

# **Optimization Applied to Transportation Systems**

**Strobel, H., Genser, R. and Etschmaier, M.**

**IIASA Collaborative Paper  
December 1977**



Strobel, H., Genser, R. and Etschmaier, M. (1977) Optimization Applied to Transportation Systems. IIASA Collaborative Paper. Copyright © December 1977 by the author(s). <http://pure.iiasa.ac.at/736/> All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage. All copies must bear this notice and the full citation on the first page. For other purposes, to republish, to post on servers or to redistribute to lists, permission must be sought by contacting [repository@iiasa.ac.at](mailto:repository@iiasa.ac.at)





# OPTIMIZATION APPLIED TO TRANSPORTATION SYSTEMS

Proceedings of the IFAC-IIASA Workshop  
held in Vienna, Austria, on 17-19 February, 1976

*Edited by*

Horst Strobel\*, Robert Genser\*\* and Maximilian M. Etschmaier\*\*\*

*Sponsored by*

The International Federation of Automatic Control (IFAC)  
(Systems Engineering Committee (SECOM))  
The International Institute for Applied Systems Analysis (IIASA)  
Laxenburg, Austria

*Organized by*

The Austrian Center for Efficiency and Productivity (ÖPWZ), Vienna

CP-77-7

December 1977

- \* IIASA and Hochschule für Verkehrswesen "Friedrich List", Dresden, GDR.
- \*\* Chairman of IFAC-SECOM Working Group on Transportation and Österreichische Bundesbahnen, Generaldirektion, Vienna, Austria.
- \*\*\* Department of Industrial Engineering, Systems Management Engineering and Operations Research, University of Pittsburgh, USA.

Views expressed herein are those of the contributors and not necessarily those of the International Institute for Applied Systems Analysis.

The Institute assumes full responsibility for minor editorial changes, and trusts that these modifications have not abused the sense of the writers' ideas.

International Institute for Applied Systems Analysis  
A-2361 Laxenburg, Austria



## Preface

In 1973, the Systems Engineering Committee (SECOM) of the International Federation of Automatic Control (IFAC) decided to establish a *Working Group on Transportation* to try to improve information exchange between control and systems scientists, and experts working in the different fields of transportation. The organization of workshops dealing with the application of control sciences in transportation systems was formulated as one special task of this group.

This volume summarizes the results of the first Workshop initiated by this Working Group which took place in Vienna (Austria) 17 to 19 February, 1976, and was dedicated to problems of *Optimization Applied to Transportation Systems*. Since this was the first IFAC workshop dealing with transportation problems, it was considered useful not to restrict the program to a special mode of transportation but to start with a very broad systems analysis view. Therefore, the following three levels of transportation systems analysis approaches were chosen as a framework:

- Transportation systems planning from a socio-economic point of view;
- Operational planning, mainly focusing on routing and scheduling problems; and
- Control and guidance of transportation systems.

One day was dedicated to each of these sessions, where formal presentations dealing with the state-of-the-art as well as with more specific topics were followed by panel discussions. In these panel discussions special emphasis was placed upon the following questions:

- What methodology is or could be used for the solution of optimization problems occurring in transportation systems?
- What similarities and differences exist between optimization problems in air, rail, sea, road, and other transportation systems?
- What problems are still unsolved, and what are the most important and promising areas for future research work?

The answers which the workshop provided to these and similar questions during the discussions, have been summarized for each of the three sessions and included in this volume, together with *abstracts of all* the presentations and full-length papers of about 60% of all presentations.

Several institutions and individuals gave their assistance to attain the targets of the workshop. The IFAC Systems Engineering Committee (SECOM) as the sponsor of the workshop gave strong support from the beginning of the planning. Moreover the International Institute for Applied Systems Analysis (IIASA) in Laxenburg (near Vienna) agreed to serve as co-sponsor. The support given by IIASA in preparing the workshop program as well as in editing and publishing this volume is highly appreciated. The National Member Organization of IFAC, i.e. the Austrian Center for Efficiency and Productivity (ÖPWZ) serviced as the local organizing committee, thus providing the necessary financial and organizational backing.

In connection with the IFAC Workshop in Vienna, IIASA organized, during the same week, a Planning Workshop in Laxenburg dealing with the identification of promising future research directions in the field of Transportation Systems Analysis. The results of that meeting have been published in the IIASA paper, CP-76-11.

Horst Strobel  
Robert Genser  
Maximilian Etschmaier



## Acknowledgments

The support given by the following organizations and individuals is gratefully acknowledged:

- The International Federation of Automatic Control, Systems Engineering Committee, Working Group on Transportation Systems:

R. Genser (Austria) (Chairman), M.Y. David (France), H.V. Ellingsen (Norway), A.A. Giraud (France), T. Ishii (Japan), Y. Ohno (Japan), G. Schmidt (FRG), H. Strobel (GDR), R. Vacca (Italy), and D. Van Meter (USA).

- The International Institute for Applied Systems Analysis (IIASA) Laxenburg, Austria:

H. Raiffa (former Director), R. Levien (Director), for taking on the co-sponsorship, and H. Strobel, for organizing IIASA's contribution.

- The Austrian Center for Efficiency and Productivity (ÖPWZ) - Division for Automation and Technical Development, Vienna, Austria:

H. Reischitz (Chairman), K. Lichtblau, W. Karner, and R. Vogel, for serving on the local organizing committee.



## Contents

### TRANSPORTATION SYSTEMS PLANNING

Summary of the Panel Discussions	3
Abstracts of Papers	5
<i>Concepts and Methods Used in Multicriterion Decision Making: Their Applications to Transportation Problems</i> B. Roy and E. Jacquet-Lagrece	9
<i>The Application of System Dynamics Simulation Methodology for Analyzing the Social Impact of a High-Speed Mass Transportation System</i> T. Hasegawa, Y. Ogawa, and M. Watanabe	27
<i>Evaluation of Public Transportation</i> O.H. Jensen	53

### OPERATIONAL PLANNING

Summary of the Panel Discussions	69
Abstracts of Papers	71
<i>A Survey of Scheduling Methodology Used in Air Transportation</i> M.M. Etschmaier	77
<i>Scheduling and Maintenance Planning in Rail Transportation</i> R. Genser	91
<i>Optimal Planning of Shipment Volumes</i> O.I. Aven, S.E. Lovetsky, and G.E. Moiseenko	119
<i>Optimal Scheduling of Trucks - A Routing Problem with Tight Due Times for Delivery</i> O.B.G. Madsen	129

<i>Optimal Bus Routing: A Case Study</i> C. Mandl	137
--	-----

#### CONTROL AND GUIDANCE OF TRANSPORTATION SYSTEMS

Summary of the Panel Discussions	151
Abstracts of Papers	153
<i>Optimization and Evaluation Techniques for Street Traffic Control</i> J.L. Schlaefli	161
<i>Optimal Decongestion Techniques for Urban Traffic Networks</i> R. Camus, A. D'Amore, and P. Sipala	173
<i>Design of Control Systems in Automated Transport Systems</i> L.D. Burrow	187
<i>Optimal Control of Railway Marshalling Yard Operations</i> J. Sokolowski	195
<i>Optimal Vehicle Control from the Point of View of Energy Conservation</i> L. Skyva	211

#### APPENDIX

List of Participants and Authors	227
----------------------------------	-----

## **TRANSPORTATION SYSTEMS PLANNING**



### Summary of the Panel Discussion\*

The discussion on the planning of transportation systems was rich and varied. The number of opinions expressed as to appropriate methodologies and relevant problems in transportation planning research was too large even to be listed here.

Some people stressed the need to recognize explicitly the multi-attribute, disaggregate nature of transportation systems, in terms of both their characteristics and their consequences. Transportation decisions affect different social groups and different regions in a variety of ways. These distributional effects should be emphasized along with the overall magnitude and narrowly defined efficiency considerations of transportation investment. The evaluation of plans within a multi-criterion framework was strongly advocated.

Considerable discussion took place on broadening the scope of transportation planning models--to view transportation as an integral part of a broader socio-economic-environmental system. In practical terms, however, it was recognized that transportation investment models could not hope to embrace all of the complex consequences of transportation systems. There was considerable agreement that insofar as optimization methods were relevant to transportation plan development, their role should be restricted to that of generating broad alternative network configurations and their sequencing over time. Alternative objective functions and constraints should be used to delimit some first approximations to transportation network plans. The dynamic nature of planning should be recognized - dynamic development models rather than equilibrium or fixed horizon models should be emphasized. The results of these "rough" models would then be subjected to more intensive analysis and evaluation. A few participants were of the opinion that optimization models were of limited use in the design phase and that alternatives could be generated on a more ad hoc basis. On the whole, however, there appeared to be an agreement that mathematical programming and perhaps even optimal control methods could be useful in the design phase, keeping in mind that subsequent analysis, evaluation and adjustments to the initial design would be mandatory.

With respect to evaluation and plan selection, there was a strong feeling that the decision- or policy-maker should be closely involved in the modelling process. This is in part a problem of effectively communicating the results of models in terms that are meaningful to the policy maker. That is, the

---

\*Prepared by R.D. MacKinnon.

models should have a transparency that allows the policy-maker to understand the basic assumptions and principles underlying the technical analysis. Even more ambitious and rewarding, perhaps, would be the establishment of a framework in which the policy-maker could assist in the model design itself. If done effectively, this would almost certainly increase the chances of successful plan implementation.

Other speakers stressed the need for follow-up evaluation studies. All too often plans are implemented without the ensuing results being monitored so that subsequent plan-making can benefit from experience. This learning feedback loop should be a part of any effective on-going planning process.

There were several suggestions concerning the desirability of applying concepts of formal systems theory to transportation planning. Is it possible to model transportation/land-use interactions as formal closed-loop systems? Can the models of hierarchical control be effectively applied to the planning of transportation systems? Would game theoretic approaches be useful in resolving group conflicts that inevitably arise in transportation planning? How useful would stochastic process modelling, and more particularly stochastic programming, approaches be in the context of transportation planning? Certainly, some explicit recognition that we are dealing with relatively poorly understood systems that are subject to random shocks should be reflected in transportation planning models.

A distinction was made between descriptive (numerical) and explanatory (theoretical) models. Often in asking for advice on transportation issues, the policy-maker is less concerned with precise quantitative answers than with broad, qualitative information on which he can move towards developing policy guidelines. Theoretical, qualitative models (hopefully, but not necessarily, of a formal rigorous type) are needed to make recommendations on general policy. The precise implementation of this policy will subsequently be facilitated by larger numerical models.

Clearly, long term transportation planning is a very complex process indeed. Any attempt to model this process in formal terms is going to be partial, and ultimately inadequate. This is not to say, however, that such attempts are doomed to be purely academic exercises. Over the past two decades, our knowledge of transportation systems has certainly increased, but so has our knowledge of how much we do not know. There are a large number of interesting, promising, and relevant areas of research in transportation planning. Optimization models clearly have a major role to play in these future developments; but the researcher must be careful not to make extravagant promises or his credibility and the credibility of all optimization models in this context will be jeopardized.



### Abstracts of Papers

*Optimization Models of Transportation Network Improvements:  
Review and Future Prospects*

R.D. MacKinnon and G.M. Barber

(Already published by IIASA as Research Memorandum RM-76-28)

The paper briefly reviews the alternative approaches to spatial improvements in transportation networks from the early linear programming attempts to the more recent discrete programming approaches, the more analytical geometrical and optimal control methods, narrow cost minimization models, and the more comprehensive attempts to incorporate a broad range of economic and social impacts. Finally, some personal remarks are made concerning the most promising areas of future research with respect to practical relevance, computational feasibility, and theoretical interest.

*Concepts and Methods Used in Multicriterion Decision-Making:  
Their Applications to Transportation Problems*

B. Roy and E. Jacquet-Lagrezé

(This volume, pp. 9-26)

The purpose of the paper is to present, with the help of a few examples of transportation problems, such as airport location, what the contribution of multicriterion decision-making is in the field of transportation studies. Attention is given to the various aspects of modelling connected with the main objective of getting a good fit between the study and the decision process.

- First, how can we build and model the set of alternatives? And is it always alternatives in an exclusive sense: i.e. if the decision-maker chooses alternative A, he will not choose alternative B? And has the problem formulation always to be to select the optimal alternative? The answer is no. There are other operational issues based on the notion of "potential actions". Different problem formulations can then be proposed: (1) select one and only one action considered as the "best", (2) select all actions which seem "good" among those studied, (3) select several actions among the "best studied".
- Secondly, one is confronted with the problem of modelling the various consequences of potential actions (cost investment, travel time and its various aspects, noise, etc.). The concept of a "consistent family of criteria"

is presented, jointly with a discrimination between different types of criteria according to the nature (ordinal, cardinal, measurable) of their significance, and the specific role played by thresholds.

- Thirdly, three different attitudes to the way of aggregating the different criteria are presented: (1) assess a value or utility function which is the traditional issue involved in cost-benefit analysis or in the more recent multi-attribute utility theory, (2) assess an outranking relation which is a binary relation whose purpose is to represent the reliable part of the preference of the decision-maker, and (3) aggregate the criteria in an interactive man-model way to explore, in a sort of learning process, what the preferences of the decision-maker are - how he balances the various criteria.

*The Application of System Dynamics Simulation Methodology for Analyzing the Social Impact of a High-Speed Mass Transportation System*

T. Hasegawa, Y. Ogawa, and M. Watanabe

(This volume, pp. 27-52)

This report deals with an analysis of the social impact of a high-speed mass transportation system, namely, the Shinkansen new rapid train system, on a local city when it is connected thereby to an extremely large city. A system dynamics model of the population behavior of the two cities was developed, by using a simulation model for Tokyo (the large city) and Utsunomiya (the local city) and carrying simulation runs with DYNAMO II. The results of the simulation runs show that the completion of the high-speed mass transportation system, Tohoku Shinkansen, that will connect Tokyo and Sendai via Utsunomiya and that is now under construction, will cause a destruction of the features of Utsunomiya as a regional center.

*Mathematical Modelling and Optimization of Complex Transportation Systems*

Z. Bubnicki

(Abstract only)

The complex industrial transportation system considered is composed of a number of sources of the transported material, receipt points, middle junctions, and transportation ways and means. The structural and nonstructural characteristics of the distinguishing elements were described, and a mathematical model for the whole system in the form of a set of linear and nonlinear equations and inequalities with continuous and binary variables presented. The optimization and control problems for this

transportation system were formulated and led to the problem of nonlinear binary continuous programming. The applications of the optimization technique were discussed, and a concept of the decomposition method given. Algorithms based on this concept were determined, and some experiences with the use of a digital computer for problem solving presented.

An example of the mathematical description and optimization of the transportation system in the raw material preparation department of a cement plant illustrated the possible practical applications of the general approach presented.

*Evaluation of Public Transportation*  
O.H. Jensen

(This volume, pp. 53-66)

The paper contains 3 main parts:

- an analysis of public transportation in rural areas in Denmark,
- an OR model to be used in the process of planning public transportation, and
- an example of the application of the model in a county in Denmark.

Owing to increasing car ownership and increasing costs of operation, the economic basis of public transportation is growing still worse. This has forced public authorities to intervene with subsidies, etc. The role of municipalities and counties as planners of public transportation is discussed.

The OR model consists of 4 submodels, which are of:

- the region,
- the household,
- car ownership of the household, and
- public transportation.

The aim is to emphasize the total system and the interaction between the submodels, rather than isolated planning and evaluation of public transportation.

The model has been used in a study in a county in Denmark and the results of this are presented and discussed.

*Methods of Defining Reliability and Safety in Traffic*  
K. Pierick

(Abstract only)

The criterion for assessing traffic safety in numerical form until now has been merely a description of the accident, either in absolute numbers (i.e. the total number of accidents, the total number of specific consequences from the accidents - body damage, injured, dead, etc.), or in statements on the accidents or their consequences in relation to transport performance (the number of accidents per driving kilometers, the number of injured persons per passenger kilometers, etc.). These numbers for the description of traffic safety result solely from observation and statistics, and are therefore only possible for transport systems already in operation. The question arises of how one can arrive at other numerical comparisons of traffic safety and reliability - comparisons which do not result solely from incidents leading to actual damage, and which are applicable for conventional as well as for newly planned transportation systems.

In principle, the problem of advance calculation of the probabilities of failures presents itself in transportation systems just as it does in mechanical and electrical engineering, where, as a result of the nature of the service-performance, one distinguishes between

- the probability of dangerous failures (synonymous with traffic safety), and
- the probability of operational-obstructing failures (synonymous with traffic reliability).

For reliability theory methods for the evaluation of transportation systems to be usable, the method should, on the one hand, distinguish in which way failures of individual elements operate on the whole system, and, on the other, be organized in such a way that for the smallest elements of the system, either a numerical description of their deficiency-behavior already exists, or one can be prepared at small cost. Calculations of the reliability of complicated transport systems are naturally not simple, and have not yet been completely realized. Preliminary investigations, however, suggest that the approach indicated above could be successful.

Concepts and Methods Used in Multicriterion  
Decision Making: Their Applications to  
Transportation Problems

B. Roy and E. Jacquet-Lagrezé

Transportation studies concerned with investment or exploitation problems largely use optimization techniques and cost-benefit analysis. Both are decision-aid techniques in the sense that they refer to the attempt, by more or less formalized models, to help a decision-maker to "make" a decision, or, at least, to give him a good knowledge of the feasible alternatives, their consequences, his own objectives, and his overall preference. It is the purpose of multicriterion approaches to focus attention on the modelling of preferences, for which it is necessary to assess an overall preference (the so-called global preference in this paper). Which criteria should be used in an optimization technique, and must one always aggregate subcriteria in an overall function? Which consequences or impacts should be taken into account in cost-benefit analysis, and must one always try to express these in monetary terms in order to aggregate them in an overall function? Multicriterion decision-making, without rejecting these traditional techniques, gives new concepts with which to assess preferences, and new methods with which to aggregate preferences (criteria). Often these appear to be easier to implement and to allow a better use of the study in the decision process. (See Godard (1973) and Roy, et al. (1975) for more work in the field of transportation.)

Three stages can be considered in modelling work:

- *Subject matter of the decision*: formal definition of the set A of potential actions, and choice of a problem formulation.
- *Consequences and valuation*: formal description of the consequences: dimensions or attributes, scales (units), valuation on each dimension, thresholds, and choice of a consistent family of criteria.
- *Global preference and comparison*: properties of the global (overall) preference: comparability or incomparability, transitivity or intransitivity, and choice of an operational attitude.

After the third stage, the analyst can start again at the first with a new set of actions because of the evolution of the decision process or because of the better knowledge of the problem that both the analyst and the decision-maker then have.

In this paper we present some of the aspects of each of these three stages which seem most important in the field of transportation studies. The reader will find a more detailed presentation and discussion of these three stages in Roy (1975).

SUBJECT MATTER OF THE DECISION

Usually decision-aid in transportation problems is understood as "find the optimal solution among a set of feasible alternatives". This approach supposes that the analyst should build a set A of potential actions, each action  $a \in A$  being an alternative solution to the transportation problem. The alternatives are, of course, mutually exclusive and the problem formulation can be to select the "best" alternative. To do this, the usual approach supposes also that the analyst is able to build an overall criterion or preference function which is a model of the global preference of the decision-maker. This is the case of a traditional cost-benefit analysis carried out to compare different sites for an airport location problem (see Frost, 1971) when each site is by itself a proper solution to build an airport. When it is not the case, exclusive alternatives are obtained by mixing the feasible sites, for instance (for an example, see Metra Consulting Group, 1972). The more recent approach of decision analysis (Raiffa, 1968) which uses multi-attribute utility theory does not change anything in this traditional problem formulation. In the study of airport development for Mexico (de Neufville and Keeney, 1972) over 4000 alternatives have been obtained and an overall utility function has been assessed on six attributes in order to select the best one.

In these examples the set A is globalized and fixed (see definition in Table 1) and for all of them the problem formulation is: select one and only one action considered as "the best".

Table 1. Four cases for the modelling of the set A of potential actions.

		The elements of A are mutually exclusive	
		YES	NO
The set A is a priori defined in a strict and exhaustive manner (by a rigid frontier, non ambiguous test of membership, etc.)	YES	case globalized and fixed	case programmed and fixed
	NO	case globalized and evolutive (or flexible)	case fragmented and evolutive (or flexible)

But this very common case is not the only one the analyst should consider (see Tables 1 and 2). Very often problems can either not be formulated in this first way or if they can, only

Table 2. Different problem formulation of A.

The objective of the problem formulation is to select		$\alpha$ "one"	one and only one action considered as the "best"
		$\beta$ "all"	all those actions which seem "good" among those studied
		$\gamma$ "some"	several actions among the "best" studied

in a very artificial and complicated way. Consider for instance the decision problem of an administration such as the Direction des Routes in France, responsible for investments in building new roads or improving existing roads. A certain amount of money is available each year, and a certain number of investments projects arrive throughout the year at the central administration for examination and acceptance or rejection. If A is the set of projects presented in one year, it appears at the end of the year that a decision is in fact an element of  $Q(A)$  (set of all subsets of A). But to help the administration make a decision, one cannot simply find an alternative (an element of  $Q(A)$ ) maximizing some benefit under a cost constraint. Projects are submitted at various times during the year and it would be a great loss of time and simply unrealistic to wait until the end of the year to make a decision. So, in fact, the Direction des Routes has to decide for each potential action  $a \in A$  (each project), whether or not it is accepted. In this example, potential actions are not mutually exclusive, and the projects are not alternatives. Nevertheless, it is useful to help the decision-maker to choose by considering whether A is defined in an exhaustive manner (A is then evolutive) or as a set of mutually exclusive alternatives (A is then fragmented). The problem formulation can be here  $\beta$  (see Table 2). More precisely, the analyst can use a procedure building a trichotomy of A:  $A = A_1 \cup A_2 \cup A_3$ ,  $A_l \cap A_h = \emptyset$  for  $l \neq h$ , an action  $a \in A$  being:

- in  $A_1$  if it is accepted without special intervention by the decision-maker;
- in  $A_3$  if it is rejected without any special intervention;
- in  $A_2$  if it requests a complementary examination (information, discussion, decision-maker judgment on this particular case, etc.).

## CONSEQUENCES AND VALUATIONS

The purpose of the second stage in the modelling is an analysis of the consequences or impacts of the potential actions (projects, etc.) in order to build a family of criteria. We first present the general concepts and then illustrate these with an airport location problem.

### General Concepts and Definitions

Step 1: Qualitative description of all the consequences of potential actions occurring both in the short and long-range: economic, social, and environmental consequences or impacts. In transportation studies, it is often easier to study these at three levels (see Roy, et al., 1975):

- individual level: travel time, comfort, travel cost, waiting time and its uncertainty, etc.;
- collective level: social and economic impacts, environmental impacts, etc.;
- structural level: long range modifications caused by the project studied (location of new industries, housing, etc.).

Step 2: Definition of a set of dimensions  $v = \{1, \dots, i, \dots, \bar{n}\}$  considered as necessary and sufficient for the problem to make a formal description of the consequences listed in Step 1. A scale  $E_i$  must be associated to each dimension  $i$ . The scale can be physical units (e.g. travel time in minutes) or a qualitative ordinal scale (expert judgments). We shall call a grade  $e \in E_i$ , an element of the scale.

