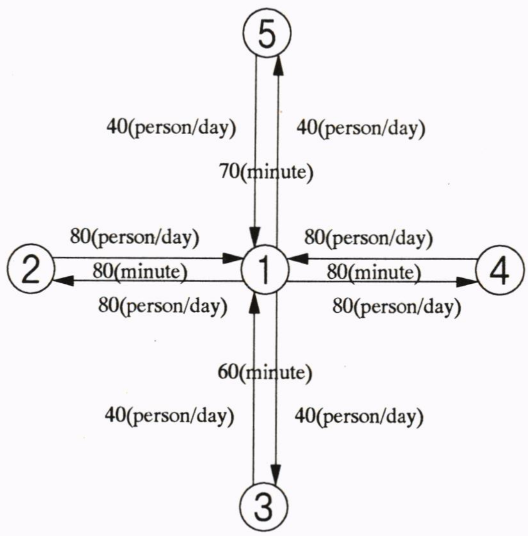

Figure 1. Airline Network

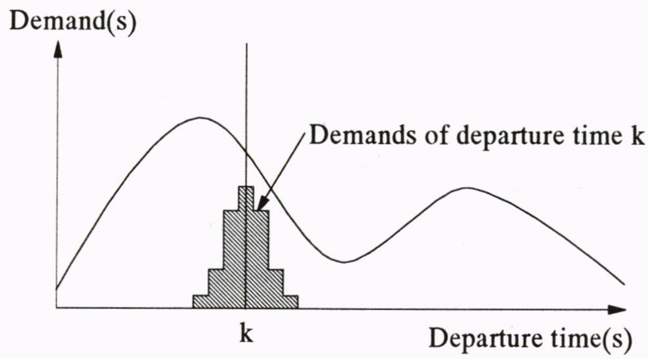

Figure 2. Departure time(s) and demand(s)

2.4 Formulation of GA

For the formulation of GA, the following criteria would be discussed here:

a) coding of design variables
b) fitness function and reproduction(selection)
c) crossover
d) mutation
e) convergence

a) Coding of design variables

The route of aircraft can be represented by design variables. One design variable has 16 sub-variables from 0 to 15, because it can express the 4 bits length of the binary scale. Then, those 16 variables are applied to the 4 routes (Fig. 1) as follows:

route		design variables
1 - 2 - 1	;	0, 4, 8,12
1 - 3 - 1	;	1, 5, 9,13
1 - 4 - 1	;	2, 6,10,14
1 - 5 - 1	;	3, 7,11,15

For example, it string with the arrangement of numbers as { 2 8 5 4 1 6 3 7 } means that the aircraft flies the route 1 - 4 - 1 (design variable is 2) first, and then it flies the route 1 - 2 - 1 (design variable is 8), 1 - 3 - 1, ... , 1 - 4 - 1 and 1 - 5 - 1.

In this study, the use of 4 length variables was applied for binary scale, because it is appropriate for efficient calculation. Furthermore, utilizing two lengths in the algorithm will be limited in the cross-over stage. As such, the 4 length variables were employed instead. This will provide more allowances in the iteration process. This, therefore, will yield efficient computations of larger lengths (5, 6, etc.) that require much calculation.

The maximum length of the string is 160 figures in the binary scale (40 design variables) in this study. More than 28 variables are enough, because a total of 4 aircraft with the limit of 1 flight route per day is being analyzed.

In this study, the design variable and strings have to be arranged, because the calculations include several aircraft and each aircraft has a different acquired actual time. Therefore, the string according to night time, route, number of aircraft, and acquired actual time are also arranged. In addition, the maximum flight, which is estimated from the flight time zone, is given as a condition. The strings equivalent to conditions in the GA process can be arranged. As this is a characteristic point of GA, the formulation of GA is considered to be very important.