Step 3: Valuations of the action  $a \in A$  on each dimension. The state indicator  $\gamma_i(a)$  is the operational means of estimating the valuation of  $a$  on the dimension  $i$ . In some easy cases,  $\gamma_i(a) \in E_i$  and the valuation is said to be single point. But often,  $\gamma_i(a) \subseteq E_i$  and the valuation is said to be nonsingle point. In the latter case, the analyst must obtain some additional information with the help of the modulation indicator  $\delta_i(a)$ .

The modulation indicator  $\delta_i(a)$  might be:

- Distributional:  $\delta_i(a)$  is then a distribution which quantifies the relative importance of each grade of the scale. If we denote by  $\delta_i^a(e)$  the mass assigned to each  $e \in \gamma_i(a) \subseteq E_i$ , then  $\delta_i(a) = \{\delta_i^a(e)\}$ .



- Relational:  $\delta_i(a)$  is the additional information making it possible to build a preference relation (preference, indifference, incomparability) between  $\gamma_i(a)$  and  $\gamma_i(a')$ . (For more details see Jacquet-Lagrez, 1975.)
- Single "indexed event": For a set  $E$  of exclusive events obtained for instance when building scenarios to describe the future, then  $\delta_i(a)$  characterizes each single point valuation  $\gamma_i^E(a) \in E_i$ . If it is possible to assess a probability distribution on  $E$ , then  $\delta_i(a)$  is a probability distribution on  $\gamma_i(a)$ .
- Complex "indexed event": Let  $E$  be a set of events. For each event the valuation might be a nonsingle point, so  $\gamma_i^E(a) \subset E_i$ .  $\delta_i(a)$  is then more complicated and has to give, for instance, the modulation for each of the  $\gamma_i^E(a) \subset E_i$ .

Step 4: Assessment of a consistent family of criteria.

A consistent family of criteria is by definition a family of  $n$  functions called criteria  $g_1, \dots, g_j, \dots, g_n$ , where the  $g_j(a)$  for  $a \in A$ , are real-valued functions defined with the help of the valuations  $[\gamma_i(a), \delta_i(a)]$  established on the  $\bar{n}$  dimensions; the number of criteria  $n$  and dimension  $\bar{n}$  might be different since some dimensions can be subaggregated into one criterion or one dimension can be divided into several criteria (see below). To be consistent, a family of criteria must be subject to the following three conditions:

- Exhaustivity condition:

$$g_j(a) = g_j(a'), j = 1, \dots, n \rightarrow a \text{ and } a' \text{ are indifferent.}$$

- Nonincomparability condition:

If  $g_j(a) = g_j(a') \forall j \neq k$  and  $g_k(a') > g_k(a)$  then  $a'$  is preferred to  $a$ , or  $a'$  and  $a$  are indifferent. The indifferent case might occur when  $g_k$  is a semicriterion\* - that is to say, when there exists<sup>k</sup> an indifference threshold  $\Delta(x)$  such that:

---

\*Roy (1975) presents also the concepts of precriterion and pseudocriterion.

$g_k(a) = x < g_k(a') \leq g_k(a) + \Delta(x) \rightarrow a$  and  $a'$  are indifferent ,

$g_k(a) = x \leq g_k(a) + \Delta(x) < g_k(a') \rightarrow a'$  is preferred to  $a$

$x + \Delta(x) \geq y + \Delta(y)$  if  $x \geq y^*$  .

If  $\Delta(x) = 0 \forall x$ , the indifferent case cannot occur and the criterion is a true criterion.

- Nonredundancy condition:

There exist some pairs of actions ( $a, a'$ ) real or fictitious for which one of the two preceding conditions becomes false if we drop any one of the  $g_j$  of the family.

To assess such a family of criteria, the analyst has to accomplish from the raw material whether the valuations are single point or nonsingle point:

- If the valuation is single point and if  $E_i$  is expressed in real numbers, the analyst can often build easily and directly criteria from the valuations:

$g_i(a) = \gamma_i(a)$  if dimension  $i$  is a "benefit" ,

$g_i(a) = -\gamma_i(a)$  if dimension  $i$  is a "cost" ,

$g_i(a) = f [0_i, \gamma_i(a)]$  if the preference order is not compatible with the order of the grades on  $E_i$ ,  $0_i \in E_i$  is then the most preferred grade on the scale.

- If the valuation is distributional, the analyst might find a single point equivalent to the distribution on the dimension  $i$ . The general form for making such a reduction is given by:

$$g_i(a) = \sum_{e \in \gamma_i(a)} \delta_i^a(e) u_i(e)$$

where  $u_i(e)$  is a utility or value assigned to each grade of the scale  $E_i$ . All the techniques, such as computing

---

\*The preference relation defined by  $g_1$  is then a semi-order.

a mean value, a present value (discounting techniques), an expected value (utility theory), an accessibility index, a monetary value in cost-benefit analysis, are of this form.

Let us drop subscript  $i$  and comment on these techniques:

- *Mean value* (weighted value)

$\delta^a(e)$  is the weight assigned to  $e \in E$  (number or percent of people, of experts, etc.). In some cases we have  $u(e) = e$ , and then of course  $E$  is a numerical scale and  $g(a)$  is estimated with the same units (i.e. a mean travel time). In other cases,  $u(e)$  is a numerical transformation of the scale  $E$ , necessary when  $E$  is ordinal with letters for each grade (see an example in the next subsection for the dimension agreement with the planning policy). The other techniques are special cases of weighted values.

- *Utility theory* (see Raiffa, 1968; von Neumann and Morgenstern, 1967)

$\delta^a(e)$  is the probability of obtaining a certain consequence  $e$  when choosing  $a$ , and  $u(e)$  is the utility function assigned to  $e \in E$ .  $g(a)$  is then the expected utility of  $a$ .

- *Cost-benefit analysis* (see Frost, 1971; Beesley, 1973; Georgi, 1973; Lesourne, 1972; Pearce and Dasgupta, 1973).

$\delta^a(e)$  is a quantity of consequence  $e$  (number of people concerned, etc.) and  $u(e)$  is the monetary value assigned to  $e$  (value of time, cost of people killed, cost of noise, etc.).

- *Preferences over time (discounting)* (see Lancaster, 1963; Koopmanns, 1960)

If  $\delta^a(t)$  is the monetary value (eventually a utility, see Meyer, forthcoming) to be accounted for year  $t$ , then the present value  $g(a) = \sum_t \delta^a(t) \frac{1}{(1+i)^t}$  is of the general form, where  $u(t) = \frac{1}{(1+i)^t}$  is the utility of one monetary unit for year  $t$ ,  $i$  being the discount rate.

- *Accessibility theory* (see Koenig, 1974)

Accessibility indices provide an interesting example of weighted value in the field of transportation studies. For instance, the accessibility to employment for a transportation alternative  $a$ , is given by  $g(a) = \sum_t \delta^a(t) \exp(-\alpha t)$ , where  $\delta^a(t)$  is the number of employments which can

be reached within a travel time  $t$ ,  $\alpha$  is a given number and  $\exp(-\alpha t)$  is the utility assigned to one employment which can be reached within travel time  $t$ .

If the analyst assumes the axioms (mostly independence axioms) to be verified, and if  $\delta^a(e)$  is not too difficult to estimate, he might nevertheless encounter real difficulties in assessing a correct  $u(e)$ . Everyone knows the problems of choosing a correct discount rate, for instance, and has seen cases in which a slight change in the discount rate could inverse the rank-order of two alternatives. This situation might occur very easily when applying these techniques and we shall see an example of such an inversion on the noise dimension in the next subsection. Jacquet-Lagreze (1975 b) has proposed an assessment technique of a fuzzy preference relation which enables the analyst to point out such situations in which inversion of rank-order could easily happen.

Another means of assessing criteria from distributions is to explode one dimension  $i$  into several criteria, by using the information of  $\gamma_i(a)$  and  $\delta_i(a)$  directly. We shall see an example of this in the next subsection for the quality of time dimension.

- Another issue is to aggregate directly some dimensions or the whole set of dimensions into a value function or utility function. This is the issue involved in decision analysis when applying multi-attribute utility theory (see Raiffa, 1968; Keeney, 1974; Von Winterfeld and Fischer, 1975).

#### Application of the General Concepts to an Airport Location Problem

Among the consequences to be examined in step 1 of such a problem are:

- individual ones: travel time, quality of travel
- collective ones: noise, employment generated
- structural ones: agreement with the planning policy.

To take into account the uncertainty of the future, we suppose here that it can be sufficiently described by two scenarios or events:

- scenario  $\epsilon = 1$ : an important highway is built between the city and the airport;
- scenario  $\epsilon = 2$ : the highway is not built, but ordinary roads exist between airport site and city.

Of course, we assume here that it is not the same authority that decides the location of the airport and the possibility of building a special highway for the airport (the investment of the highway could result from a decision of a bank or a private firm). If it were the same authority, then the highway would enter in the definition of the potential actions (here alternatives) with for instance:  $a_1$ , site a with the highway;  $a_2$ , site a without the highway;  $a_3$ , site b with a highway, etc.

#### Travel time

Scale  $E_1$ : t in minutes.

State indicator  $\gamma_1(a)$ : A traffic forecasting model allows estimation of the travel time. Necessarily the output of such a model is a distribution  $\delta^1(a)$  of all the different travel times and  $\delta_1^a(t)$  is the number of users that can reach the airport site a within travel time t. Now as we have assumed two different scenarios depending whether the highway can be built, we need to take into account these scenarios to work out a traffic forecasting model because the highway has an impact on travel time. So in fact we are in the complex indexed event case and we have two distributional evaluations  $[\gamma_1(a), \delta_1^a(a)]^\epsilon$  with  $\epsilon = 1, 2$ .

If we do not know any probability distributions of the two scenarios, we can build a criterion which will take two different values in each. For instance, we might compute a mean travel time. But if there are thresholds in the travel time perception by users (e.g. users do not really consider as a significant benefit a saving of 2 or 3 minutes for a journey which takes 15 to 60 minutes) then the criterion is a semicriterion and not a true criterion.

#### Quality of travel time

Scale  $E_1$ : a qualitative scale defined by:

- $e_0$ : standing position for which no activity is possible,
- $e_1$ : sitting position for which some activity such as reading an easy newspaper or having an ordinary conversation is possible, and
- $e_2$ : comfortable sitting position in a quiet environment allowing difficult reading, professional conversation, writing, etc.

State indicator  $\gamma_i(a) = \{e_0, e_1, e_2\}$  will be in general the set of grades (non-single valuation). A modulation indicator  $\delta_i(a)$  can be an estimation of the percentage of users in each of the three situations  $e_0, e_1, e_2$ :  $\delta_i(a) = \{\delta_i^a(e_0), \delta_i^a(e_1), \delta_i^a(e_2)\}$ .

To assess criteria the analyst can explode the dimension quality of time into two criteria:

$g_{i1}(a) = - \delta_i^a(e_0)$  : percentage of users travelling in bad conditions ,

$g_{i2}(a) = \delta_i^a(e_2)$  : percentage of users travelling in good conditions .

Noise

In the study "An airport for Yorkshire" (see Metra Consulting Group, 1972), the scale  $E_i$  is a noise number index (NNI), the state indicator is an estimation of the households lying inside NNI contours, this estimation being made more precise by a modulation indicator. For two of the feasible sites these distributions were those of Table 3.

Table 3. Evaluation of two airport sites by the dimension noise.

A \ E	50 NNI	45-50 NNI	35-45 NNI	Total
(a) Balne Moor	46	293	700	1039
(b) Winterset	2	5	1521	1528

In our notation, the valuation of Balne Moor is

$$\gamma_i(a) = \{> 50 \text{ NNI} , 45-50 , 35-45\} ,$$

$$\delta_i(a) = \{46, 293, 700\} .$$

As this study was a cost-benefit analysis, a criterion was assessed by means of the general form given in the previous section. The costs  $u_i(e)$  adopted in the study were in pounds sterling:

$$u_i (> 50 \text{ NNI}) = - \text{£}2562$$

$$u_i (45-50) = - \text{£}2075$$

$$u_i (35-45) = - \text{£}1595 \quad (\text{note that } u_i (< 35) = 0) \quad .$$

Thus the contribution to the general cost made by "noise" was, for (a) and (b)  $g_i(a) = - 1,842,000$  and  $g_i(b) = - 2,441,000$ , since for example:

$$g_i(a) = \sum_{e \in E} \delta_i^a(e) u_i(e) = 46 \times 2562 + 293 \times 2075 + 700 \times 1595 \quad ,$$

$$g_i(a) > g_i(b) \quad .$$

As noted before, this aggregation may be very sensitive to other monetary values. For instance, if we choose the set of values:

$$u_i (> 50 \text{ NNI}) = - 4000$$

$$u_i (45-50) = - 3000$$

$$u_i (35-45) = - 1000$$

$$u_i (< 35 \text{ NNI}) = 0,$$

then we obtain the new costs

$g_i(a) = - 1,763,000$  and  $g_i(b) = - 1,544,000$  which inverse the rank order:  $g_i(a) < g_i(b)$  .

#### Employment generated

Scale  $E_i$ : N, number employed.

State indicator  $\gamma_i(a)$ . After observation of the employment generated by other airports already built, the number employed was

estimated by the relation  $N = 300 + 520 T$ , where  $T$  is the level of air-traffic for each year measured in millions of passengers.  $T$  is estimated by a model of traffic forecasting with a precision of  $\pm 40\%$ . If we choose the number of employees generated for the year 1990, this procedure defines the state indicator:

(a) Balne Moor:  $\gamma_i(a) = [1320, 2680]$  ,

(b) Wintersett:  $\gamma_i(b) = [1400, 2880]$  .

As we do not know any probability distribution on these intervals, we are in the case where the modulation indicator could be relational. From these data, the analyst can build for instance a preference relation (here  $b \succ_i a$ ) eventually a fuzzy preference relation (see Jacquet-Lagrez 1975 b).

#### Agreement with planning policy

Scale  $E_i$  is a qualitative scale defined by:

- $e_0$ : severe disagreement between the probable consequences of the site on urban development and planning policy,
- $e_1$ : impact on urban development not very important; the site is neither in agreement, nor in disagreement with planning policy,
- $e_2$ : very good agreement; the probable consequences of the urban development induced by an airport at this site are in good agreement with planning policy.

State indicator  $\gamma_i(a)$ . A group of experts are asked to evaluate the different alternatives on this scale. Here, either the analyst explodes the dimension into several criteria (maybe as many criteria as there are experts) and obtains a single point value for each criterion, or he might consider the distribution of the values in the dimension (especially if there is good agreement between the experts). Then the modulation indicator  $\delta_i(a)$  would be the distribution of the experts on the three grades  $e_0, e_1, e_2$ . A criterion could then be assigned a quantitative score  $u(e)$  for each grade.



### GLOBAL PREFERENCE AND OPERATIONAL ATTITUDE

This is concerned with the problem of aggregating the consistent family of criteria  $g_1(a), \dots, g_j(a), \dots, g_n(a)$  into the global preference. It is the most subjective part of the modeling since the global preference considered at this stage is that of a well identified decision-maker.

Roy (1971, 1974) considers four fundamental mutually exclusive situations when comparing two potential actions  $a$  and  $a'$ :

Indifference: Two actions are indifferent in the sense that there exist clear and positive reasons to choose equivalents.

Example:  $g_j(a) = g_j(a')$  for all  $j$  .

Strict preference: One of the two actions is strictly preferred to the other.

Example:  $g_j(a) = g_j(a')$  for all  $j \neq k$  and  $g_k(a') - g_k(a)$  is a significant difference.

Large preference: One of two actions is not strictly preferred to the other but it is impossible to say if the other is strictly preferred to or indifferent from the first one because neither of the two former situations dominates.

Example:  $g_j(a) = g_j(a')$  for all  $j \neq k$  and  $g_k(a') - g_k(a)$  is neither small enough to justify indifference, nor large enough to justify strict preference.

Incomparability: The two actions are incomparable in the sense that none of the three former situations dominates.

Example:  $g_j(a) > g_j(a')$  for  $j = 1, \dots, p$  and  $g_j(a') > g_j(a)$  for  $j = p + 1, \dots, n$  , most of the differences being significant.

How can one discriminate between these four different situations?

#### First Operational Attitude: Aggregate the $n$ Criteria Into One Criterion

To assess the global preference, the analyst often adopts the axiom of complete transitive comparability:

- (a) indifference situations define on  $A$  a binary relation  $\sim$ , symmetric and transitive,
- (b) preference situations define on  $A$  a binary relation  $\succ$ , anti-symmetric and transitive,

(c) large preference and incomparability situations are excluded.

Let us call  $\succeq$  the complete relation defined by:

$$a \succeq a' \iff a > a' \text{ or } a \sim a' .$$

If the three conditions of the axiom are satisfied, then  $\succeq$  defines a weak order on A (and vice versa). In real-world problems it is always possible (see Fishburn, 1970) to characterize such a weak order by means of a function

$$V_A(a) = V_A[g_1(a), \dots, g_n(a)] \in \mathbb{R}$$

such that

$$V_A(a) = V_A(a') \iff a \sim a'$$

$$V_A(a) > V_A(a') \iff a > a'$$

Such a function appears as a true criterion aggregating the n criteria of the family.

To make this function explicit is an operational attitude which has proved its effectiveness, particularly when the analyst adopts the problem formulation  $\alpha$  on a globalized and fixed set A. As we have mentioned in the first section, the main methodologies in transportation problems are cost-benefit analyses, and multi-attribute utility theory.

Second Operational Attitude: Accept Incomparability and Aggregate the Family of Criteria in an Outranking Relation

For a lot of decision problems, Roy (1971, 1975) has outlined the operational advantage of allowing incomparability and eventually large preference situations as the response to the fundamental question about global preference modelling.

An outranking relation is a binary relation  $S_A$  defined on A such that:

- (a)  $a S_A a'$ , and  $a' S_A a$ : the analyst estimates that for the decision-maker a and a' are indifferent,
- (b)  $a S_A a'$ , not  $a' S_A a$ : the analyst estimates that for the decision-maker a is preferred to a' (strict or large preference),

- (c) not a  $S_A a'$ , not a  $a' S_A$ : the analyst cannot, does not want to, or does not know how to choose between the two preceding cases.

Various techniques have been proposed for assessing such outranking relations, based either on ordinal or cardinal properties of the criteria (see Bertier and Roy, 1972; Jacquet-Lagrezze, 1975; Moscarola and Roy, 1976). Very often, for the same problem, a family of nested outranking relations are assessed, and the weaker the outranking relation is, the more risky it is in its preference assessment. The analyst's conviction to opt for an outranking relation may in certain cases need to be formalized through a degree of credibility in the outranking which leads to the concept of fuzzy outranking (see Roy, 1976; Jacquet-Lagrezze, 1975 a). For some problems, it is sufficient to assess such an outranking relation and to use it directly as a decision aid as in Betolaud and Fevrier (1973) or in Bertier et al. (1972) where different highway locations (A86-87) in the western area of Paris were compared. But most often, especially when the number of potential actions is large, these relations are an intermediary and can be used in different manners, depending on the problem formulation which is chosen. For instance, in order to build a trichotomy in the problem of selecting road investment projects, special procedures have been proposed (see Moscarola and Roy, 1976).

Third Operational Attitude: Elaborate in an Interactive Way One or More Compromises Based on Local Preferences

In this third attitude, the global preference may remain widely implicit, and is characterized by the use of three types of mechanisms in an appropriate iterative sequence which lead to the interactive elaboration of one or more compromises.

- (a) Research mechanism: the analyst exploits the data gathered as a result of a previous reaction (cf. (c)) in order to make headway in the elaboration of compromises. It is concerned with:
- the analysis and comparison of new data with old,
  - research (taking into account the results of the analysis of compromise projects and/or of certain of their characteristics).
- (b) Reinitialization mechanism: this creates, with due regard for the results of (a), new conditions under which the next reaction must be performed. It concerns:
- translation, into a language comprehensible by the decision-maker or his representative, of

these results or situations (real or fictitious) they have to understand,

- obtaining the necessary conditions (understanding certain results, thinking about the antagonism between some criteria, ...) so that the regathered data during the reaction will be as significant as possible.

(c) Reaction mechanism: this gathering of information on the local preferences by the decision-maker may take extremely different forms:

- discussion as unbiased as possible,
- discussion based on a prepared questionnaire,
- voting procedure, etc.

When improvement is no longer necessary, or when it becomes impossible, we have reached a compromise, which is, more or less clearly according to the procedure adopted, a local optimum relative to an implicit criterion. Diverse methodologies have been proposed: see Aubin and Naslund (1972), Benayoun et al. (1971), Geoffrion et al. (1971), Vincke (1975), Roy (1975).

#### REFERENCES

- Aubin, J.P., and B. Naslund (1972), *An Exterior Branching Algorithm*, working paper 72-42, European Institute in Management, Brussels.
- Beesley, M.E. (1973), *Urban Transport: Studies in Economic Policy*, Butterworth, London.
- Benayoun, R., et al. (1971), Linear Programming with Multiple Objective Functions: STEP Method (STEM), *Mathematical Programming*, 1, 3.
- Bertier, P., and B. Roy (1972), La méthode ELECTRE II: une application au media-planning, VIIème Conférence Internationale de Recherche Opérationnelle, Dublin.
- Bertier, P., et al. (1972), *Choix de tracés autoroutiers en milieu suburbain: faut-il vraiment endommager des sites urbains et/ou forestiers? Lesquels?*, Paper presented at the Journées de Formation Transports et Agglomérations, Nice, October.
- Betolaud, Y., and R. Fevrier (1973), Conservation des forêts suburbaines et passage des autoroutes, L'exemple du tracé de l'autoroute A 86-A 87, *Revue Forestière Française*, May-June.
- Fishburn, P.C. (1970), *Utility Theory for Decision-Making*, Wiley, New York.

Frost, M.J. (1971), *Values for Money: the Techniques of Cost Benefit Analysis*, Gower Press, London.

Geoffrion, A.M., et al. (1971), *An Interactive Approach for Multi-criterion Optimization, with an Application to the Operation of an Academic Department*, working paper no. 16, Western Management Science Institute, University of California, Los Angeles.

Georgi, H. (1973), *Cost-Benefit Analysis and Public Investment in Transport: a Survey*, Butterworth, London.

Godard, X. (1973), *Methodologie de l'analyse multicritère appliquée aux transports urbains*, Rapport de recherche no. 9, Institut de Recherche des Transports, Arcueil, France.

Jacquet-Lagrezé, E. (1975), *How We Can Use the Notion of Semi-orders to Build Outranking Relations in Multicriteria Decision Making*, in *Utility, Probability and Human Decision Making*, Reidel, Boston, Mass.

Jacquet-Lagrezé, E. (1975 a), *La modélisation des préférences-Préordres, quasi-ordres et relations floues*, thèse, Université de Paris V, Paris.

Jacquet-Lagrezé, E. (1975 b), *Modelling Preferences Among Distributions Using Fuzzy Relations*, paper presented to the 5th Research Conference on Subjective Probability, Utility and Decision Making, Darmstadt.

Keeney, R.L. (1974), *Multiplicative Utility Functions*, *Operations Research*, 22, 1.

Koenig, G. (1974), *Théorie économique de l'accessibilité urbaine*, *Revue Économique*, no. 2.

Koopmanns, T.C. (1960), *Stationary Ordinal Utility and Impatience*, *Econometrica*, 38, 2.

Lancaster, K. (1963), *An Axiomatic Theory of Consumer Time Preference*, *International Economic Review*, 4, 2.

Lesourne, J. (1972), *Le calcul économique: théorie et applications*, Dunod, Paris.

Metra Consulting Group (1972), *An Airport for Yorkshire*, M.C.G., London.

Meyer, R.F., *State-Dependent Time Preference*, in D. Bell, R. Keeney, and H. Raiffa, eds., *Conflicting Objectives in Decisions*, Wiley, New York (forthcoming).

Moscarola, J., and B. Roy (1976), *Procédure automatique d'examen fondée sur un surclassement trichotomique en présence de critères multiples*, in *Rairo*, série verte.

- de Neufville, R., and R.L. Keeney (1972), *Use of Decision Analysis in Airport Development for Mexico City, Analysis of Public Systems*, MIT Press, Cambridge, Mass.
- von Neumann, J., and O. Morgenstern (1967), *Theory of Games and Economic Behavior*, Wiley, New York.
- Pearce, D.W., and A.K. Dasgupta (1973), *Cost Benefit Analysis - Theory and Practice*, MacMillan, New York.
- Raiffa, H. (1968), *Decision Analysis - Introduction Lectures on Choice under Uncertainty*, Addison-Wesley, Reading, Mass.
- Roy, B. (1971), Problems and Methods with Multiple Objective Functions, *Mathematical Programming*, 1, 2.
- Roy, B. (1973), How Outranking Relation Helps Multiple Criteria Decision-Making, *Selected Proceedings of a Seminar on Multi-Criteria Decision Making*, The University of South Carolina Press, Columbia.
- Roy, B. (1974), Critères multiples et modélisation des préférences (l'apport des relations de surclassement), *Revue d'Économie Politique*, no. 1.
- Roy, B. (1975), A Conceptual Framework for a Normative Theory of "Decision-Aid", paper presented to the XXII International Meeting of the Institute of Management Science, 24-26 July, to be published in *Management Sci.*, 1977.
- Roy, B. (1976), Outranking Partial Preference Analysis and Decision Aid: the Fuzzy Relation Concept, in D. Bell, R. Keeney, and H. Raiffa, eds., *Conflicting Objectives in Decisions*, Wiley, New York (forthcoming).
- Roy, B., et al. (1975), Elaboration de critères permettant une intégration des divers aspects liés au temps dans l'aide à la décision en matière de transports, Research Report 7400020002257501 D.G.R.S.T. - S.A.E.I., ATP Socio-Economie des Transports, Paris.
- Vincke, Ph. (1975), *Une méthode interactive en programmation linéaire à plusieurs fonctions*, Communication au Congrès EURO I, Bruxelles.
- Von Winterfeldt, D., and G.W. Fischer (1975), Multi-Attribute Utility Theory: Models and Assessment Procedures, *Utility, Probability and Human Decision Making*, Reidel, Boston, Mass.

The Application of System Dynamics Simulation  
Methodology for Analyzing the Social Impact  
Of a High-Speed Mass Transportation System

T. Hasegawa, Y. Ogawa, and M. Watanabe

INTRODUCTION

As the high-speed mass transportation system between major Japanese cities developed, so those cities' developments have been affected. Thus, the construction of the Tokaido Shinkansen of the Japanese National Railway between Tokyo and Osaka has had a destructive impact, not only on the economy of Osaka, but also on the culture of the Osaka district, despite the fact that the Osaka area has been the cultural and economic center of Japan until just recently for more than 10 centuries and that people there are still taking advantage of their long history.

The city of Osaka has more than 2.5 million inhabitants now and still leads Japan in various activities. Although a big city, it has suffered destructive impacts, and the smaller cities along the high-speed mass transportation system have obviously been suffering severer impacts. This report deals with a simulation of the dynamic behavior of populations in an extremely large city and a medium sized city that is about 100 km from the larger city and supposedly the local center of its region, by the technology of system dynamics. The model is constructed in order to analyze how the population structure in the two cities will be affected by the completion of a high-speed mass transportation system connecting the two cities. The development of a larger city is due to more complicated factors and the impacts of the transportation system under consideration may not be so large.

Selecting Tokyo as the larger city and Utsunomiya as the smaller (now the local center of the Tochigi Prefecture) the authors have developed a system dynamics model which runs on an IBM 370-158 computer system with DYNAMO II. The model reproduces the history of the population behavior of the two cities fairly well from 1955 to 1975. As is well known, there are many problems concerning the validity of the system dynamics model. The authors think that it is essential that the model reproduce the history of the population structure, too. The results of the simulation runs are not definite yet; however, they have shown that the completion of the Tohoku Shinkansen, under construction to connect Tokyo and Sendai via Utsunomiya, may well take the wealth and brains of Utsunomiya to Tokyo. So the Tohoku Shinkansen will accelerate the centralization of Japan in Tokyo and disturb the local cities. The data used in this report are mainly from census reports [1] and the Regional Economy Statistics [2].

### SYSTEM DYNAMICS MODEL FOR TWO CITIES

In this report, 23 wards of the Tokyo Prefecture are selected as the main city and Utsunomiya is selected as the sub-city because the Tohoku Shinkansen, which is a super express railway connecting Tokyo and Sendai via Utsunomiya, is now under construction and because Utsunomiya has been the local center of the Tochigi Prefecture.

In order to analyze the characteristics of the cities, the inhabitants of the main city have been classified into three groups:

- Top population: top executives and managerial;
- Labor population: all workers except merchants and employees in service business; and
- Merchant population: merchants and employees in service business.

While for the sub-city, an additional group is added:

- Agricultural population.

In addition, populations at night and in the daytime are considered for the main city. For the sub-city, only the population at night is considered since there is not much difference between night and day population there.

For the identification of the direction of population movement in both cities, Figure 1 shows, with the use of abbreviated terms for the variables in the system dynamics model, the direction of movement from the sub-city to the main city and Figure 2 from the main city to the sub-city. As a measure of the activities in each city, the number of factories is used.

The model is divided into 12 sectors, the top, labor, merchant, main city top population in daytime (TWP), main city labor population in daytime (LWP), main city merchant population in daytime (MWP), sub-city agriculture population (APS) and factory. The top, labor, merchant and factory sectors are subdivided into two sectors for main city and sub-city respectively.

As there are many similar points in these sectors, the top sector is explained in fair detail and others are explained by emphasizing the differences from the top sectors and from each other.

#### The Top Sector

The top sector is a model for the top population at night. In this section, a subsector for top population of the main city



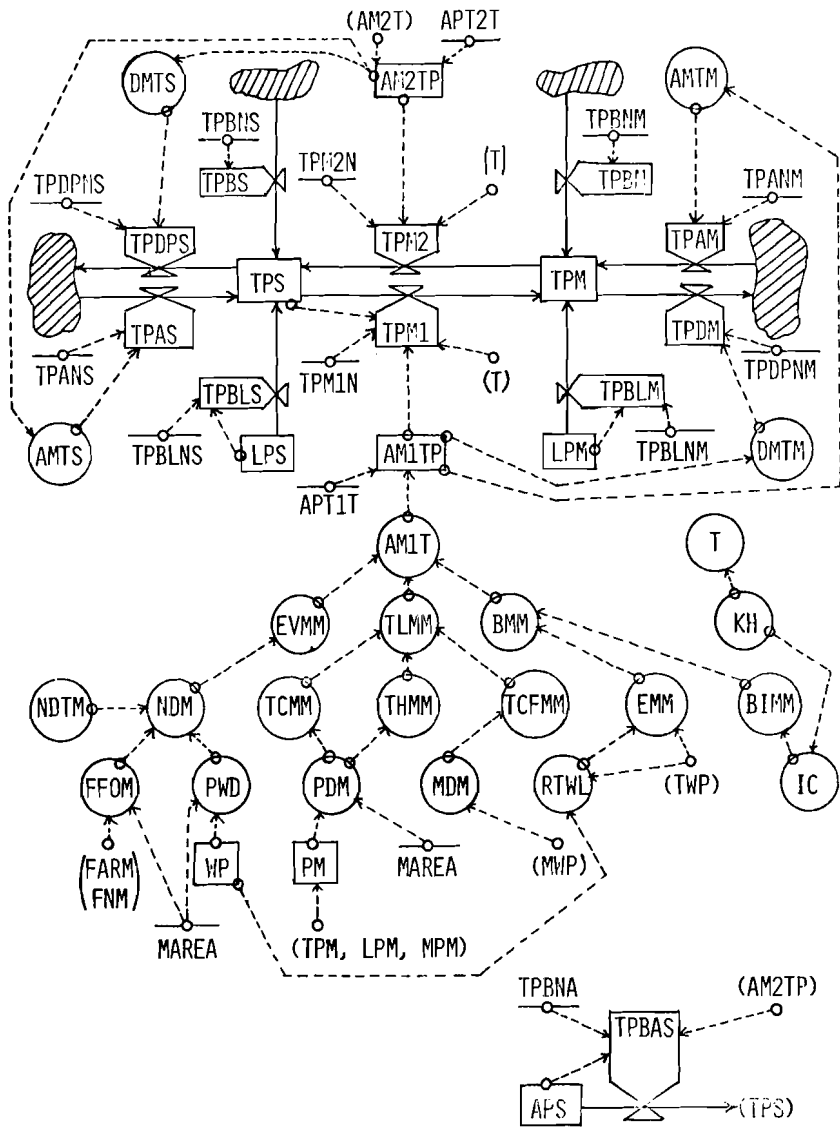


Figure 1. Flow diagram for the top sector.

at night is explained, i.e. TPM. Top population in the sub-city, TPS, is in contrast to TPM; the TPM sector is explained here.

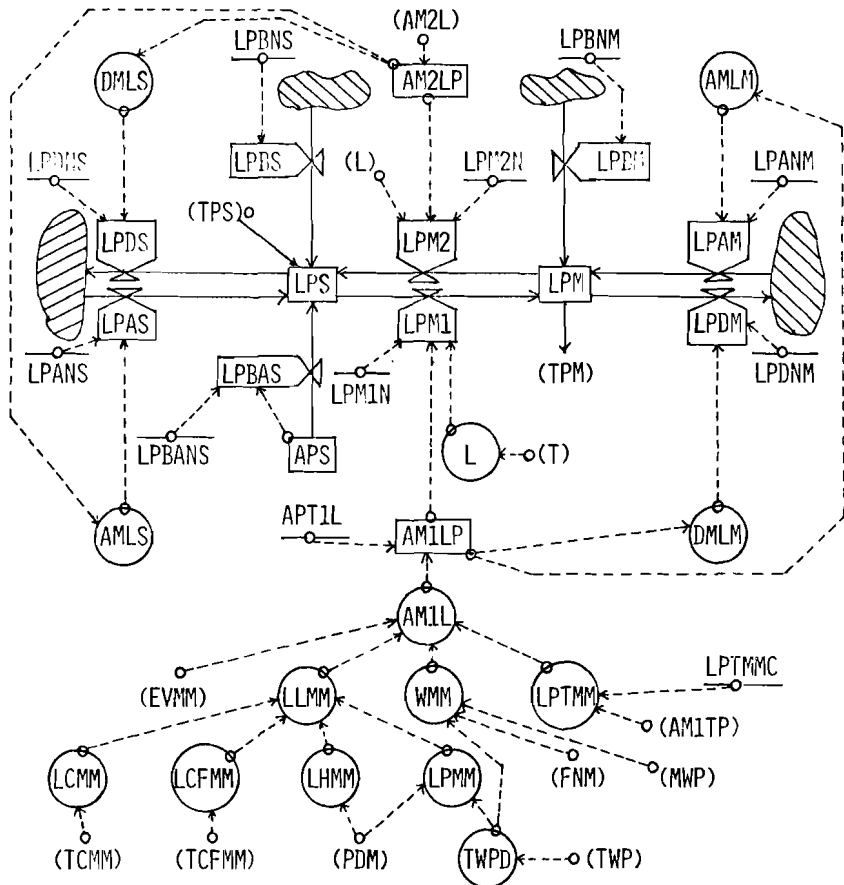


Figure 2. Flow diagram for the labor sector.

First, TPM is given by

$$L \quad TPM.K = TPM.J + (DT) (TPM1.JK + TPAM.JK + TPBLM.JK + TPBM.JK$$

$$X \quad - TPM2.JK - TPDM.JK) \quad ,$$

where TPAM and TPDM are rate variables for arrival and departure, respectively, between this system and the environment, and TPBM and TPBLM are also rate variables for natural growth of TPM and arrival from LPM to TPM, respectively. The dimension for the above rate variables is number of persons/year.

TPBM is estimated by the growth rate of the labor population in Japan shown by the census reports of the Japanese Government [1].

The social growth is supposed to be determined by the relative attractiveness of each city, i.e. people move from one place to another according to the relative attractiveness perceived by considering various merits and demerits of the places. The rate variables that determine the social growth are TPAM, TPDM, TPM1, and TPM2. TPM1 and TPM2 are rate variables which are derived by the relative attractiveness of the main city and sub-city. These four rate variables are given as:

$$R \quad TPAM.KL = (TPANM) (TPM.K) (AMTM.K)$$

$$R \quad TPDM.KL = (TPDNM) (TPM.K) (DMTM.K)$$

$$R \quad TPM1.KL = (TPM1N) (AM1TP.K) (TPS.K) (T.K)$$

$$R \quad TPM2.KL = (TPM2N) (AM2TP.K) (TPM.K) (T.K) \quad ,$$

where  $T.K = 1$ . Further detail of  $T.K$  is given in the impact sector.  $AMTM$  is a multiplier for arrival to the main city from the outside environment and  $DMTM$  is a multiplier for departure from the main city to the outside environment.  $AM1TP$  and  $AM2TP$  are attractiveness perceived by the top population moving from sub-city to main city and main city to sub-city, respectively.  $AMTM$  and  $DMTM$  are determined by the TABLE functions of  $AM1TP$ :

$$A \quad AMTM.K = TABLE(AMT, AM1TP.K, 0, 2.6, .2)$$

$$A \quad DMTM.K = TABLE(DMT, AM1TP.K, 0, 2.6, .2) \quad ,$$

where  $AMT$  and  $DMT$  show the values of TABLE functions. For more detail, readers may obtain copies of the source program from the author.  $TPANM$ ,  $TPDNM$ ,  $TPM1N$  and  $TPM2N$  are constants which stand for the multiplier in the normal states of the cities.  $AM1TP$  is determined by attractiveness multiplier  $AM1T$  with delay time  $APT1T$ :

$$L \quad AM1TP.K = AM1TP.J + (DT/AP1T.J) (AM1T.J - AM1TP.J).$$

$AM1T$  is the attractiveness multiplier for moving from sub-city to main city which is supposed to be determined by various factors, among which the following three are chosen in this model.

- Natural environment, clean air and water (EVMM);
- Living conditions for top population (TLMM); and
- Business activities (BMM).

Then,  $AM1T$  is given by:

$$A \quad AM1T.K = (EVMM.K) (TLMM.K) (BMM.K).$$

As  $AM1T$  concerns the movement from sub-city to main city, these factors relate to the main city. The natural environment

multiplier of the main city, EVMM is derived by a TABLE function of the natural environmental destruction multiplier, NDM, by:

$$A \quad EVMM.K = TABLE(EVMT, NDM.K, 0, 2.25, .25).$$

DRM is determined by a weighted sum of two TABLE functions as:

$$A \quad DRM.K = (TABLE(NDT1, PWD.K, 0, 11E3, 1E3) + (2)(TABLE(NDT2, FFOM.K,$$

$$X \quad 0, .1, .01)))/(3) \quad ,$$

where PWD is the population density of the main city in daytime and FFOM is the fraction of factory areas in the main city.

The living condition multiplier of the top population in the main city, TLMM, is derived by a product of commutation multiplier of top population in the main city, TCMM, clothes and food multiplier of the top population in the main city, TCFMM, and housing multiplier of top population in the main city, THMM:

$$A \quad TLMM.K = (TCMM.K)(TCFMM.K)(THMM.K) \quad ,$$

where TCMM, TCFMM and THMM are derived as follows:

$$A \quad TCMM.K = TABLE(TCMT, PDM.K, 0, 16000, 1000)$$

$$A \quad TCFMM.K = TABLE(TCFMT, MDM.K, 0, .8, .05)$$

$$A \quad THMM.K = TABLE(THMT, PDM.K, 0, 1E4, 1E3).$$

The business multiplier of the main city, BMM, is given by a product of the employment multiplier of the main city, EMM, and the business information multiplier of the main city, BIMM:

$$A \quad BMM.K = (EMM.K)(BIMM.K) \quad ,$$

where EMM and BIMM are derived as:

$$A \quad EMM.K = TABLE(EMT, RTWL.K, 0, .2, .02)$$

$$A \quad BIMM.K = TABLE(BIMT, TWP.K, 0, 1E6, 2E5) + IC.K.$$

The variable IC is a factor characterizing the main city as the political and economic center of Japan.

TPM2, AM2TP are derived similarly for the sub-city with one difference: that is, in the sub-city, there exists a migration from the agriculture population to the top population of the sub-city:

$$R \quad TPBAS.KL = (APS.K)(TPBNA)(AM2TP.K).$$

The flow diagram of the top population sector is shown in Figure 1.

Labor Sector

The flow diagram of the labor sector at night is shown in Figure 2. Similarly to the explanation of the top sector, the labor population of the main city at night, LPM, is explained mainly. LPM is given as:

$$L \quad LPM.K = LPM.J + (DT) (LPM1.JK + LPAM.JK + LPBM.JK - LPM2.JK - TPBLM.JK)$$

$$X \quad - LPDM.JK).$$

LPM1 is mainly determined by the attractiveness of the main city for the labor population in the sub-city, AM1L, which is given by the four factors:

- EVMM;
- Living conditions for the labor population in the main city, LLMM;
- Working conditions for the labor population in the main city, WMM; and
- Prospects for the labor population in the main city, LPTMM.

LLMM is a function of the four multipliers, LPMM, LCMM, LHMM and LCFMM, where LPMM is the multiplier of public service, LCMM is the multiplier of commutation, LHMM is the multiplier of housing condition, and LCFMM is the multiplier of clothes and foods, all for the labor population in the main city.

$$A \quad LLMM.K = (LPMM.K) (LCMM.K) (LHMM.K) (LCFMM.K)$$

$$A \quad LPMM.K = (TABLE(LPMT1, PDM.K, 0, 1E4, 1E3) + (4) (TABLE(LPMT2, TWPD.K,$$

$$X \quad 0, 2000, 200)))/(5)$$

$$A \quad LHMM.K = TABLE(LHMT, PDM.K, 0, 1E4, 1E3)$$

$$A \quad LCMM.K = TCMM.K$$

$$A \quad LCFMM.K = TCFMM.K .$$

WMM is given by the density of the top population in daytime, TWPD, the number of factories, FNM, and the merchant population in daytime, all in the main city:

$$A \quad WMM.K = (TABLE(WMT, TWPD.K, 0, 1200, 200) + TABLE(WMT1, FNM.K, 0, 1E5,$$

$$X \quad 1E4) + TABLE(WMT2, MWP.K, 0, 16E5, 2E5))/(3) .$$

LPTMM is given by

$$A \quad LPTMM.K = (LPTMMC) (AM1TP.K) \quad ,$$

where LPTMMC is a constant. Then, AM1L is given by:

$$A \quad AM1L.K = (EVMM.K) (LLMM.K) (WMM.K) + LPTMM.K \quad .$$

As shown in Figure 2, the labor population is explained similarly to LPM except the point that LPS has an inflow from APS.

### Merchant Sector

The flow diagram of the merchant sector at night is shown in Figure 3.

The attractiveness of the main city for the merchant population in the sub-city, AM1M plays an important role in the merchant sector. The attractiveness for the merchant population in the main city is mainly determined by the four factors:

- EVMM;
- The living conditions for merchants in the main city, MLMM;
- Commercial conditions for the merchant population in the main city, COMM; and
- Prospects for the merchant population in the main city, MPTMM.

MLMM is given by a product of four multipliers for the merchant population at night in the main city:

$$A \quad MLMM.K = (MPMM.K) (MCMM.K) (MHMM.K) (MCFMM.K) \quad ,$$

where MPMM is the multiplier of public service, MCMM is the multiplier of commutations, MHMM is the multiplier of housing, and MCFMM is the multiplier of clothes and foods.

$$A \quad MPMM.K = (TABLE(MPMT1, PDM.K, 0, 1E4, 1E3) + (4)(TABLE(MPMT2, TWPD.K,$$

$$X \quad 0, 2000, 200)))/(5)$$

$$A \quad MCMM.K = TABLE(MCMT, PDM.K, 0, 16000, 1000)$$

$$A \quad MHMM.K = TABLE(MHMT, PDM.K, 0, 1E4, 1E3)$$

$$A \quad MCFMM.K = TABLE(MCFMT, MDM.K, 0, .8, .05) \quad .$$

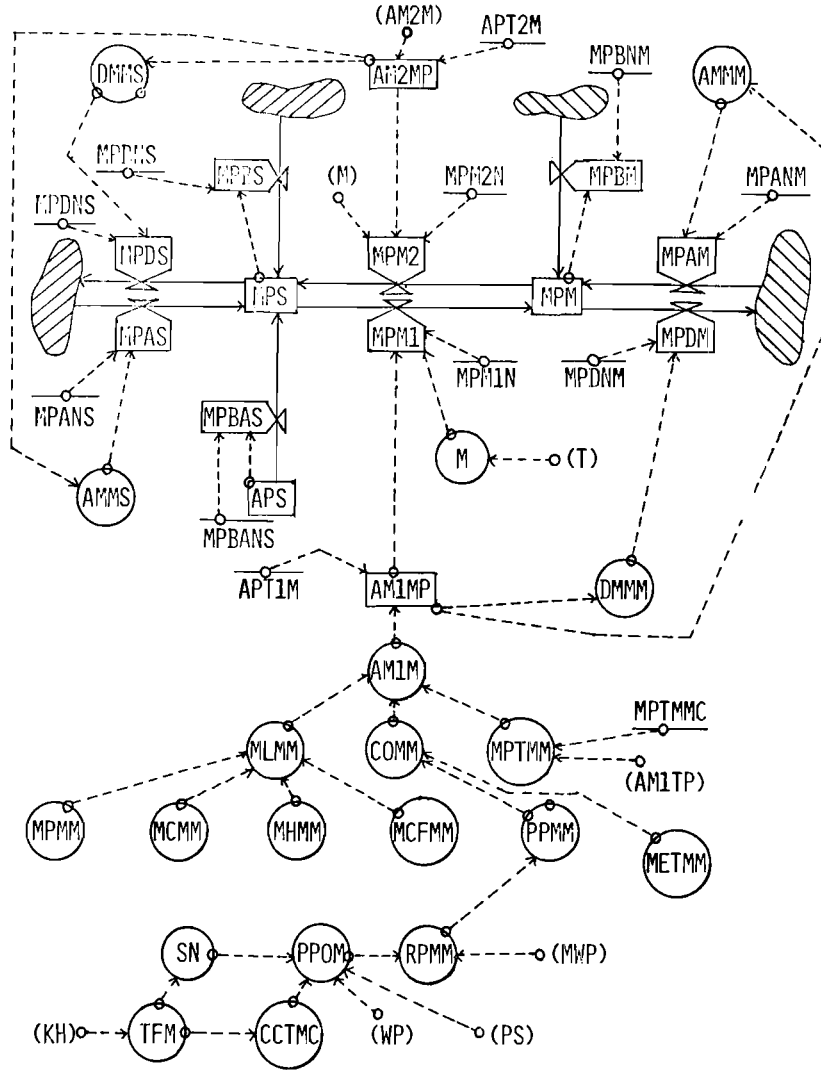


Figure 3. Flow diagram for the merchant sector.