GA considers many strings simultaneously (the number of strings being called population size), and in the beginning of optimization these strings are created by complete random choices using a random generator.

b) Fitness function and reproduction (selection)

Since the objective is the maximization of passengers per day in this study, the objective function can be used directly as a fitness function. However, in the early stage of optimization, the fitness values will present generally a larger dispersion, and on the contrary, in the later stage of optimization the fitness values of strings in a population may be very close to each other. A large dispersion of fitness values may cause some strings with low fitness to die easily, and this is undesirable at the start of GA runs. On the other hand, the evolution may become more difficult during the end of GA runs without the dispersion in fitness of strings. Therefore, it is necessary to transform the number of passengers into a fitness function by using the following equation (1).

$$F_i(t) = \max(a \cdot \Phi_i(t) - b, 0) \qquad (i = 1, N) \qquad (1)$$

where t is the number of generation, and coefficient a, b are defined as Equation (2).

$$a = \frac{\Phi_{avg}(c-1)}{\Phi_{\max} - \Phi_{avg}}$$
$$b = \frac{\Phi_{avg}(c \cdot \Phi_{avg} - \Phi_{\max})}{\Phi_{\max} - \Phi_{avg}} \qquad (2)$$

where ϕ_{avg}, $\phi_{\max}$ are the average and maximum number of passengers in a given generation, respectively.

Equation (1) is shown in Figure 3. The number of passengers (objective function) ranged $\phi_{\min}$ to $\phi_{\max}$ will be transformed into the value of fitness function ranged 0 for the lowest to $c \cdot \phi_{avg}$ for the highest.

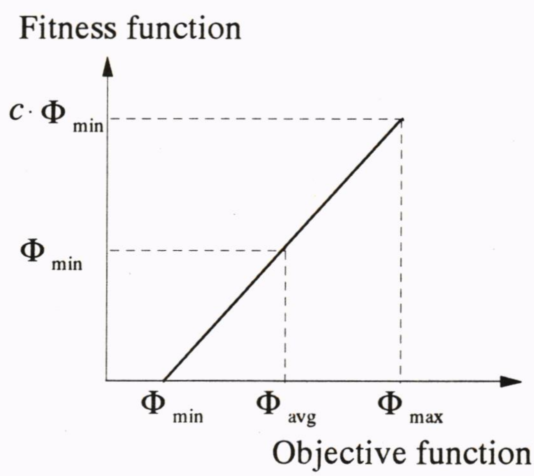

Figure 3. The objective function transformed into the fitness function

In the process of reproduction, the selection of strings for the next generation is accomplished by their fitness values calculated by Equation (1). At first, the average F_{avg} of $f_i (i = 1, N)$ is calculated and the strings with a higher fitness than F_{avg} will produce their children first. Equation (3) gives this number of children of *i-th* string left preferentially in the following generation:

$$n_i = INT\left[F_i / F_{avg}\right] \qquad (i = 1, N) \qquad (3)$$

where INT[•] means to take the integer from the bracket [ ]. Fitness value F_i varies with the value of c. As explained below, if c is assigned it large value, the string with the highest fitness will produce the most copies to the next generation.

Based on Equation (3), the number of each string left preferentially in the following generation is decided, and generally the sum of these numbers is less than N (population size). Therefore, the rest of the copies will be selected from the population in their own probability. This probability is defined in the following equation:

$$\overline{F_i} = F_i - n_i \cdot F_{avg} \qquad (i = 1, N) \qquad (4)$$

Corresponding to the value of $\overline{F_i}(i = 1, N)$, the remaining strings are determined. In this stage, the value of $\overline{F_i}$ of the string selected once is set to zero here.

New strings for the next generation are selected by the two steps explained above, and then the procedure moves to the crossover.

c) Crossover

Once parents (paired strings) have been selected, the crossover operator is applied to all parental pairs, which create the filial strings. Two crossover operators are considered in this study.

A crossover site is selected at random, then the parental pairs are separated into two part from this point. The longer parts are then received by the children pairs. The digits in the shorter portion of the strings are taken by another parent. Then these digits become the last digits of the succeeding children pairs.

The following example illustrates the procedure. In this particular case, the digits of the

Parent A : 2 8 5 4 1 ⋮ 6 3 7
Parent B : 7 4 *6* 3 8 ⋮ 1 2 5

right (shorter) part of Parent A are taken by Parent B where the last digit of the former became the first digit of the latter and the remaining digits (*6 3*) placed in the middle. Similarly, if Parent B's string is considered, its right (shorter) part becomes the left portion of Parent A's string. However, the order may not necessarily be the same. Finally, as mentioned earlier, the parent's strings are received by the children pairs as shown in the following example. The above sequential iteration is repeated for the whole crossover process.