COMM is given by the three multipliers for merchant population at night in the main city:

$$A \quad \text{COMM.K} = (\text{PPMM.K}) (\text{METMM.K}) + \text{MIC*IC.K} \quad ,$$

where PPMM is the multiplier of purchasing in the main city, METMM is the multiplier for the merchant employment in the main city, and MIC\*IC is the multiplier for the accessibility of information concerning commercial activities. PPMM is given by the ratio, RPMM, of the purchasing population, PPOM to the population in daytime, MWP:

$$A \quad \text{PPMM.K} = \text{TABLE}(\text{PPMT}, \text{RPMM.K}, 0, 16, 2).$$

PPOM is given by

$$A \quad \text{PPOM.K} = \text{WP.K} + (\text{CCTMC.K}) (\text{PS.K}) (\text{SN.K}) \quad ,$$

where WP is the population of the main city in daytime, PS is the population of the sub-city, and SN is the number of satellite towns of Tokyo. SN and CCTMC will be explained later in the impact sector. METMM is derived from the ratio, WMD, of the merchant population in daytime to the sum, WP, of the populations of top, merchant and labor in the daytime.

$$A \quad \text{METMM.K} = \text{TABLE}(\text{METMT}, \text{WMD.K}, 0, .5, .05).$$

MPTMM is given by

$$A \quad \text{MPTMM.K} = (\text{MPTMMC}) (\text{AM1TP.K}).$$

Then, AM1M is given by

$$A \quad \text{AM1M.K} = (\text{EVMM.K}) (\text{MLMM.K}) (\text{COMM.K}) + \text{MPTMM.K}.$$

The merchant population in the sub-city is given similarly as shown in Figure 3.

#### Top Population Sector of the Main City in Daytime

The top population of the main city in daytime, TWP, is given by

$$L \quad \text{TWP.K} = \text{TWP.J} + (\text{DT}) (\text{TWA.JK} - \text{TWD.JK} + \text{TWB.JK} + \text{TWBL.JK} + (\text{TPM1.JK}$$

$$X \quad \quad \quad - \text{TPM2.JK}) (\text{TWP.J}) / (\text{TPM.J})$$

$$R \quad \text{TWA.KL} = (\text{TWAN}) (\text{TWP.K}) (\text{AMTWP.K})$$

$$R \quad \text{TWD.KL} = (\text{TWDN}) (\text{TWP.K}) (\text{DMTW.K})$$

$$L \quad \text{AMTWP.K} = \text{AMTWP.J} + (\text{DT}/\text{APTWP.J}) (\text{AMTW.J} - \text{AMTWP.J}).$$



The attractiveness multiplier for the top population in daytime,  $AMTW$ , is given by

- Convenience of commutation for the top population in daytime ( $TWCM$ ), and
- $BMM$ ,

where

$$A \quad TWCM.K = TABLE(TCMT, PWD.K, 0, 16000, 1000).$$

Then,  $AMTW.K$  is given by

$$A \quad AMTW.K = (TWCM.K) (BMM.K).$$

The flow diagram of the TWP sector is shown in Figure 4.

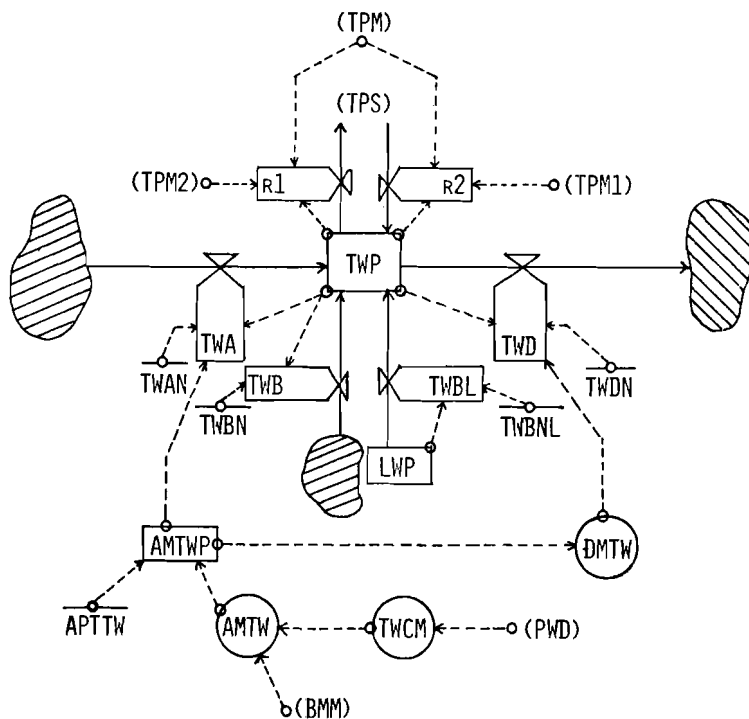


Figure 4. Flow diagram for the top in daytime (working) sector.  
(R1 and R2 are undefined rate variables.)

Labor and Merchant Populations in Daytime

These sectors are explained in Figures 5 and 6.

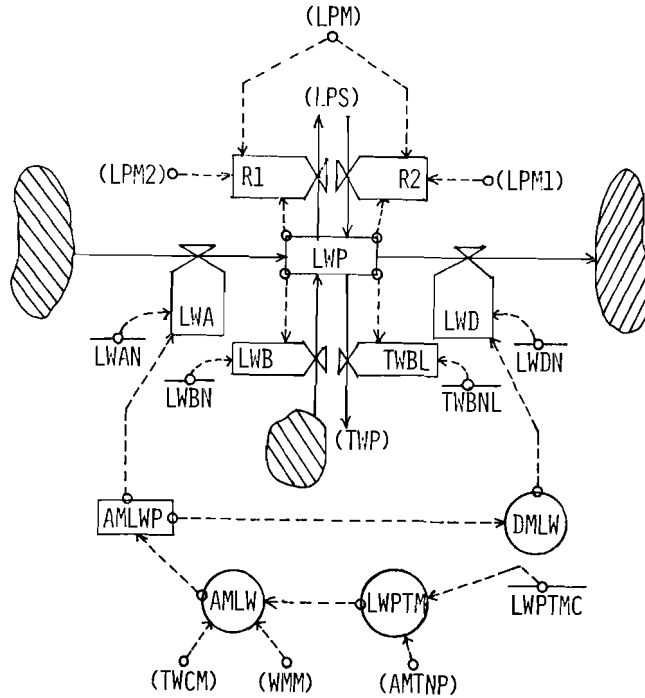


Figure 5. Flow diagram for the labor in daytime (working) sector. (R1 and R2 are undefined rate variables.)

Agriculture Population Sector

The agriculture population, APS, in the sub-city is explained in Figure 7 and in the APS sector of the source program. APS is given by

$$L \quad APS.K = \text{MAX}(APS.J + (DT)(APBS.JK - TPBAS.JK - LPBAS.JK), 0)$$

where APBS is the natural growth of the agriculture population and TPBAS, LPBAS, and MPBAS are the out flow rate to top, labor and merchant populations, respectively.

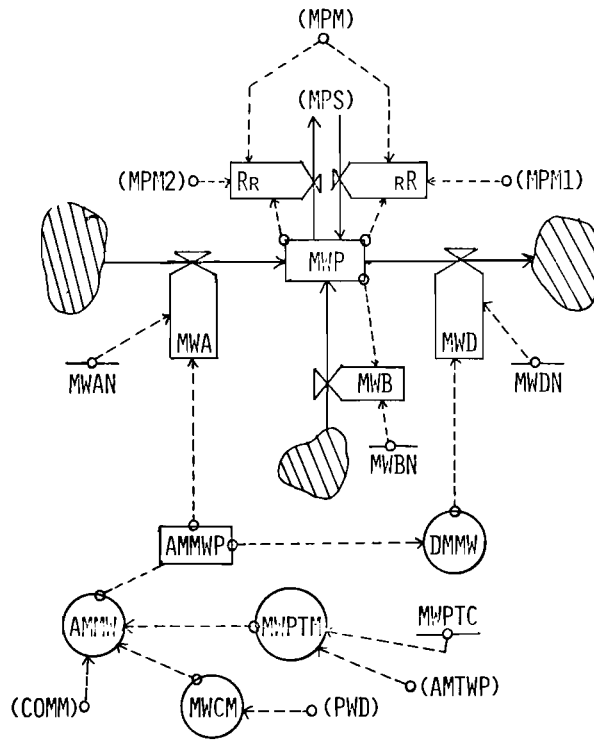


Figure 6. Flow diagram for merchant in daytime (working) sector.  
(RR and RR are undefined rate variables.)

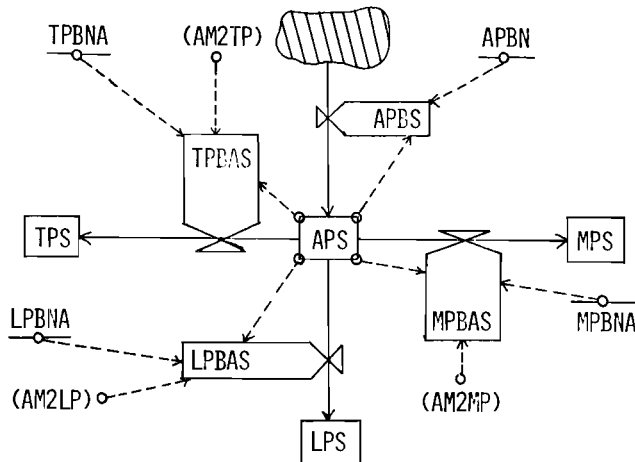


Figure 7. Flow diagram for the agriculture population sector.

The city area of the sub-city, SAREA expands according to the decrease of APS and is given as

$$A \quad SAREA.K = TABLE(SART, APS.K, 0, 3E4, 1E4)$$

$$T \quad SART = 300/250/180/100.$$

### Factory Sector

The number of factories in the main city, FNM is given by

$$L \quad FNM.K = FNM.J + (DT) (FCM.JK - FDM.JK) ,$$

where FCM is the rate of construction of factories and FDM is the rate of demolition. FCM is given by

$$R \quad FCM.KL = (LCRM.K) (FCDM.K) ,$$

where LCRM is the construction rate of factories given by the ratio RLJM of the number of laborers in daytime to the number of employees in the factories in the main city.

$$A \quad LCRM.K = TABLE(LCRT, RLJM.K, 0, 4, .5)$$

FCDM is given by

$$A \quad FCDM.K = (FNM.K) (FMM.K) (FCNM).$$

FMM is the multiplier for the factories and FCNM is a constant. FMM is derived by

- Land fraction occupied by the factories, FLMM, and

- Job multiplier at the factories, FWMM, as:

$$A \quad FMM.K = (FLMM.K) (FWMM.K).$$

FLMM is given by

$$A \quad FLMM.K = TABLE(FLMT, FFOM.K, 0, .1, .01) ,$$

and FWMM is

$$A \quad FWMM.K = TABLE(FWMT, RLJPM.K, 0, 4, .5).$$

RLJPM is

$$L \quad RLJPM.K = RLJPM.J + (DT/RLJPTM) (RLJM.J - RLJPM.J) ,$$

while FDM is given by

$$R \quad FDM.KL = (FNM.K) (FDMPM.K) (FDNM) ,$$

where  $FDNM$  is a constant. The factory demolition multiplier perceived,  $FDMPM$  is given by the factory demolition multiplier with time delay,  $FDMPPT$ .

$$L \quad FDMPM.K = FDMPM.J + (DT/FDMPPTM) (FDMM.J - FDMPM.J).$$

$FDMM$  is given by

$$A \quad FDMM.K = TABLE(FDMMT, FMM.K, 0, 2.6, .2).$$

For the factories in the sub-city,  $FNS$ , the similar model is constructed as shown in Figure 8.

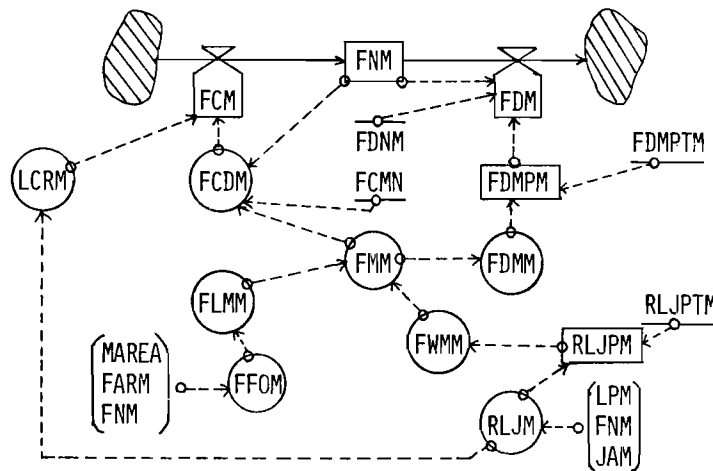


Figure 8. Flow diagram for the factory number in main city sector.

Impact Sector

In this model, the impact of a high-speed train is given by the maximum train speed improvement from 100 km/h to 180 km/h. That is,

$$T \quad KHT = 60/70/80/85/90/95/100$$

RUN BASERUN

$$T \quad KHT = 60/70/80/85/90/95/180$$

RUN IMPACTRUN.

KH is supposed to affect T, L, M, IC, ICS and TFM.

A T.K = TABLE(TTT, KH.K, 60, 180, 20)

A L.K = T.K

A M.K = T.K .

TFM is the mobility or transfer multiplier given by

A TFM.K = TABLE(TFMT, KH.K, 60, 180, 20).

TFM affects SN and CCTMC. SN is the number of satellite towns of Tokyo and CCTMC is the availability multiplier of business information.

A SN.K = 10 + TABLE(SNT, TFM.K, 0, 2, .4)

A CCTMC.K = TABLE(CTMT, TFM.K, 0, 2, .4)

#### RESULTS OF SIMULATION RUNS

Simulation runs start from 1955. As all the data are derived from the census reports of the Japanese government, they are available for every five years from 1955 to 1970. Figures 9 to 18 show the results of simulation runs. In Figures 9, 11, 13, 15 and 17, solid lines for the first fifteen simulation years indicate the corresponding actual data.

The model proposed here simulates the actual behavior of the populations in both cities fairly well as is shown by the relatively small discrepancies between the actual data and the results of simulation runs. Figures 10, 12, 14, 16 and 18 show the results of impact runs, where the impact of completion of the Tohoku Shinkansen is given at the 30th simulation year, i.e. 1985. The maximum speed of the railway connecting both cities will be 180 km/h.

Figures 9 to 18 show the values of the following level variables.

Figures 9 and 10: The total population of the main city at night, M.

The total population of the main city in daytime, W.

The number of employed in the sub-city in daytime, S.

Figures 11 and 12: The top population of the main city at night, T.

The labor population of the main city at night, L.

The merchant population of the main city at night, M.

Figures 13 and 14: The top population of the main city in daytime, T.

The labor population of the main city in daytime, L.

The merchant population of the main city in daytime, M.

Figures 15 and 16: The top population of the sub-city at night, T.

The labor population of the sub-city at night, L.

The merchant population of the sub-city at night, M.

The agriculture population of the sub-city at night, A.

Figures 17 and 18: The number of factories in the main city, M.

The number of factories in the sub-city, S.

In these figures, the abscissa scaling T stands for  $10^3$ , and the simulation years go downwards.

If the Tohoku Shinkansen is not constructed, the base run of this simulation forecasts the future behavior of the populations as shown in Figures 9, 11, 13, 15 and 17: the population of the main city at night reaches its peak in 1968 and the population in daytime reaches its peak in 1970. For the future, the population in daytime will reach a stable state at about 5 million in about 1995 and the population at night will reach a stable state at about 4 million also about 1995. The population of the sub-city will keep increasing with a decreasing rate of increase.

As for the structure of the population, the top population of the main city at night will keep increasing slowly, the labor population reaches its peak in about 1967 and will reach a stable state in about 1995 and the merchant population behaves similarly to the labor population, except in its magnitude. The top

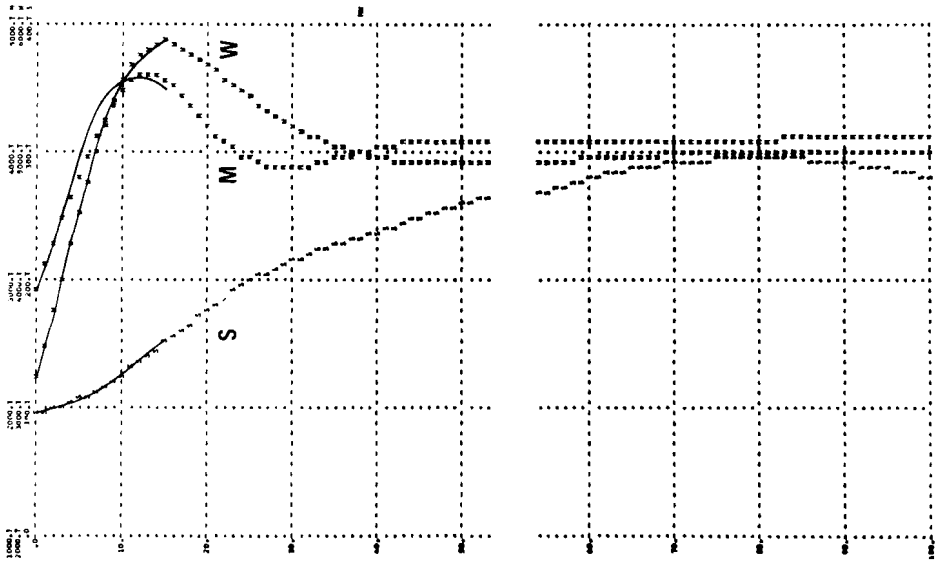


Figure 9. Total population of the main city at night and in daytime and the number of employed in the sub-city, without impact.

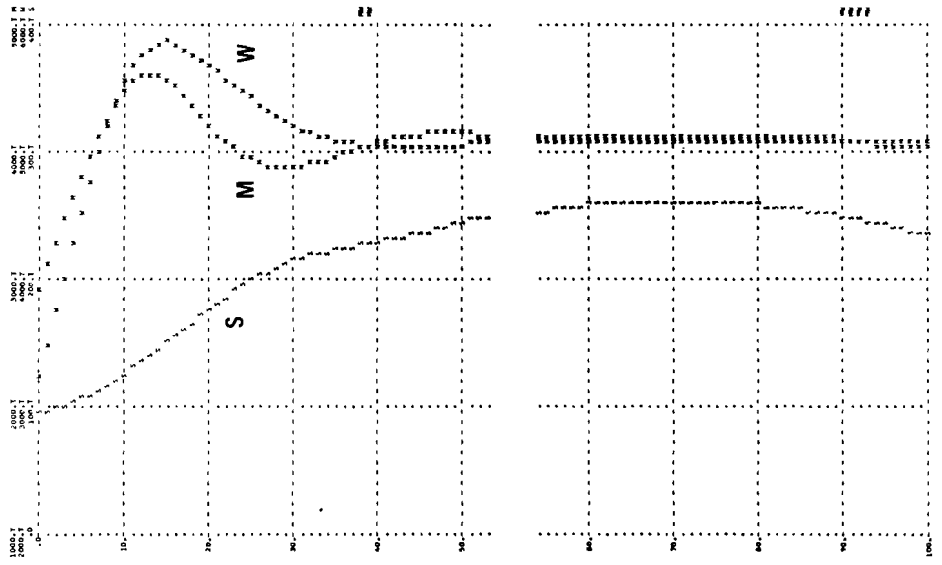


Figure 10. Total population of the main city at night and in the daytime and the number of employed in the sub-city, with impact.



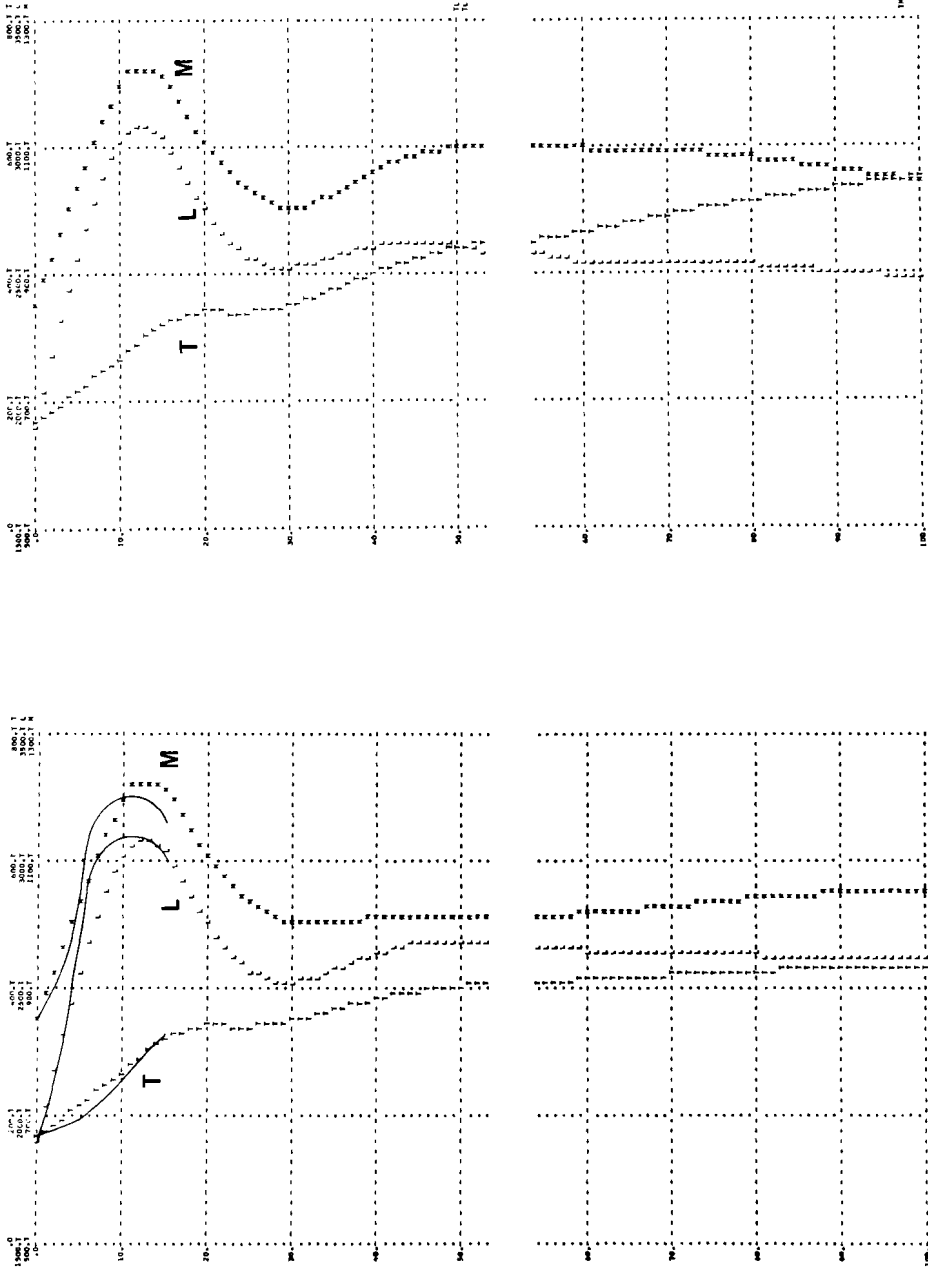


Figure 11. Population structure of the main city at night, without impact. Figure 12. Population structure of the main city at night, with impact.

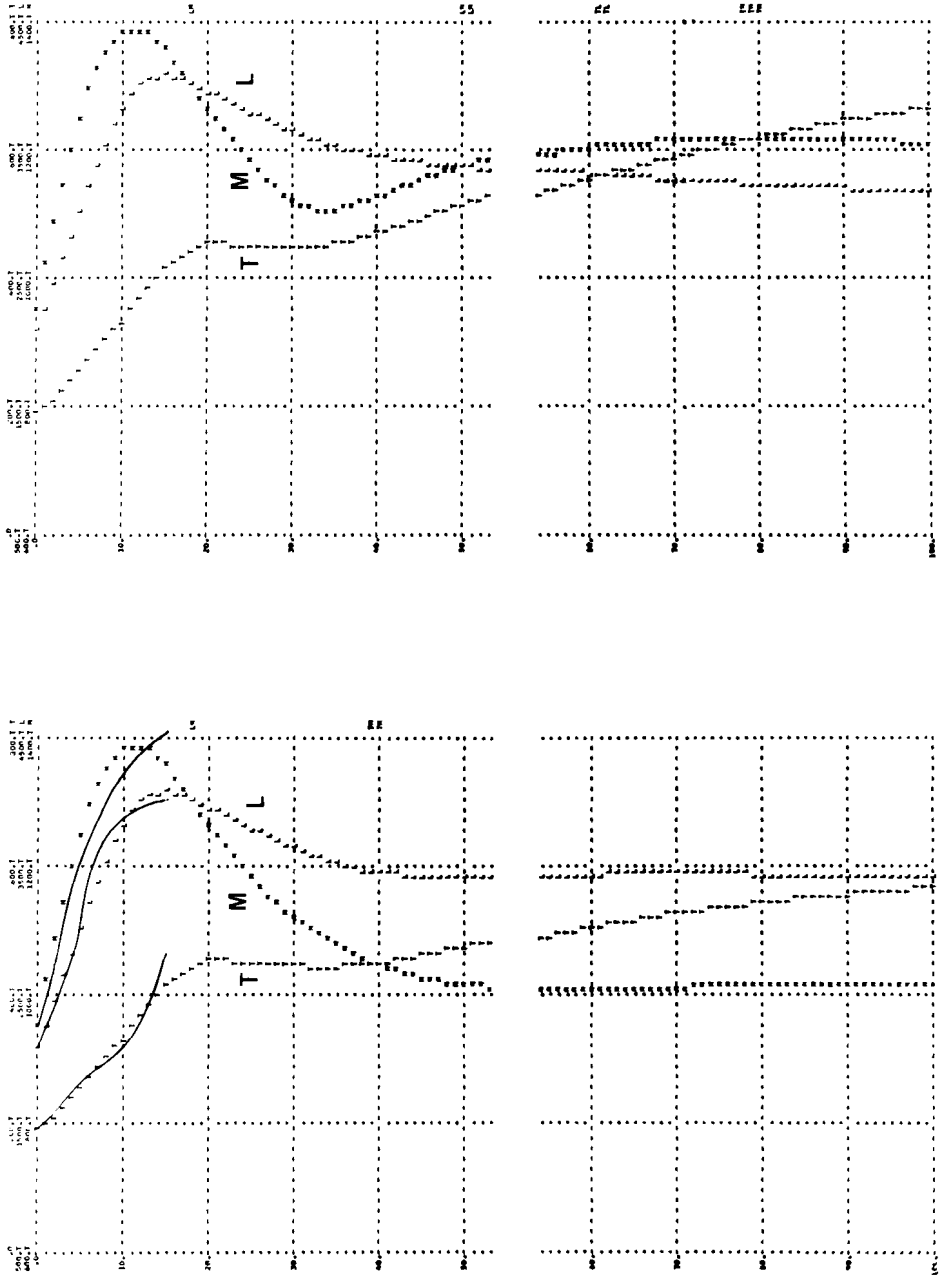


Figure 13. Population structure of main city in daytime, without impact. Figure 14. Population structure of main city in daytime, with impact.

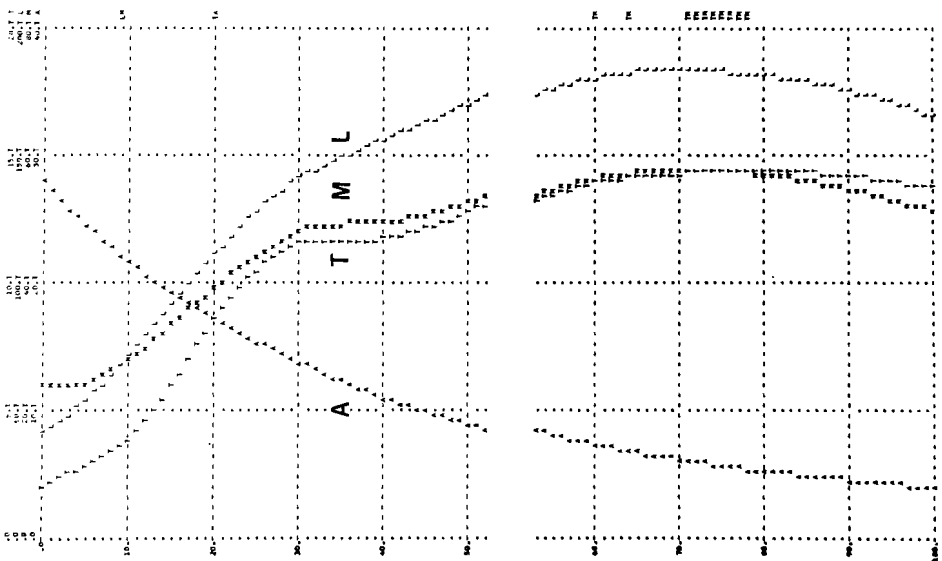


Figure 15. Population structure of sub-city at night, without impact.

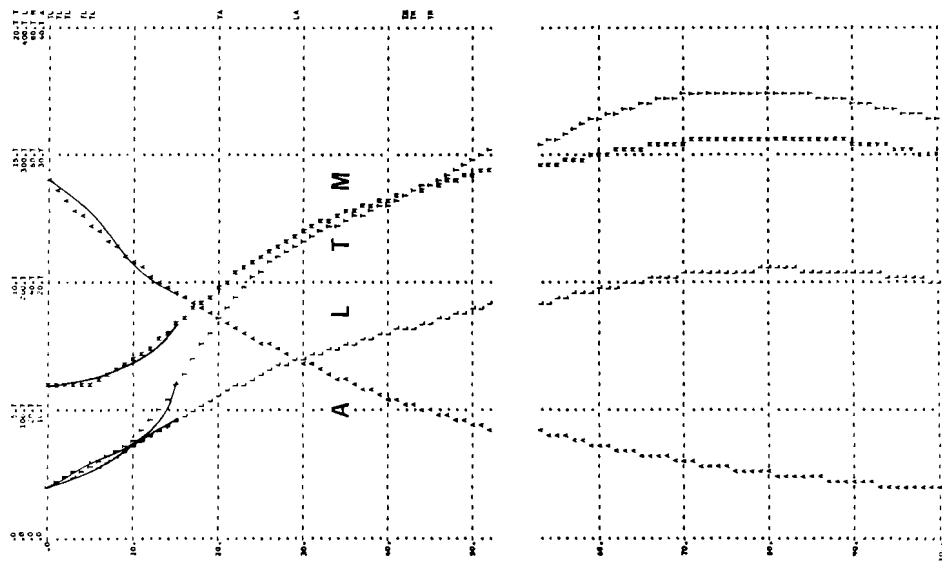


Figure 16. Population structure of sub-city at night, with impact.

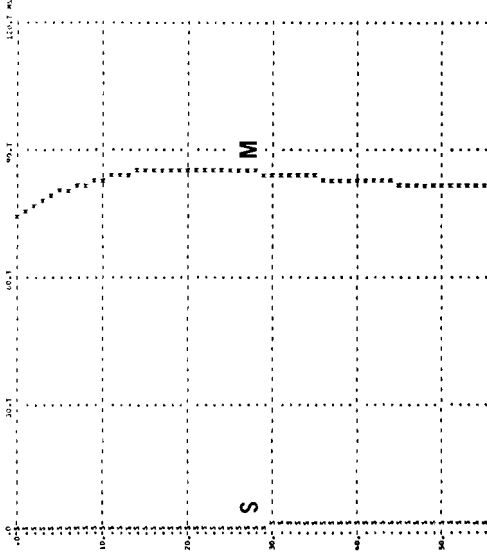


Figure 18. Number of factories in both cities, with impact.

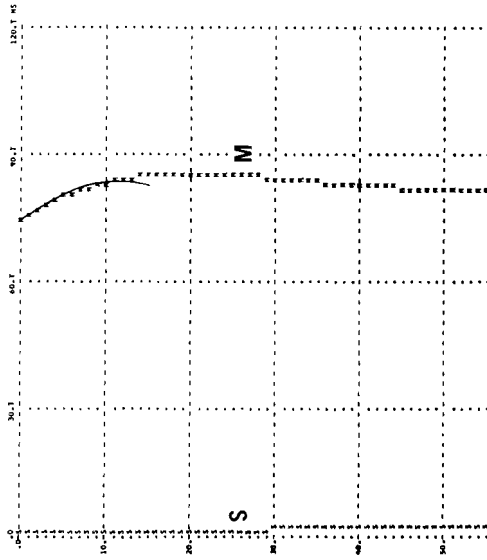


Figure 17. Number of factories in both cities, without impact.

population of the main city in daytime will keep increasing slowly. The labor population in daytime reaches its peak in 1970 and will reach a stable state at about 3.4 million in about 1995. While the merchant population in daytime reaches its peak in 1966 and approaches the merchant population at night.

For the sub-city, the populations will keep increasing with decreasing rates except the agriculture population. The agriculture population decreases with a decreasing rate.

If the Tohoku Shinkansen starts operation with a scheduled speed of 180 km/h between Tokyo and Utsunomiya in 1975, the behavior of the population in both cities is simulated by the impact run and shown in Figures 10, 12, 14, 16 and 18. There are many differences compared with the results of the base run of the simulation: namely, the population of the main city at night increases; the population in daytime also increases more than at night; in 1995, there will be 150,000 persons more than without the Tohoku Shinkansen, while in the sub-city, the increase rate of its population becomes about a half in the period from 1985 to 1995. As for the population structure, the top population of the main city at night increases with the rate twice that of the result of the base run, and the labor population decreases. The merchant population increases and reaches a stable state of 1.2 million. For the sub-city, the increase rate of the number of employed decreases fairly well and reaches a stable state. The top population stagnates from 1985 to 1995 and then increases with a rate smaller than that of the base run. The labor population increases with a rate smaller than the base run and reaches a stable state.

#### CONCLUSION

The results of the simulation runs show the validity of the simulation model fairly well as a system dynamics model. Comparing the results of the base run and impact run, we may say that the sub-city will be affected by the completion of the Tohoku Shinkansen. For instance, the development of the managerial functions in the sub-city and the growth of the tertiary industries will be suppressed. While in the main city, the concentration of the managerial function will be accelerated. This implies that the wealth, knowledge and information of Japan will concentrate more and more in Tokyo while the local cities will lose their own special features to Tokyo.

The model reported here is not a definite one yet, but, it does give us various ideas concerning the impact of a high-speed mass transportation system on a local city. There are still many problems besides the refinements of the model. For instance, some local cities other than Utsunomiya need to be dealt with. What is more important is that we should construct a method to perform an assessment of a high-speed railway from social,

economic and technological points of view, and if a high-speed mass transportation system turns out to be indispensable, we should find a set of policies to minimize the detrimental effects of this transportation system to the local cities.

The authors believe that the expected excess centralization of Japan to Tokyo is harmful for the future of Japan.

The authors would like to express their sincere appreciation to Mr. Hideo Mitsuhashi of Kyoto University (at present Nippon Electric Co., Tokyo, Japan) for his contributions to the coding of the model and his valuable comments on this report.

#### REFERENCES

- [1] Bureau of Statistics, Japanese Government, *Census reports, 1955, 1960, 1965 and 1970, and partial report for 1975*, Japanese Government.
- [2] Toyo Keizai Press, *Regional Economy Statistics, 1974*, Toyo Keizai Press, Tokyo, 1974.

Appendix

Definitions of terms not given in the text

AM1(2)LP	Attractiveness for Migration multiplier of 1(2) type Labor Perceived
AM1(2)MP	Attractiveness for Migration multiplier of 1(2) type Merchant Perceived
AMLM(S)	Arrival Multiplier of Labor to M(S)
AMLW	Arrival Multiplier of Labor Work
AMLWP	Arrival Multiplier of Labor Work Perceived
AMMM(S)	Arrival Multiplier of Merchant to M(S)
AMMW	Arrival Multiplier of Merchant Work
AMMWP	Arrival Multiplier of Merchant Work Perceived
APBN	Agriculture Population Birth Normal
APT1(2)L	Attractiveness Perception Time of 1(2) type Labor
APT1(2)M	Attractiveness Perception Time of 1(2) type Merchant
APTLW	Attractiveness Perception Time of Labor Work
APTMW	Attractiveness Perception Time of Merchant Work
DMLM(S)	Departure Multiplier of Labor from M(S)
DMLW	Departure Multiplier of Labor Work
DMMM(S)	Departure Multiplier of Merchant from M(S)
DMMW	Departure Multiplier of Merchant Work
FARM(S)	Factory AREA of M(S)
FPSMM(S)	Factory Public Service Multiplier of M(S)
ICMT	Information Constant of M Table
ICST	Information Constant of S Table
JAM(S)	Job Average of M(S)
LPANM(S)	Labor Population Arrival Normal of M(S)
LPBNM(S)	Labor Population Birth Normal from Agriculture
LPM1N	Labor Population Migration of 1 type Normal
LWA	Labor Work Arrival
LWAN	Labor Work Arrival Normal
LWB	Labor Work Birth
LWBN	Labor Work Birth Normal

LWDN	Labor Work Departure Normal
LWP	Labor Work Population
LWPTM	Labor Work ProspeCT Multiplier
LWPTMC	Labor Work ProspeCT Multiplier Constant
MAREA	M AREA
MPAM(S)	Merchant Population Arrival of M(S)
MPANM(S)	Merchant Population Arrival Normal of M(S)
MPBAS	Merchant Population Birth from Agriculture of S
MPBM(S)	Merchant Population Birth of M(S)
MPBNM(S)	Merchant Population Birth Normal of M(S)
MPDM(S)	Merchant Population Departure of M(S)
MPDNM(S)	Merchant Population Departure Normal of M(S)
MPM(S)	Merchant Population of M(S)
MPM1(2)	Merchant Population Migration of 1(2) type
MPM1(2)N	Merchant Population Migration of 1(2) type Normal
MWA	Merchant Work Arrival
MWAN	Merchant Work Arrival Normal
MWB	Merchant Work Birth
MWBN	Merchant Work Birth Normal
MWCM	Merchant Work Commutation Multiplier
MWD	Merchant Work Departure
MWDN	Merchant Work Departure Normal
MWPTM	Merchant Work ProspeCT Multiplier
MWPTMC	Merchant Work ProspeCT Multiplier Constant
NDTM(S)	Nature Destruction Multiplier of M(S)
PPMT	Purchasing Power Multiplier Table
RTLS	Ratio of Top to Labor of S
TPBNLM(S)	Top Population Birth Normal from Labor of M(S)
TPBNM	Top Population Birth Normal of M(S)
TWBN	Top Work Birth Normal
TWBNL	Top Work Birth Normal from Labor
WMPTM(S)	Work Multiplier Perception Time Multiplier of M(S)
M	Main city
S	Sub-city



## Evaluation of Public Transportation

O.H. Jensen

### INTRODUCTION

Planning public transportation in the larger Danish cities has for years been the subject of much work. It is generally accepted that, because of capacity and environmental factors, public transportation is absolutely necessary in these areas. However, in rural areas and in the smaller provincial Danish towns, public transportation has not been controlled and planned according to the very wide rights that by legislation are given to public authorities. However, today it is becoming more and more important for public authorities to intervene because of the still declining public transportation service.

This, combined with a demand for more public transportation at the expense of the private car, implies that today planning public transportation is one of the main subjects for many municipalities and counties in Denmark. However, as the role of public transportation planner and controller is new to them, they have no well established way of handling this subject. The lack of quantitative planning methods is very perceptible and it is the aim of this paper to fulfill part of the acute demand for research in this area.

### BACKGROUND

In terms of the number of vehicle kilometers, public transportation in Denmark plays a less significant role than before. For instance, if we consider 1960 and 1973 we have:

	1960	1973
Public transp. ( $10^6$ bus km)	174	186
Private cars ( $10^6$ car km)	7150	19800

While the number of car km increased 177%, the number of bus km increased only 7%.

The increase in the number of car passengers occurred especially in rural areas, where there is no capacity problem involved. It has not been possible to counterbalance increasing costs by rationalization of public transportation in sparsely populated

areas, as larger buses in this context are pointless. Furthermore, it has been shown (Pedersen, 1974) that there is no economy of scale in the operation of public transportation in Denmark, so that a concentration in fewer and bigger transportation companies is no solution to the cost problem.

In Denmark, public transportation has for decades been operated by a large number of small private companies and to some extent by the Danish State Railways. For various reasons this has been advantageous for as long as it was profitable to operate public transportation. But this implied that the extent of public control of public transportation was very small. Public transportation was an autonomous system, which for years worked fairly well. However, today, the financial basis of many bus services is declining and public authorities are forced to intervene, if they want to preserve a system of public transportation with a satisfactory level of service.

Formally, such an intervention has always been possible and in fact presupposed by legislation, but has not actually occurred. Thus, the role of controllers of public transportation is new to most Danish counties and municipalities, and the necessary quantitative methods for planning public transportation have not been sufficiently developed.

The present state of the transportation system gives the municipalities and counties very specific problems. For instance, nearly all sparsely populated municipalities run a separate school bus system, to which only pupils entitled to transportation have access. This raises the obvious question of whether the school bus system and the public transportation system may be substituted by a single public transportation system with only small changes in the use of resources but, it is hoped, with improvements in the service. A case study on this is given as an example at the end of this paper to illustrate one application of the model of accessibility.

Three case studies (Schlichtkrull and Ørum, 1974; Christensen and Friis, 1976; Juul Frederiksen, 1976) have been run at the institute (IMSOR) concerning four different municipalities in Denmark, and in one of these an interview survey was carried out; some of the results are shown in Figure 1 and Figure 2.

Figure 1 shows that the users of public transportation are mainly the young and old, and mainly women. Figure 2 shows that nearly all trips are home-based, and gives the other ends of trips.

#### A TOTAL MODEL OF ACCESSIBILITY

Transportation is not an end in itself, but a resource to satisfy some superior object. Therefore a model of public transportation necessarily must describe factors other than pure public transportation. One of the most fundamental things to

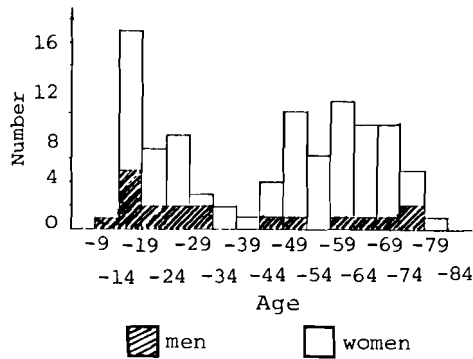


Figure 1. Distribution of sex and age of passengers using public transportation (pupils excluded) (94 interviews) in a rural area.  
Ref: Christensen and Friis, 1976.

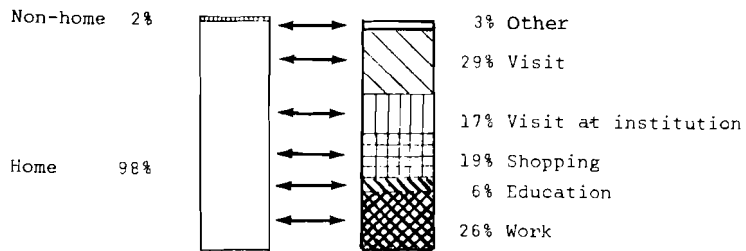


Figure 2. Distribution of trip purposes for passengers using public transportation in a rural area (94 interviews) (pupils excluded).  
Ref: Christensen and Friis, 1976.

transportation demand is the location of traffic generating facilities. Several surveys have indicated that access to private cars is another very important factor to include when describing the modal split between public and private transportation.

Therefore a total model of accessibility must contain a submodel of private car ownership, a submodel of the location pattern in the area and a submodel of public transportation. Furthermore, the model must include a submodel of the household and its composition, one of the reasons being that some trips are generated by the individual and some by the household in general. Besides it is the household, which eventually has the use of a car.

The model of accessibility therefore contains the following submodels:

- A submodel of the area
- A submodel of the household
- A submodel of private car ownership
- A submodel of public transportation.

If we do not include all four submodels in the model of accessibility, we shall not be able to obtain an overall optimum. Optimization assumes that interferences may be done just as well in one subsystem as another. Therefore one should not talk about planning public transportation, but rather about planning the total system in which public transportation is a subsystem. This does not impede any application of optimization within the subsystem of public transportation, but this must be done with due consideration of the total system.

THE SUBMODELS

The Submodel of the Area

We shall not give any detailed, mathematical description here, but only mention that the area is divided into zones, and that the model is based on network theory with the zones considered as nodes of the network.