Child A : 2 8 5 4 1 1 2 5
Child B : 7 4 6 3 8 6 3 7

In the crossover methods explained above, crossover is carried out with the same probability $pc(<1)$. When crossover is not carried out, two parental strings remained in the following generation without any change.

d) Mutation operator

Mutation is performed in two ways in this paper. One way is to select two locations along a string, and the bits between the two locations are reversed. The other way is by selecting two bits along a string at random and exchange them.

Furthermore, if the objective function (number of passengers) is no longer improved after a number of generations, the optimization may be convergent, or premature convergent may occur. In this study, the probability of mutation is set as 0.001.

e) Convergence criterion

There must be some convergence criteria to stop GA runs before all strings become the same, since it is not necessary to run GA until all strings become equal. Finally the best result through the whole optimization history is used as the optimum. In the present paper, the optimization procedure is stopped when one of the conditions shown below is satisfied:

a) Maximum generation (100th generation in this paper).
b) The number of passengers has not been updated through 20 consecutive generations.

3. CASE STUDY

3. 1 Comparison of GA and the enumeration method

Using the data from Figure 1, GA and the enumeration method can be compared. The program for the application of GA in this study was developed by the authors using FORTRAN language. The results, shown in Table 1, give the same optimum routes for 2 planes as for 3 planes. Table 2 shows the calculation time needed for 2, 3, and 4 planes using Fujitsu M-380 computers. It is clear that the calculations using GA take a short time as compared with that using the enumeration method. The reason for the time lapse in the enumeration method in each additional route is considered only after the preceding route has been decided. In case of GA all data is considered simultaneously.

Likewise, when considering 4 planes and the 162 routes generated, the enumeration method creates an overflow making calculation no longer possible. GA, however, manages to complete the calculation in only 3 seconds.

Table 1. The optimal solution (enumeration method and GA)

	Routes	Number of passengers
2aircraft	121215121 151314141	295
3aircraft	151414131 131412121 121515141	381

Table 2. Comparison of calculation time (enumeration method and GA)

	enumeration method (second)	GA (second)
2aircraft	4	3
3aircraft	1 1 3	1
4aircraft	2 4 0	3

(2) Application

Using the network in Figure 4, the optimum routes, with the longer time frame of 5 a.m. - 11 p.m. can be calculated. The more complicated two-stop route for the application example is employed. The increase in complexity of this route precludes the use of the enumeration method. Therefore, GA is used.

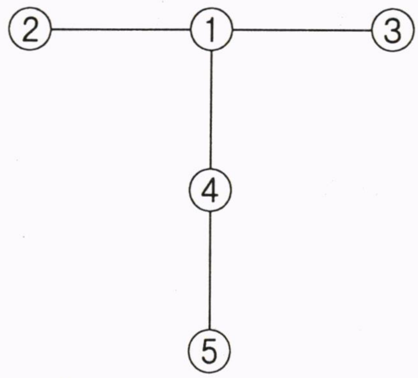

Figure 4. Airline Network (2)

In the previous network, four sub-routes were possible, while in this network the sub-routes increase to six. Although only two sub-routes have been added, the calculation has increased significantly by the squaring of the sub-routes. Table 3 shows the network data for OD, demand, and actual time. Actual time includes maintenance time and transfer time. The aircraft used can accommodate 100 passengers. The distribution of demand data has been extrapolated from the data of the previous network. The origin of the aircraft each day is indicated by 1. GA's probability is again set at 0.001.

Table 3. OD-Demands and Flight time

OD	Demand (person)	Flight times (minutes)
1-2, 2-1	880	90
1-3, 3-1	820	90
1-4, 4-1	370	40
4-5, 5-4	200	60

Because the aircraft must always end at 1, the 16 variables have been applied to the 6 routes as follows:

route		design variables
1 - 2 - 1	;	5 ,11
1 - 3 - 1	;	0 , 6 , 12
1 - 4 - 1	;	1 , 7 , 13
4 - 5 - 4	;	2 , 8 , 14
1 - 4 - 5 - 4	;	3 , 9 , 15
1 - 4	;	4, 10

The enumeration method deals will with continuous routes only but has difficulty with non-continuous routes and with directing all aircraft back to 1 as the final destination. On the other hand, GA has no such problems. The calculations of airplane 2, 3, and 4 are presented in Table 4. Because of the size of the calculations, no results can be given in this table for the enumeration method. Several points concerning GA are evident.