The Submodel of the Household

Several alternative models describing the size and the composition of the household have been examined and a comparison based on the degree of fit between the model and observation has led to the following Poisson model.

$$Y_{ij\ell} = Y_{ij\ell}^+ + Y_{ij\ell}^-, \quad i=1,2 \quad j=1,2,3,4 \quad \ell=1,2\dots, \quad (1)$$

$$Y_{ij\ell}^+ \in \text{Bin}(n_{ij}, (\ell - p_{ij}^0) p_{ij\ell}^+) \quad , \quad " \quad , \quad (2)$$

$$Y_{ij\ell}^- \in \text{Bin}(n_{ij}, p_{ij}^0 p_{ij\ell}^-) \quad , \quad " \quad , \quad (3)$$

where  $i$  is the number of principal persons (holder of the apartment + evt. spouse) in the household,  
 $j$ , the age category of the head of the household  
 (4 categories: <25, 25-39, 40-64,  $\geq 65$ ),  
 $\ell$ , the number of residents in the household,

- $Y_{ij\ell}$ , the number of households in category (i,j) with  $\ell$  residents,
- $Y_{ij\ell}^+$ , the number of households with children in category (i,j) with  $\ell$  residents,
- $Y_{ij\ell}^-$ , the number of households without children in category (i,j) with  $\ell$  residents,
- $n_{ij}$ , the number of households in category (i,j),
- $P_{ij}$ , the fraction of households in category (i,j) without children in the household.

$$P_{ij\ell}^+ = \Pr\{X_{ij}^+ = \ell\} , \quad (4)$$

$$P_{ij\ell}^- = \Pr\{X_{ij}^- = \ell\} , \quad (5)$$

$$X_{ij}^{+'} = X_{ij}^+ - i - \ell , \quad (6)$$

$$X_{ij}^{-'} = X_{ij}^- - i , \quad (7)$$

$$X_{ij}^{+'} \in P(\lambda_{ij}^+) , \quad (8)$$

where  $X_{ij}$  is the number of residents in a household in category (i,j),

$X_{ij}^{'}$ , the number of residents in a household in category (i,j) in excess of the number given by the category (i,j), and

$P(\lambda_{ij})$ , the Poisson distribution of parameter  $\lambda_{ij}$ .

The superscripts '+' and '-' refer to households with and without children respectively.

It appears from (1)-(8) that the model is a category model with the households divided into 8 categories (1 or 2 principal persons, 4 age categories for the age of the household).

The model operates with principal persons, which are the holder of the apartment plus the spouse if any, and other persons, which we denote subpersons. The principal persons are classified according to economic activity depending on sex, age, marital status and the district category of the dwelling (urban

area, semiurban area and rural area). The subpersons are classified as children or other subpersons. The children are also classified by age and whether they obtain education or not. The statistical distributions used here are polynomial. Note that the structure of the model outlined implies that it is possible to stratify the population according to many different criteria, because the empirical interdependencies between the variables are described at the micro-level, the individual household. In some circumstances it may be of interest to calculate the number of married, unemployed women with no children in order to assess the availability of such labor.

Finally because of the interdependence of income and private car ownership, the household model uses a rough description of the empirical connection between the income of the household and the economic activity of its members.

#### The Submodel of Public Transportation

The submodel of public transportation is based on network theory, but as bus services in rural areas have low frequencies, we must use the discrete times to tell how the branches of the network are served. Average numbers as frequencies etc. make no sense for public transportation in rural areas. This submodel also contains a method of evaluating feasible services in matrix form.

We shall not give any further description of the model here, but refer the interested reader to Jensen (1975).

#### The Submodel of Car Ownership

Traffic surveys in Denmark have shown that car ownership to a great extent can be correlated to the income of the household, the type of the household and the district category of the dwelling.

Consider a given type of household and a given district category of the dwelling. The income is taken into account by introducing:

- I as the income of the household,
- C , the number of cars in the household,
- $f(I|C)$ , the income distribution for households with C cars,
- $g(C|I)$ , the probability that a household (or fraction of households) with income I has the use of C cars,
- $g(C)$  , the probability that a randomly selected household has the use of C cars, and

$f(I)$  , the income distribution for all households.

Now let us use

$$g(C|I) \cdot f(I) = f(I|C) \cdot g(C) \quad \Longleftrightarrow \quad (9)$$

$$g(C|I) = \frac{f(I|C) \cdot g(C)}{f(I)} \quad . \quad (10)$$

However

$$f(I) = \sum_C f(I|C) \cdot g(C) \quad \Longleftrightarrow$$

$$g(C|I) = \frac{f(I|C) \cdot g(C)}{\sum_C f(I|C) \cdot g(C)} \quad . \quad (11)$$

The applicability of this model depends on whether we can find a fair expression of  $f(I|C)$ . A statistical analysis, based on two traffic surveys in Denmark (Egnsplanraadet, 1968; Nyvig, 1972), has shown that the gamma distribution fits the data fairly well. The expression is

$$f(I|C) = \frac{1}{\Gamma(k) \cdot \beta^k} x^{k-1} e^{-x/\beta} \quad , \quad (12)$$

for which we denote

$$I_C \in G(k, \beta) \quad . \quad (13)$$

This illustrates that the gamma distribution contains two parameters  $k$  and  $\beta$ . These may be interpreted, which is the great advantage of using the model ((11)-(12)). The parameter  $\beta$  is a scale parameter, implying that  $\beta$  may be interpreted as a measure of household income and that therefore  $\beta$  increases with time (inflation) and is higher for areas with high costs of living. The gamma distribution is additive to the parameter  $k$ , that is

$$X \in G(k_1, \beta) \wedge Y \in G(k_1, \beta) \quad \Longrightarrow$$

$$X+Y \in G(k_1+k_1, \beta) \quad . \quad (14)$$

Thus if the household income is interpreted as the sum of a series of gamma distributed contributions, then  $k$  increases with increasing economic activity of the household members.

These interpretations of parameters in the gamma distribution increase the applicability of the household model by enabling transfer of information from one survey to another and thus making savings in the collection of data as well as obtaining better forecasts in situations where there is no possibility of getting data.

The interpretations have been tested on other surveys with good results.

#### ACCESSIBILITY

The model is based on accessibility demand and not on transportation demand. This is because we assume the following relation:

Accessibility demand + location + institutional factors  
⇒ transportation demand.

We must plan to satisfy accessibility demand not transportation demand. This can be done equally well by locational or institutional factors as by transportation factors.

Accessibility is defined by specifying the facility to which access is required and the population that demands access. For instance, we may have the following demands for accessibility:

Facility	Population
School	Children 7-14 years
Shopping	Pensioners and married non-economically active persons
Municipal office	All households

The structure of the total model of accessibility is shown in Figure 3 in which direct accessibility means that the dwelling is situated in the same zone as the facility.

#### THE MODEL OF ACCESSIBILITY: A CASE STUDY

The model is at present being used in a survey in the Danish municipality of Middelfart on the island of Funen.



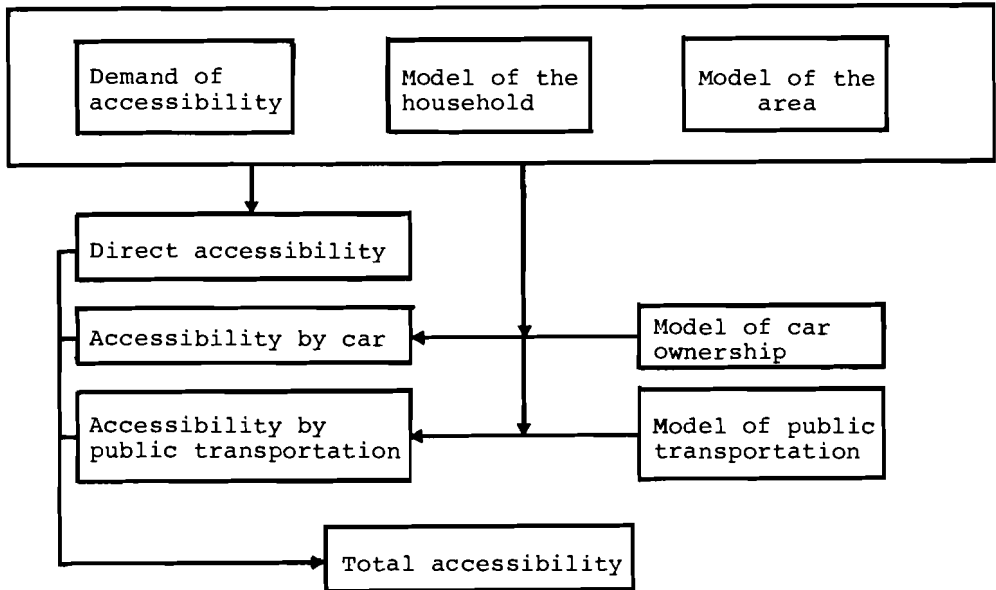


Figure 3. An outline of the model of accessibility describing the socio-economic system and the transportation system of an area. For a given demand of accessibility it is immediately possible by means of the model of the household and the model of the area to calculate the direct accessibility, where no transportation is needed. By applying the model of car ownership, the accessibility by car is calculated and finally the accessibility by public transportation is found by the aid of the model of public transportation. By addition the total accessibility to the population is found.

A procedure for using the model in the technical administrations of the municipalities and the counties has been formulated, but since it is adapted to particular Danish conditions, it is not reproduced here. It generates alternative plans for public transportation in the area in cooperation with an experienced public transportation planner so that such things as organization, existing laws and transportation services etc., not included in the model, can be taken into account. It is this combination of quantitative results from the models and qualitative judgments by people that gives the best results.

By emphasizing the calculation and the presentation of the consequences of alternative plans we will now illustrate how the model may be used by decision-makers in a political decision process. Consider two alternatives, the existing public transportation system and a system, "the plan", elaborated by the procedure just mentioned. The alternatives can be compared by a cost-benefit analysis as follows.

Middelfart has been divided into 9 zones as shown in Figure 4, which also shows the routing network according to the

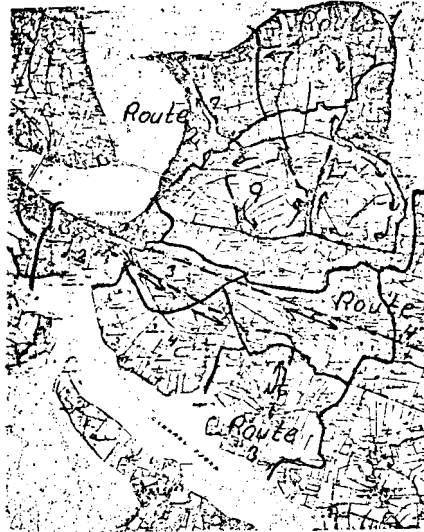


Figure 4. The municipality of Middelfart divided into 9 zones. x, the location of a public school. —, the routing network (planned).

elaborated plan. The benefits of the two alternatives are illustrated by the eight accessibility demands mentioned in Table 1.

Table 1. Accessibility demand corresponding to a series of selected facilities important to transportation.

Facility	Population group
Public school	Children 7-16 years
Business school	Children 15-19 years
High school	
Shops, special goods Visit in or via Middelfart	Pensioners and married non-economically active persons
Hospital	All residents
Places of work	Economically active persons
Municipal office	All households

For any of these demands we have calculated how and to what extent they were satisfied by the actual and planned systems. As an example detailed consequences for shopping are shown in Table 2 and Table 3. We must point out that the calculation of accessibility by public transportation is based on a standard trip with a certain stay-time at the destination and at a certain time of day.

Table 2. The accessibility of pensioners and married non-economically active persons to shops (special goods) (actual system).

Zone no.	Accessibility			
	car	direct	publ. transp.	none
1	0	0	0	0
2	224	2190		
3	38		366	
4	7		51	
5	11		105	
6	12			110
7	59	566		
8	7			63
9	33		160	159
Total	391	2756	682	332
	9%	66%	17%	8%

Table 3. The accessibility of pensioners and married non-economically active persons to shops (special goods) (planned system).

Zone no.	Accessibility			
	car	direct	publ. transp.	none
1	0	0	0	0
2	224	2190		
3	38		366	
4	7		51	
5	11		105	
6	12		110	
7	59	566		
8	7		63	
9	33		319	
Total	391	2756	1014	0
	9%	66%	25%	0%

Tables 2 and 3 show all the accessibility demands and make it possible to evaluate the benefits where distributions, marginals and extremes are included and not just calculated averages. Transportation is an offer to the individual and therefore you cannot rely on averages.

In this case study we emphasized four aspects of the consequences:

- by geographic sectors,
- by travel purposes,
- by population groups, and
- a total evaluation for the whole area.

The first three aspects are highlighted by the relevant conclusions from Tables 2 and 3 for the chosen accessibility demands. Tables 4 and 5 give a total evaluation for the whole area.

Table 4. Accessibility consequences of the actual and planned public transportation system (%).

Basic population group	The group mentioned in Table 1					Only these w. access by public transp. or none	
	Accessibility					Accessibility	
Travel purpose	car	pub. transp.		none		none	
		actu.	plan	actu.	plan	actual	plan
Public school	0						100
Busin. school	0	40	42	2	0	4	0
High school	0	40	42	2	0	4	0
Shops (sp.goods)	9	17	24	8	0	33	0
Visit	9	30	34	8	4	22	11
Hospital	64	11	13	4	2	29	10
Municip. office	56	11	14	6	3	33	18

Table 4 shows that when considering a given population group within the municipality, the improvements made by the planned system over the actual system are fairly small. However if persons with direct access or access by private car are excluded, then for the remaining population which has an acute demand for public transportation, the relative improvements are much bigger. Another remarkable conclusion from Table 4 is that public transportation today is still important. This stresses the fact that public authorities must keep an open eye on public transportation development: it is too risky to lose control of it. On the other hand 50% of the population presumably does not care whether there is a public transportation system or not.

Table 5. The consumption of resources as a result of the actual and planned public transportation system.

System	Total consumption Monday-Friday 6 a.m.-7 p.m.							
	school weeks				non school weeks			
	Vehi- cle km	Hours in sched- ule	Driver hours	No. of vehic.	Vehi- cle km	Hours in sched- ule	Driver hours	No. of vehic.
Actual	5095			min 6	2380	90	123	2
Planned	3945	163	223	4	2495	104	163	4

The calculation of the consumption of resources by the two alternatives shown in Table 5 raises some problems, as the actual is composed of three separate systems - a regional system, a local system and a school bus system. The planned system is an integrated system and thus Table 4 is only partly completed because of a lack of comparative data.

Which system to prefer depends on a political discussion of whether the improved service as expressed by Tables 2-4 and the three distributional aspects mentioned above counterbalances the consumption of resources shown in Table 5, the last implying extra expenses to public funds which depends on local negotiation with transportation companies as to price.

Note that Tables 2-4 are for local use, while Tables 4 and 5 can be used to compare different municipalities in order to get an overall view of the county.

CONCLUSION

The total model of accessibility makes it possible to evaluate the consequence of change in the public transportation system in a very varied way by specifying which travel purposes, which areas etc. benefit and which lose. The example shows that public transportation is not to be considered as any substitute for the private car. The task is to find the right division of work between public and individual transportation. Therefore optimization in transportation requires an overall view - a systems view.

It is stressed in this paper that public transportation is an offer to the individual, and therefore it makes no sense to calculate by averages.

It must be emphasized that the example given illustrates only one application of the model of accessibility, namely, a comparison of alternative plans of public transportation service. The model may equally well be applied to evaluate alternative locations of institutions, alternative patterns of settlement and alternative opening hours of institutions, etc.

#### REFERENCES

- Christensen, J., and A. Friis (1976), *Public Transportation in A Rural Municipality; An Analysis in the Municipality of Hadsund*, IMSOR, Technical University of Denmark, Lyngby (in Danish).
- Egnsplanraadet (1968), *Greater Copenhagen Regional Transportation Study*, Hovedstadsraadet, Copenhagen.
- Frederiksen, A.J. (1976), *Municipal Transportation Planning; An Analysis of the Public Transportation in the Municipality of Middelfart*, IMSOR, Technical University of Denmark, Lyngby (in Danish).
- Hutchinson, B.G. (1974), *Principles of Urban Transport Systems Planning*, McGraw-Hill, New York.
- Jensen, O., *The Interaction Between Public Transportation and Other Social Activities, A Systems Approach* (forthcoming).
- Jensen, O. (1975), *A General Method of Evaluating Transportation Systems of Low Frequency*, IMSOR, Technical University of Denmark, Lyngby.
- Kobayashi, K., et al. (1975), *A Method for Evaluating Urban Transportation Planning in Terms of User Benefits*, *Transportation Research*, 9, 1.
- Koelman, F.H.J. (1972), *Transportation Planning Integrated in the Overall Planning Process*, *Traffic Quarterly*, 26, 1.
- Nyvig, A. A/S (1972), *The Traffic in Aalborg 1970* (in Danish), A. Nyvig, Hørsholm, Denmark.
- Pedersen, P.O. (1974), *The Cost Structure of Buses Operated by Private Railways*, Report 26, IVTB, Technical University of Denmark, Lyngby (in Danish).
- Schlichtkrull, P., and P. Ørum (1974), *The Coordination of Public Transportation in a Sparsely Populated Area*, IMSOR, Technical University of Denmark, Lyngby (in Danish).

## **OPERATIONAL PLANNING**





### Summary of the Panel Discussion\*

The papers of this session contain reviews and original contributions to scheduling and routing methodology applied to virtually all modes of transportation. The most striking conclusion is the common philosophy and basic approach to a number of modes. Although quantitative techniques, especially optimization, are considered important, useful, and deserving of further research, it is recognized that they cannot be expected to solve scheduling problems completely. Instead they are used for suboptimizations in an iterative process. Most essential in this process is the interaction of the planner, who has a number of qualities not readily matched by computer programs. Some of these are the ability:

- to recognize special features in complicated patterns,
- to deal with vaguely defined objectives and constraints, and
- to consider multiple objectives and to resolve conflicts between them.

The computer on the other hand has the ability:

- to perform large numbers of logical and mathematical operations with high speed, and
- to display data in a multitude of arrangements.

It is expected that future developments in hardware and software will lead to significant improvements in data display facilities, as well as data manipulation. This will further increase the importance of the interactive approach to scheduling.

The development of optimization algorithms will of course remain an important part of the work in scheduling, but it was agreed that the availability of algorithms that can perform really meaningful optimizations is many years off. Since it appears impossible ever to include all the objectives and constraints, it may not even be desirable to have an optimization algorithm which produces the optimal solution. Instead it might be preferable to have an algorithm which produces large numbers of good solutions. After all, a solution which is optimal for only some of the objectives and constraints, need not even be

---

\*Prepared by M.M. Etschmaier

good if all objectives and constraints are considered. Since the selection of the best solution from among the good solutions generated is to be made by the planner, great care has to be devoted to the process of generating good solutions. If it is to be useful, the set of good solutions has to be small, yet contain solutions with as many significant differences as possible. Further research in this area would be highly desirable.

The question was raised whether the scheduling problem can be considered as a problem that by its very nature is unsolvable. This might be suggested by the depth of detail that can be included as possibly relevant, and by the fact that, if any detail is studied and included in the solution, the awareness of it in the company is likely to be changed. However, it was agreed that, while it may never be possible to solve the scheduling problem to complete satisfaction, it probably cannot be viewed as a problem that is truly unsolvable.

Abstracts of Papers

*A Survey of Scheduling Methodology Used in Air Transportation*  
M.M. Etschmaier

(This volume, pp. 77-89)

This paper gives an overview of the general procedure followed by most airlines for the development of flight schedules. Models developed for the various stages of this process are described. Special emphasis is placed on the direct and indirect impact that these models have had on the airline industry.

*Some Elements of the Airline Fleet Planning Program--Or, Why Human Fleet Planners (and Not Computers) Make Airline Fleet Plans*

M. Pollack

(Abstract only: a working draft of this paper is available as Report Number ARP/1998, from the author - his address is in the Appendix, p. 231)

Although many optimizing models have been presented for fleet planning, little success at implementation has been achieved. Examined in detail in this paper were some possibly important problem elements that have either been left out, modelled too simply, or not considered at all. Included were: year-to-year continuity conditions, integer variables, isolation of aircraft, nonlinear cost and revenue functions, passenger flows, minimum service levels, multiple criteria, and problem size. Some suggestions were outlined for future model development.

*Scheduling and Maintenance Planning in Rail Transportation*  
R. Genser

(This volume, pp. 91-118)

Levels and stages of complex "scheduling" are outlined for railway systems. The present approaches are discussed for:

- time scheduling (timetable),
- rolling stock rostering,
- manpower planning,

- operating schedules of marshalling yards and stations, and
- maintenance planning.

The connection with subsystems like traffic control, workshop control, etc., and the influence on safety and reliability of the railway system are considered.

Examples of possible, further developments such as the use of

- pattern recognition,
- on-line learning,
- active data base systems, etc.,

are given for unsolved problems.

*Methods for Optimizing Networks of Regular Passenger Transportation Systems*

K.-D. Wiegand

(Abstract only)

Transit planning in rail passenger transportation depends on the advantageous organization of routes, and the appropriate selection of timetable connections. So the parameters service frequency and travel-time (the time interval between successive trains; connections), as well as expenditure on traffic installations (including operational connections, e.g. at level crossings) may be influenced, in theory, by technical timetable connections.

The organization of railway routes has a direct influence on travel time (the shortest or fastest routes; the choice of the largest possible number of direct connections) and the operational expenditure (satisfaction of the traffic demand through a railway system which is economical in its number of lines, but nevertheless covers all sections adequately).

The method used for assessing the best structuring of the railway system is graph-theory. With its help, it is possible to decide on the minimum number of lines in transit networks (the necessary number of open chains for realization of the network-function). It is possible, moreover, to limit the number of chains from the outset, since only such chains as are identical to optimal paths are possible components of the railway system. The decision of by what criteria such paths should be determined, can be made - the rule is that the shortest routes in time are the best.

The real organization of railway-line networks is then represented as a combinatorial optimization exercise, for the solution of which a branch-and-bound algorithm has been developed. Simplified, this treats every judgment on the acceptance of an individual line into the line network as a branch; and every characteristic maximum value for the number of passengers able to travel direct, (a number still in the deciding stage) as a bound.

The planning of railway-line systems with the help of the methodology described is at present being carried out for passenger transportation by the German Railways.

*Scheduling Problems in Ship Operation*  
G. Rognstad

(Abstract only)

Each kind of employment has its own particular scheduling problem, although they all consist of minimization of transportation costs or maximization of the profit for a fleet for a given period of time.

The problems to be found in world-wide tramp shipping, where the prospective voyages are unpredictable, are different from so-called "contract-trading", where the cargo to be transported and destination is known, and are also different from liner service, where the sailing schedule is fixed, but the "cargo" is unknown almost until one reaches the various ports. But still they all have one thing in common: an enormous amount of money can be saved by better planning or scheduling of the fleet.

The existing methods of solution, especially for a combined tramp-contract scheduling problem and liner trades are attempts based on linear programming, network analyses, and simulation techniques and they are more or less successful. However, these methods are only in practical use to a very small extent in shipping today. Some experiences from methods tried, or in practical use, were discussed, as were the reasons for the limited utilization of more advanced methods in shipping today, and the reasons for their failure.

*Optimal Planning of Shipment Volumes*  
O.I. Aven, S.E. Lovetsky, and G.E. Moiseenko

(This volume, pp. 119-128)

The planning of marine transportation is made within a multi-level management system and is subdivided into long-term, medium-term and short-term planning.

The large size of the problems to be solved, the wide geographical dispersion of the system's operating units, and other factors make the problem of optimal transportation planning quite complex.

The report presents a system of two-level models for optimal long-term and medium-term shipment planning. The models are based on the decomposition principle: the planning problems are represented as sets of interconnected subproblems.

There is a certain correspondence between the degree of data aggregation and the range of planning. A medium-range plan is a more detailed version of the long-term one and uses more precise information about transportation process.

An iterative algorithm is proposed which solves the set of problems and provides the optimal allocation of cargo to be shipped among the operating units of the system. It also coordinates the plans of different levels as well as with different ranges.

The problem of long-term plan correction and reformulation of these models to take into account the random nature of the parameters is also discussed.

*Optimal Scheduling of Trucks: A Routing Problem with Tight Due Times for Delivery*  
O.B.G. Madsen

(This volume, pp. 129-136)

This paper describes a truck scheduling problem from a theoretical and a practical point of view. The problem is to determine the routes for a given number of trucks given a very tight time schedule. The problem is first considered from a theoretical viewpoint. Then two heuristic methods are presented and a summary is given of their application in practice.

*Optimal Bus Routing: A Case Study*  
C. Mandl

(This volume, pp. 137-147)

In Aarau, the routes of public buses are to be reorganized. Out of the street network, a subgraph is constructed of those streets along which buses are allowed to travel. The city is divided into small areas for which the movements of people using public transportation between all pairs of areas is given. It is assumed that these movements are independent of any particular network of bus routes. A special restriction is that all bus routes have to pass by the railway station located in the center of the town. Together with the local authority, two objectives were worked out: to minimize travel time for each passenger and

to minimize necessary changes of buses during passengers' travel. Clearly travel time depends also on waiting time due to bus changes, but this could not be included in the model because the frequencies of buses were not known before the bus routes were constructed, and that depended strongly on the actual bus routes. So the problem had to be broken into two parts: first, to decide on the bus routes and secondly, (not part of this study) to decide about bus frequencies. A heuristic algorithm is then presented that minimizes the number of passengers who have to change buses, and that produces bus routes close to the shortest paths within the given graph. These bus routes are then measured against various other criteria, and a sensitivity analysis carried out.

In connection with the case study, some general remarks on various assumptions and restrictions within the model are discussed, namely:

- the assumption that demand for a public transportation system is independent of the actual transportation network existing and travel time,
- the restriction that all routes or even many should pass through the center of a town, which leads to an extremely high traffic density in the center, and
- the restriction that, for example, all buses should travel with the same speed on all routes, which, as is shown, also leads to high traffic density in the center.

Finally, some ideas on the problem of defining objectives for public transportation systems are presented.





A Survey of Scheduling Methodology Used  
In Air Transportation

M.M. Etschmaier

A case can be made for describing the contribution of operations research to airline scheduling as outstanding or as barely existing. Airline scheduling is probably one of the first areas towards which operations researchers pointed their attention. One might think that this happened because the schedule is the part of an airline most often seen by the general public. Perhaps operations researchers were experiencing difficulties with flight connections and thought it would be a simple task to rearrange the schedule a little and obtain a much more efficient way of going from one place to another. I believe they thought for quite some time that scheduling airlines was relatively easy and could be carried out by a large linear model, linear programming being the obvious method. It might also be necessary to do some integerization - not a difficult task either. However, they soon discovered that the number of variables to consider increased with the detail of the model, and that number was increasing more rapidly than the memory of available computers. But people still hoped that eventually the size and efficiency of computers could catch up and permit some sort of closed form of this problem to be solved.

Unfortunately, although many such closed form endeavors were begun, none of them were ever completed, because the learning process and increase in insight into the problem was much faster than what modellers could add to models to accommodate it.

As work progressed, researchers started to realize the tremendous difficulty of developing a schedule for an airline, and the consensus today is that the problem of airline scheduling is more or less unsolvable, at least by quantitative techniques. For one thing, scheduling an airline virtually amounts to the optimization of the whole airline operation: its effects extend into almost every department. And, for another (as with the application of optimization to any kind of industrial enterprise), there are a number of objectives and constraints which are highly ambiguous. These are presented to the model builder as hard and solid requirements, and he develops a model which takes them into consideration. Inspecting the results may show that some constraint is limiting the solution rather severely and anybody in charge can say "maybe something can be done about changing this constraint; maybe it is not that important after all; maybe we can go back and talk to the people". In airline scheduling especially, many constraints come from what one can do with people: union contracts are a typical example. Often one can offer the

union representative concessions on other matters if they will agree to a change in a constraint. This kind of give-and-take is not something which one can expect to formalize or incorporate in a model, but depends largely on extraneous factors like the current situation or climate in the company, or the personal relationship of the person in charge of one department to the person responsible for some other department, etc.

On the other hand, it is easy for a human planner to recognize these possibilities; it does not take him long to determine whether something can be done about relaxing too restrictive a constraint. His judgment is clearly not the result of rigorous reasoning but is rather determined by his feelings based on a good perception of the relevant environment. People working on airline scheduling have realized the great contribution the human mind can make to developing complex statements such as airline schedules. And an airline schedule is indeed a very complex statement: it is not just a list of numbers, but a list of numbers which have to interact in many different ways.

The ability of the human mind to grasp complex relationships depends on the way in which these relationships are presented. Data have to be reduced to the bare bones, so that the human can immediately relate to them and process them mentally. This kind of reduction is systematic. It is difficult to specify the best form of reduction for any given problem, but the kind of reduction which is required does not change as rapidly as the kinds of conclusions which have to be made. So it is possible to arrive at computer models which perform a great deal of data reduction, manipulation and presentation of data in a variety of different ways and permit the human to grasp these data and work with them directly.

As a consequence, work in airline scheduling has concentrated on the development of a systematic interaction between man and computer. The computer, especially in the initial stages, is confined to handling the very tedious arithmetic chores involved in determining the feasibility of a schedule, while the man's role is recognizing patterns and suggesting new solutions or alterations to existing solutions in the light of certain evaluations.

To date, the role of the computer has been rather limited and the full potential of the approach sketched above has not been realized. There are no systems in practical use which provide an optimal interface between man and computer, and there are few which employ formal operations research models for optimization and simulation. However, one cannot say that operations research has made no great and significant contributions in this area, just that the contributions made lie mainly in the conceptualization and formalization of the scheduling process. This by no means simple or easy task involves definition of the steps through which the process evolves as well as of the kind of interaction between man and computer. Significant work has been carried out in both areas.

## THE ENVIRONMENT FOR SCHEDULE DEVELOPMENT

In talking about airline scheduling, we have to differentiate between the development of a completely new schedule from scratch (this usually is not the problem which an airline faces) and the modification of an existing schedule. An airline, when developing a schedule, looks at last year's schedule, and might even look at schedules as they could evolve in the future. A lot of the detail for such modified schedules already exists. This is true even for the introduction of new fleets because one fleet is not switched to another overnight, but over several years.

Completely new schedules are needed for long-range planning purposes and then their details are unimportant. They are developed merely so that one can verify whether under certain circumstances an airline can be profitably operated without violating the many difficult scheduling constraints. They are also developed to assist aircraft manufacturers in their market research, and regional and national governments in the formulation of air transportation plans and policies. The question is what type of aircraft in what numbers can be operated profitably in a given region, which means considering the likely future demand for travel on different routes so that a good use can be made of all aircraft throughout the day. At present, details of possible schedules are needed for this. The problem is that governmental agencies and aircraft manufacturers do not have the expertise to draw up a schedule like an airline does, and airlines, on the other hand, do not have the time and cannot afford to take experts out of the regular planning process and assign them to developing a schedule which will very definitely never be used in the form in which it is developed.

In this presentation we will briefly describe the schedule development process in use by airlines, and then scan a few of the models which have been developed. It is clear from what has been said that most of these models are not used frequently or at all in the actual schedule development process of an airline, but more for planning purposes of governmental agencies or airline fleet production. Of course, the day-to-day scheduling by an airline tries to draw from these models and incorporate as many of their features as possible.

## THE SCHEDULE DEVELOPMENT PROCESS

The schedule is one of the central elements in any corporate plan of an airline. It exists in different forms depending on the detail required. Generally one can distinguish between

- a route and frequency plan, which gives the routes and the frequency with which the different types of aircraft fly on them, and
- a complete schedule, which includes times of arrival and departure.

A schedule evolves over several years from a simple statement of objectives through numerous refinements of route, frequency plans, and complete schedules to a final executable schedule. The process is outlined in Figure 1.

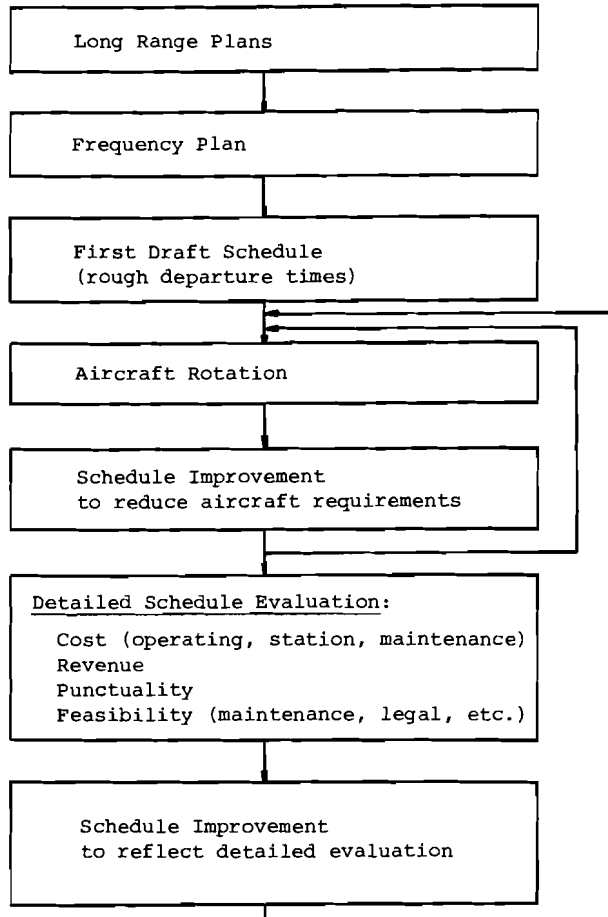


Figure 1. Airline scheduling overview.

The time over which this process occurs varies from airline to airline. An example is shown in Figure 2. At any point in time there exists a schedule for each year within the planning horizon, each of which is in a different stage of development. It is apparent that much interaction takes place between the schedules for the different years within the planning horizon. The problem of optimizing the schedule is not a one period problem but a problem of viewing many periods in perspective. Thus the schedule construction process is not only one of evolving a schedule through the planning horizon, but also one of evolving schedules for future periods from those for preceding periods.

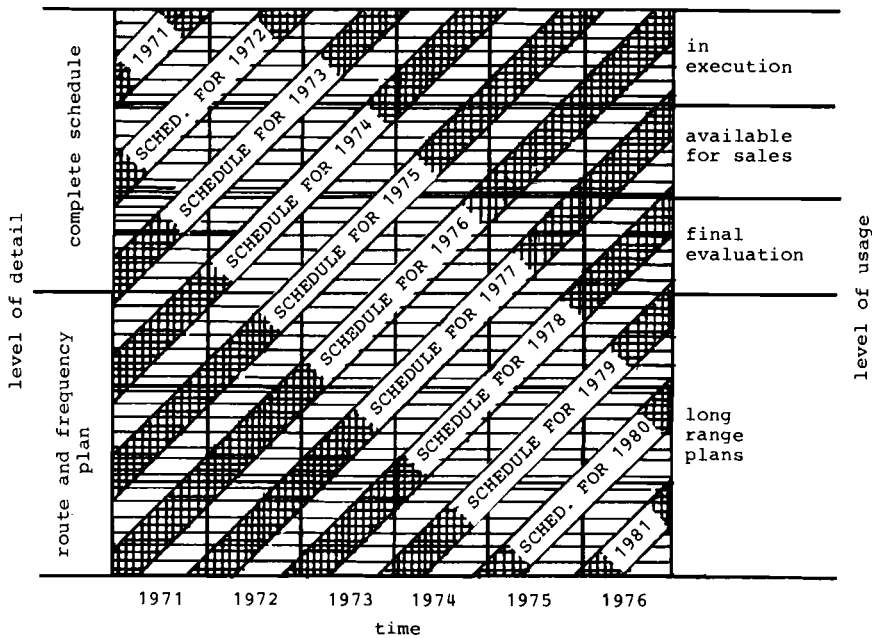


Figure 2.

Many flights planned for one year might not in themselves be profitable in that year, but are included to develop a new market, to have a head start over the competition, or for other reasons.

It follows from Figure 1 that the evolution of a schedule for a certain period starts with a general statement of long term objectives. From this, route and frequency plans are developed which may be viewed as schedules without times attached to them. They contain routes on which one wants to fly and the frequencies and types of aircraft with which one wants to fly on them.

In the next step, departure times are attached to the frequencies to yield a rough draft schedule. Times are mainly chosen for their market desirability, and operational objectives and constraints are largely neglected. These are considered in the next step, when one tries to fit the chosen flights into aircraft rotations with as high an aircraft utilization as possible. So one tries to change flights to increase aircraft use or, what comes to the same thing, reduce the number of aircraft required. The construction of aircraft rotations and subsequent schedule improvement is carried out iteratively at the central scheduling or planning department.

After a reasonably realistic schedule has been arrived at, it is sent out to the various operating departments for their

detailed evaluation, a process which would have been much too complex to be considered in the first development of the schedule. The subsequent inclusion of their recommendations in the optimization usually only requires quite minor changes of the schedule. The operating departments feed back how much it would cost for them to participate in the execution of the proposed schedule and possibly what kinds of additional investment or manpower would be required to do that. The feedback from all departments is collected by the central planning unit which then comes up with a revised schedule that is again sent out to the various operating departments and initiates another round of evaluation, another iteration. The number of iterations which are gone through for any schedule vary from airline to airline. Some go through as many as three or four, the number being limited by how quickly they can process all the information which is generated that way.

#### SOME MODELS FOR THE SCHEDULE DEVELOPMENT PROCESS

##### Frequency Optimization

The problem considered in frequency optimization is that of finding the number of flights per aircraft type and route that maximize the total profit for the system. Obviously this formulation is a gross simplification of the real problem, since frequencies alone determine neither revenue nor cost. For this reason it does not pay to put too much sophistication into frequency optimization and only the crudest models are used. One can only hope that solutions arrived at by a frequency optimization can be developed into an optimal schedule.

The basic frequency optimization model dates back to a paper by Dantzig and Ferguson and is described in Dantzig's classic book on linear programming (1963). The model assumes that the profit on a route is strictly linear and only constrained by a fixed and known demand and by the availability of aircraft. The problem can be stated as follows:

$$\text{Max } \sum_i \sum_j x_{ij} p_{ij}$$

subject to:

$$\sum_i c_i x_{ij} \leq D_j \quad \text{for all routes } j ,$$

$$\sum_j h_{ij} x_{ij} \leq H_i \quad \text{for all aircraft types } i ,$$

where

$x_{ij}$  is the frequency of aircraft type  $i$  on route  $j$  ,

$p_{ij}$  , the profit per frequency  $x_{ij}$  ,

$h_{ij}$  , the block hours per frequency  $x_{ij}$  ,

$D_j$  , the demand on route  $j$  ,

$c_i$  , the capacity of aircraft type  $i$  , and

$H_i$  , the available block hours on aircraft type  $i$  .

If one ignores the fact that frequencies can only assume integer values, solutions can be obtained by linear programming. Insistence on integers with this simple formulation would not be worthwhile, since the model is unrealistic for most airlines and the need for rounding might present a welcome opportunity for manipulating the solution.

It is easy to add restrictions on the frequencies ( $x_{ij}$ ) or on any summation of them into the model. Also rather straightforward is the incorporation of multisector routes. One only needs to replace the first set of constraints from the classical model by two new sets of constraints:

- A city pair ( $j$ ) can be satisfied by several routes ( $k$ ). If  $y_{jk}$  is the portion of demand on city pair  $j$  which is satisfied by route  $k$  then we have

$$\sum_k y_{jk} \leq D_j .$$

- One route ( $k$ ) satisfies portions of several city pairs ( $j$ ). If  $S_k$  is the set of city pairs satisfied by route  $k$  and  $S_{kr}$  is the subset of  $S_k$  with the use of segment  $r$  of route  $k$ , then we have

$$\sum_i c_i x_{ik} - \sum_{j \in S_{kr}} y_{jk} \leq 0 .$$

This formulation might mean a large increase in the number of variables. Although its higher degree of realism might make it worth using integer linear programming, this is

usually ruled out by the limited ability to handle integer linear programming problems of a larger size.

Further extensions have been proposed and developed, one of which includes connecting passengers in the same way in which multisegment routes have been included. Alternatively, connecting passengers can be assigned to routes outside the model, leaving of course a crucial part of the model in the hands of the person who makes that assignment.

Another extension includes limited availability of a station (h) in the following way:

$$\sum_{j \in U_h} \sum_i f_{ij} x_{ij} \leq F_h ,$$

where  $U_h$  is the set of routes which originate at station h. Of course the same type of equation can be used if the limitation is imposed on the amount of fuel loaded at some set of stations.

Variations of the basic frequency optimization model use nonlinear cost functions and nonlinear constraints. Nonlinear cost functions are used to express the frequency dependence of cost on air traffic congestion, waiting time of passengers for a flight, and terminal operating cost. These nonlinearities are usually handled by piecewise linearization. Nonlinear constraints are used to express the frequency dependence of market shares on a route. The percentage of passengers that an airline can capture for a route is assumed to be proportional to the percentage of flights operated by that airline over that route. In fact, the percentage might be modified by differences in attractiveness of aircraft, popularity of the airline, and other factors. Solutions have been obtained by quadratic programming on a quadratic approximation and by an approximation on the Kuhn-Tucker conditions.

An entirely different approach to frequency optimization is taken in the so-called port linkage problem which has been formulated for long haul airlines on routes with low density of demand, an approach also usable for local air service in sparsely populated areas. The port linkage problem essentially consists of finding cities to be included in a set of routes between two end points such that the utilization of the aircraft capacity is optimized over the entire route. The construction of the route may be subject to a multitude of operational constraints. Solutions have been obtained by mixed integer programming with the use of Benders' decomposition method.



### Assigning Departure Times to Flights

The problem of finding optimal departure times for a given route can be divided into:

- modelling the time-of-the-day preference of passengers, and
- finding optimal departure times for a given model of the time-of-the-day preference of passengers.

The time-of-the-day preference model is really only part of a much more complicated demand model that expresses the number of passengers on some given flight as a function of all relevant factors, departure time being just one of them. When the frequency of flights on a route and the aircraft types assigned are known, then the remaining variability in the demand function could be viewed as

- the number of passengers who would be willing to board a flight at a given time, and
- the competition of flights for the same passengers.

Two essentially different models have been developed to express this remaining variability:

- a density-type representation of demand, and
- a weighting factor approach.

Although the density-type representation looks intriguing from the point of view of the schedule construction problem, linking it to the complete demand function entails considerable difficulty. The model assumes that every potential traveler has one single choice of time at which he wants to travel. Of all the flights offered he chooses the one which causes him the least inconvenience. Inconvenience is measured in terms of the time difference between the passenger's choice of time and the departure time of the flight, possibly modified by some weighting factors accounting for the attractiveness of the flight (the travel time, the number of intermediate stops, the level of comfort of the aircraft, etc.). While in some models a passenger accepts any level of inconvenience, in others this level is limited by a constant. Passengers who are inconvenienced beyond this constant are assumed simply not to choose to fly. A refinement of this approach assumes a function that, for any level of inconvenience, expresses the proportion of passengers willing to accept it. The use of even the simplest form of density-type representation leads to considerable difficulties in finding an optimal set of departure times. The problem is further complicated if one introduces multi-stop flights, especially if one segment might be included in a number of different routes. The procedures that have been developed to assign departure times to

flights therefore usually amount to some more or less arbitrary heuristic rule. Most of these rules select departure time sequentially.

The weighting factor approach enables a demand function to be applied directly. With a fixed frequency, it assumes that every departure time can attract some given number of passengers so long as the nearest flight in time is sufficiently far away. Modification of these models to account for the attractiveness of the aircraft is straightforward. Consideration of the competition between flights for the same passengers significantly increases the complexity of the models because it requires knowledge of the departure times of other flights to determine the number of passengers attracted - the same information required by the density-type models. However, the weighting factor approach provides significantly more flexibility in considering all the relevant details of the situation.

The time-of-the-day models which have been developed thus far appear rather limited in scope and realism. One of the reasons for this might be that the use which airlines expect to get out of such models is small. Departure times are influenced by many factors that cannot be considered in models, e.g. marketing strategies, the influence of regional managers and past schedules, and many operational considerations. The importance of these models lies in their use for long-term planning by governmental agencies and aircraft manufacturers, and only to a limited extent by airlines themselves. For these uses however it is only necessary to obtain rough answers: sophisticated time-of-the-day models would be unnecessarily complex.

#### Aircraft Rotation Models

Aircraft rotation planning is concerned with the assignment of flights to individual aircraft. The work to be discussed covers two complementary problems:

- how can a given schedule be realized with a minimum number of aircraft.
- how can the schedule be changed to reduce the number of aircraft required.

It has been observed (e.g. Loughran, 1972) that the minimization of the required number of aircraft is not completely realistic in airline schedule planning. The actual situation is that an airline has a more or less fixed number of aircraft available and it has to find a schedule which makes optimal use of these. The problem to be solved by the rotation model is to assign aircraft to flights and identify changes in the schedule such that it can be carried out with the aircraft available. However, defining the problem in this way might tend to obscure underuse of the fleet and the potential for introducing more flights.

Minimization of the number of aircraft required, of course, conflicts with some of the objectives of a good schedule, the most important ones being punctuality and station costs. Nevertheless, by revealing idle aircraft, such a minimization might facilitate substantial savings. After all, empty aircraft can be leased out or sold. Having whole aircraft to work with also makes manual planning of additional flights and routes much easier. Minimization of the number of aircraft required also provides a much clearer objective function suitable for mathematical programming. If the objective were solely to find feasible solutions, secondary objectives would have to be introduced to choose between the possibly large number of feasible solutions. No such objective suitable for mathematical programming has yet been proposed.