1. GA makes calculations of the new routes within 10 seconds.
2. The longer calculation time of GA reflects the increase in the length of the strings.

In addition the rate of meeting all the demands of the passengers increases with the addition of aircraft number as follows:

aircraft number	demand increased
two planes	56%
three planes	78%
four planes	96%

Finally, although the complexity of the calculations was increased though the use of a different network, GA was found to have no problem making the calculations efficiently. Future studies should be made using even more complex networks and calculations to ensure the feasibility of GA in scheduling.

Table 4. Routes and calculation time

Number of aircraft	Routes	Number of passengers (person)	Calculation time (seconds)
2	141413121312121 12121454131314141	2520	2.9
3	14131413121414131 12141414541212131 12145412131454141	3877	4.3
4	121212121213141 14541213121414541 14131413145413121 14541314541412131	4337	9.0

4. CONCLUSION

In this paper, the GA was applied to problems of airline-network scheduling and the findings indicated that its efficiency was proven. The comparison of the GA method to enumeration method showed that the former is more efficient than the latter in airline network scheduling.

The formulation of GA was tried for airline-network problem and the computer program was developed by using FORTRAN language. Although the results in the case study showed the same optimum routes for 2 planes as for 3 planes, the calculations using GA took very short time as compared with that using the enumeration method. It is the reason that in case of GA all data can be treated simultaneously.

As a result, it was shown in this paper that the specific features of GA were put into practical application, such as a) application of data of aircraft to a string; and b) use of aircraft's routes as design variables.

It is shown that GA will provide a significant capability for complex combination optimization problems. However, the paper does not show the complexity of GA adequately. Further research is needed to give the full range of possible applications of GA to optimization problems on the complete airline network, and other networks such as express railroads combined with airlines.

Further research should also design GA's environment to include more specific characteristics, such as demand structure prediction, aircraft consolidation and crew placement. However, it was shown by this paper that GA is suitable for problems with rough conditions rather than clear and analytic problems, from the viewpoint of this process.

REFERENCES

, Y.(1972) Route Network Improvement In Air Transportation Schedule Planning. **MIT Department of A&A Report 1912**.

Eriksen Steven, E.(1978) Demand models for US Domestic Air Passenger Markets. **MIT Department of A&A Report 1918**.

Etschmaier, M.M and Mathaisel, D.F.X (1985) Airline Scheduling; **An Overview Transportation Science**, Vol. 19, No.2.

Holland, J.H(1975) **Adaptation in Natural and Artificial Systems**. University of Michigan Press.

Kobayashi, S.(1993A) Perspective of Genetic Algorithm. **Journal of the Society of Instrument and Control Engineers** (material for lecture).

Kobayashi, S. (1993B) Perspective of Genetic Algorithm. **Journal of the Society of Instrument and Control Engineers**, Vol.32, No.1, pp2-9.

Oniyanagi, Y., Tokunaga, K. and Inamura, H.(1992) Integer Quadratic Programming Model for Aircraft Scheduling. **Proceeding of Infrastructure Planning 15(1)-2**, pp597-602.

Simpson Robert, W.(1969) Scheduling and Routing Models for Airlines Systems. **MIT Department of A&A Report 1969**.

Sugai, Y. and Hirata, H. (1990) Combinatorial Optimization Algorithms and Their Applications. **Journal of the Society of Instrument and Control Engineers**, Vol.29, No.19 pp1084-1091.

Tamura, T. and Inano, S.(1987) Scheduling and Routing Models for Regional Airline Systems. **Infrastructure Planning Review 5**, pp155-169

Tokunaga, K. and Inamura, H.(1990) Aircraft Scheduling Models for General Type of Network. **Proceeding of Infrastructure Planning 13**, pp607-614.

Tokunaga, K. and Inamura, H.(1992) Aircraft Scheduling Models by Using Dynamic Programming Approach. **Journal of Infrastructure Planning and Management**, No.440/IV-16, pp109-116.