Aircraft rotation models, to be realistic, have to take into consideration the maintenance requirements of the aircraft. While these requirements vary considerably with aircraft type and therefore airline, as far as aircraft rotation planning is concerned they can all be reduced to a simple formula. After a fixed number of flight hours have elapsed, the aircraft has to undergo a maintenance check of more or less fixed duration. Such maintenance checks can only be performed in one or a few selected bases. In many cases there is a hierarchy of maintenance events which have to be considered, some of which may be restricted to certain times of the day. This causes a substantial increase in the difficulty of the problem. In many cases maintenance is carried out during the night, when aircraft use is low anyway. Also many route networks are laid out such that the aircraft returns to some main base very frequently. In these the rotation plans meet the maintenance requirements automatically, and so it may be advantageous to exclude the maintenance constraints when constructing the rotation plan and simply check afterwards if they have been met. If not, small modifications of the solution may satisfy them. For these reasons the construction of aircraft rotation plans that do not include maintenance constraints have received considerable attention and comparatively little work has been devoted to models that do. Aircraft rotation models of the former sort use some assignment algorithm. Any assignment rule leads to the minimal number of aircraft required, so long as it makes use of aircraft which have already arrived rather than introducing a new aircraft to serve a flight. Models to include maintenance constraints usually follow some heuristic procedure of the branch and bound type. By using the dual of the simple assignment problem and some insight into the degenerate nature of the problem, it is easy to identify those connections that may be used in a solution that requires the minimal or some given higher number of aircraft. The heuristic procedure can then be limited to these connections.

BIBLIOGRAPHY

- Agard, J. (1968), The Air Force 'Tarage' Plan, *AGIFORS*\* 8.
- Brough, B.M. (1962), The Proper Study of an Airline, *AGIFORS* 2.
- Brough, B.M. (1966), The Integration of Management Information System Planning - The BEA IMPACT Concept, *AGIFORS* 6.
- Dantzig, G.B. (1963), *Linear Programming and Extensions*, Princeton University Press, Princeton.
- Elce, I. (1970), The Development and Implementation of Air Canada's Long Range Planning Model, *AGIFORS* 10.
- Etschmaier, M.M. (1970), Schedule Construction and Evaluation for Short and Medium Range Corporate Planning, *AGIFORS* 10.
- Etschmaier, M.M. (1973), *Projects and Implementations in the Schedule Development Process of an Airline*, Technical Report No. 17, Department of Industrial Engineering, University of Pittsburgh, Pittsburgh.
- Etschmaier, M.M. (1975), *Planning of Local Air Service*, Technical Report No. 31, Department of Industrial Engineering, University of Pittsburgh, Pittsburgh.
- Etschmaier, M.M., and R.J. Richardson (1973a), *Benders' Decomposition Algorithm and the Port Linkage Problem*, Technical Report No. 10, Department of Industrial Engineering, University of Pittsburgh, Pittsburgh.
- Etschmaier, M.M., and R.J. Richardson (1973b), *Improving the Efficiency of Benders' Decomposition Algorithm for a Special Problem in Transportation*, Technical Report No. 14, Department of Industrial Engineering, University of Pittsburgh, Pittsburgh.
- Jessiman, W.A., and D.E. Ward (1970), *Intercity Transportation Effectiveness Model*, Peat, Marwick, Mitchell, and Co., Boston, Mass.
- Kushige, T. (1963), A Solution of Most Profitable Aircraft Routing, *AGIFORS* 3.
- Lombarda, P., and B. Nicoletti (1971), Aircraft Rotations by Computer, *AGIFORS* 11.
- Loughran, B.P. (1972), An Airline Schedule Construction Model, *AGIFORS* 12.

---

\**AGIFORS: Proceedings of AGIFORS Symposium No. 10*, published by American Airlines, New York.

- Niederer, M. (1971), Frequency and Schedule Optimization by LP, *AGIFORS* 11.
- Richardson, R.J. (1973), *Benders' Decomposition Method Applied to an Airline Routing Problem*, Ph.D. Dissertation, University of Pittsburgh, Pittsburgh.
- Richter, H. (1968), Optimal Aircraft Rotations Based on Optimal Flight Timing, *AGIFORS* 8.
- Richter, H. (1970), Experience with the Aircraft Rotation Model, *AGIFORS* 10.
- Simpson, R.W. (1968), A Review of Scheduling and Routing Models for Airline Scheduling, *AGIFORS* 8.
- Simpson, R.W. (1969), *Scheduling and Routing for Airline Systems*, Flight Transportation Laboratory Report FTL R 68-3, MIT, Cambridge, Mass.
- Soudarovich, J. (1971), Routing Selection and Aircraft Allocation, *AGIFORS* 11.
- Struve, D.L. (1970), Intercity Transportation Effectiveness Model, *AGIFORS* 10.
- teWinkel, D. (1969), An Algorithm for Aircraft Scheduling in a Radial Network, *AGIFORS* 9.
- Tobin, N.R., and T.E. Butfield (1970), OR Models and their Implementation in the Airline's Planning Cycle, *AGIFORS* 10.
- Walker-Powell, A.J. (1969), Aircraft Scheduling by Computer, *AGIFORS* 9.
- Walker-Powell, A.J. (1970), The Port Linkage Problem, Unpublished paper, presented at the AGIFORS Study Group on Schedule Construction and Evaluation, Vienna.



Scheduling and Maintenance Planning  
In Rail Transportation

R. Genser

GENERAL REMARKS

For technical, economic and human reasons a railway system complex is broken down into functional subsystems, for example:

- Finance
- Operation
  - freight
  - passenger
- Personnel
- Mechanical engineering
  - rolling stock
- Civil engineering
  - track and buildings
- Signalling and communications
- Electrical engineering
- Supply
- etc.

There is a hierarchical structure of planning with long range planning at the top and operational planning at the bottom. A very simplified model with regard to the boundary conditions and actual parameters of influence is used and it relies on both forecast and predicted data. At the operational level, real states of the system have to be considered, but at the planning stage freedom in the solution domain is reduced by investigating a solution grid. The output of an upper level is an input, boundary condition, or optimization objective for a lower level. Of course feedback exists, but with increasing distance between levels, the time lag generally increases. The state of other subsystems is regarded as quasi-static, and represented by, for example, boundary conditions. It is assumed that the organization, change of objectives, etc. of another subsystem are known a priori.

In general, the decomposition found in institutions are satisfactory because they result from long experience. This means that the information flow between subsystems will be minimal in view of mutual influence on dynamics and according frequency and periodicity.

The approach of the Japanese National Railways (JNR) is given as an example in Figure 1. The diagram shows their planned and already installed railway information system [23, 62].

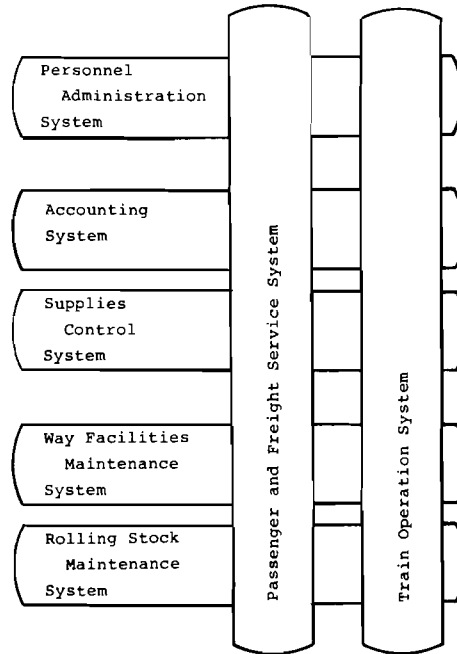


Figure 1. Railway information system of JNR [22].

The outline of the railway traffic planning system in Figure 2 [23] points out the decomposition of the problems. A block diagram for train traffic operation is given in Figure 3 [63]. Usually railway systems cover large areas. Therefore beside the functional and temporal decomposition, a topological decomposition will be found for technical and organizational reasons (see Figure 4).

The results of optimization not only depend on algorithms, but also on:

- the information system used,
- data presentation and gathering,
- the organization, and
- the possibility of adapting the system to technological progress and learning.



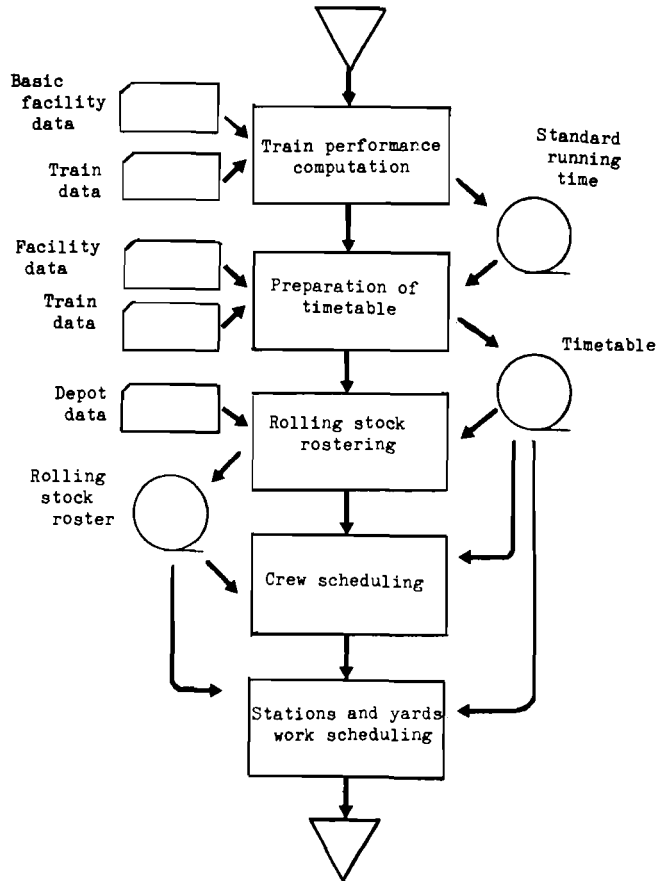


Figure 2. Railway traffic planning system [23].

The expense of implementing a new system is trivial compared with the expense of maintenance and the adaptation to change in technology during the system's lifetime.

In comparison with other transportation systems, there are very stringent safety and reliability regulations for railways in most countries for historical reasons. And it is possible for railways to be both the most safe and reliable in an economical manner, especially with heavy traffic flow. To achieve these two main objectives, every decision has to be not only optimal, but it has to fail safe!

Safety and reliability are not only gained during the operation of the railway system, but also considerable effort for



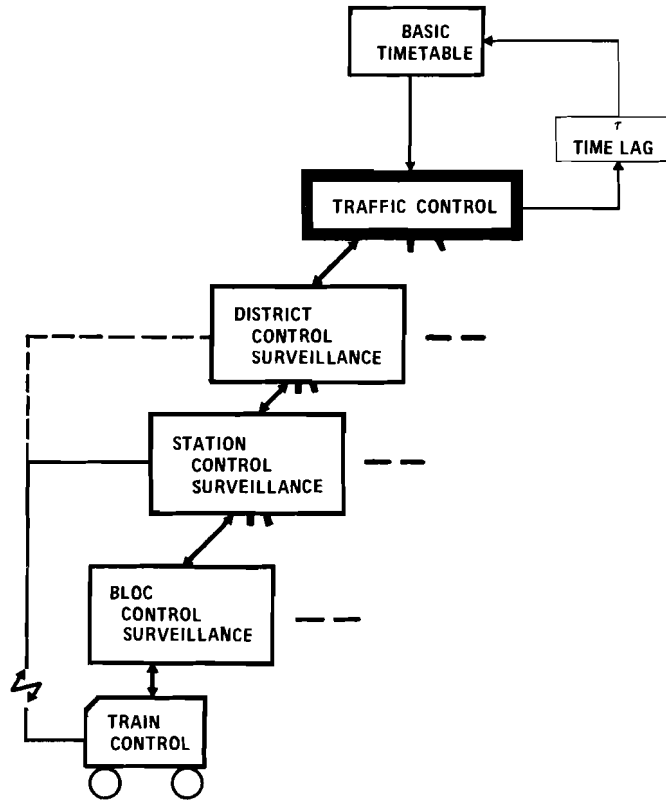


Figure 4. Traffic control.

this objective has to be applied during the planning phase for a system. A safe system should have the following qualities:

- step-by-step realization should be possible,
- fitted to step-by-step maintenance, and
- man should be able to operate the system in an auxiliary mode in case of disturbance or destruction.

Step-by-step realization has the additional advantage of gaining subgoals, and allowing the feedback of learning by experience to be used before the whole system is implemented.

The way to improve the reliability of an existing railway system by switching to modern technology is shown in Figure 5.

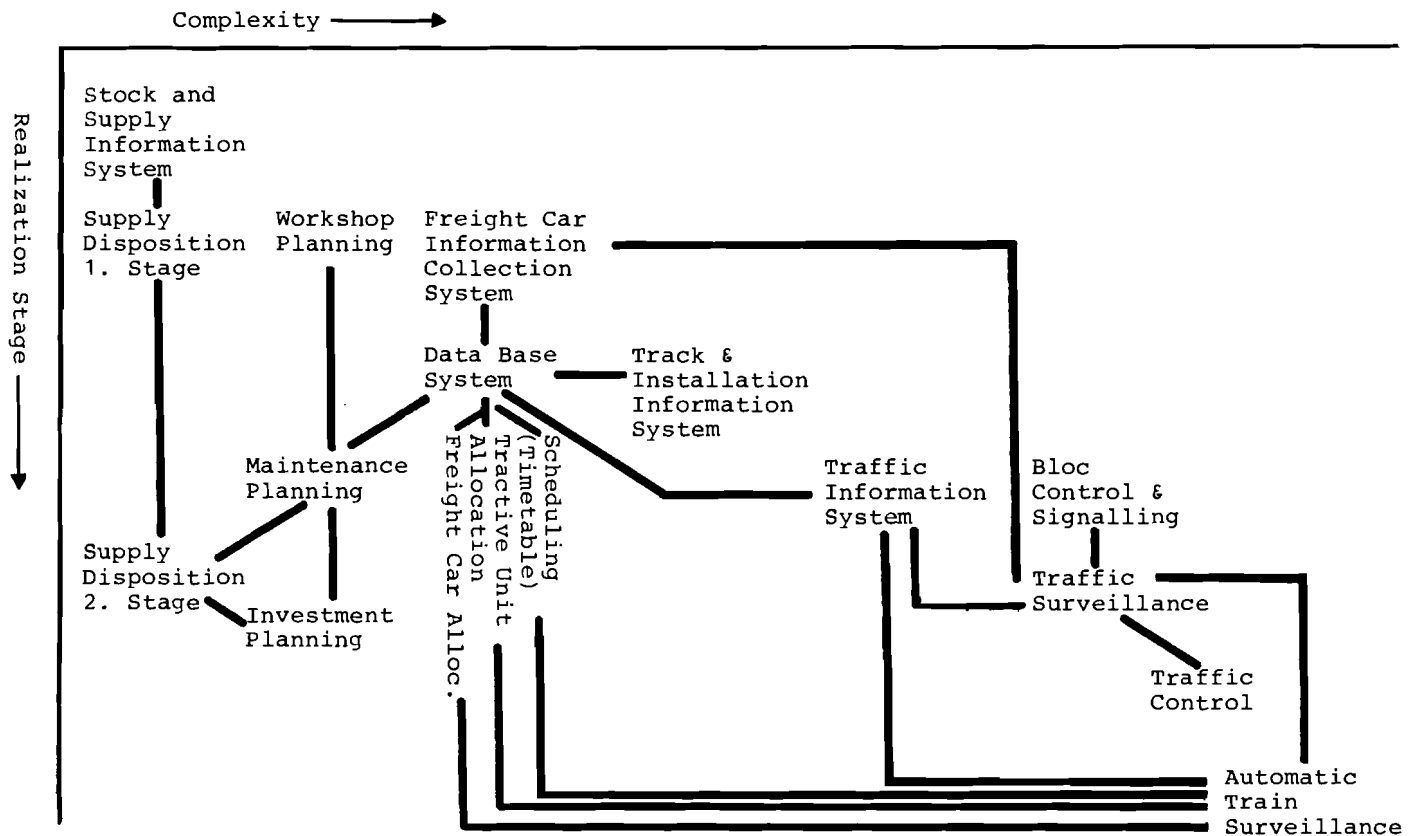


Figure 5. Improvement of an existing railway system.

The schedule of realization is influenced by the complexity of the problems, which depends on:

- the state of technology,
- the methods to be used,
- the degree of integration and influence on other subsystems or on organization, and
- the experience of the user.

In complex systems, it has to be recognized that a new method or technology can only be applied if the user has had experience with it before the design phase. This will be the rule for optimal project management in safety related systems. If there is a lack of practical operational experience, then the correct requirements for designing the system can never be given. The time lag between the solution of different subproblems is needed to get feedback by learning, as shown in Figure 5. It can be seen that an integrated approach is needed, especially if structural or organizational changes are caused.

Figures 6 and 7 should give an idea of the planning and information systems in two railways [62, 33].

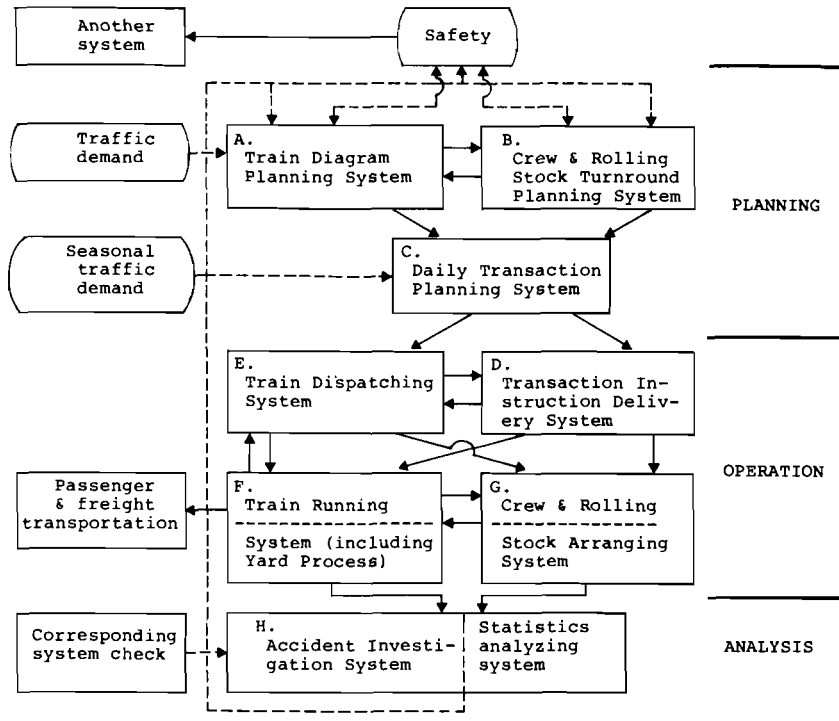


Figure 6. Train operation system OPERUN of JNR [62].

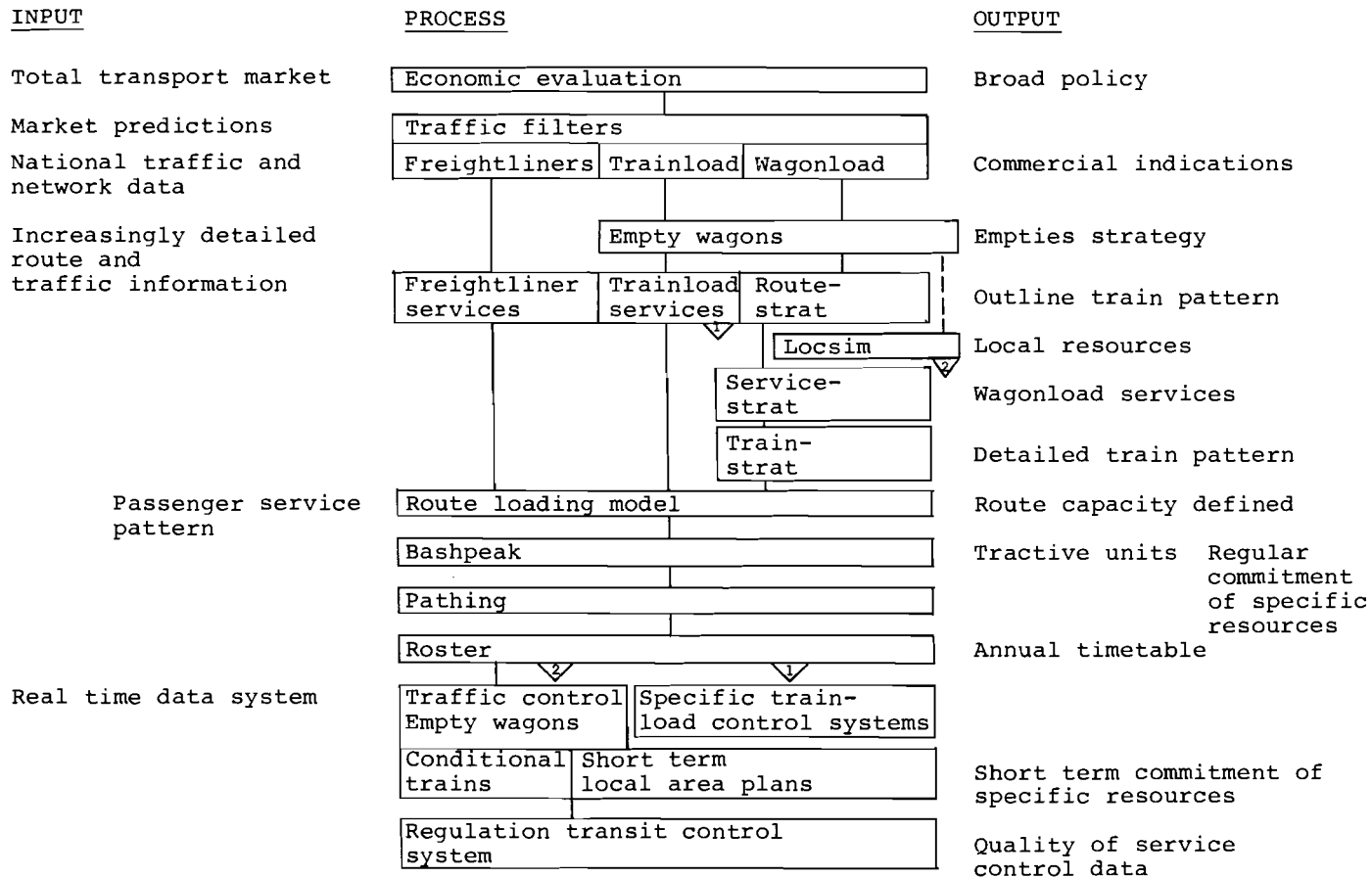


Figure 7. Planning system of BR [33].

SCHEDULING

Up to now, timetable optimization and train scheduling have been the most difficult problems in a railway system. Because of the unique condition given by using a track, scheduling has more importance for railway operation than for other transportation modes. But just optimal train scheduling with the help of computers would fill a need, since otherwise, when disturbances occur, the resultant information flow overburdens man when trying to decide what to do. Computers are at their best in the routine proofing of large complex logical operations. But the complexity of the actual situation, with moving decision boundaries and changing objectives resulting from experience with the system, does not allow the use of fully automated solutions procedures. A man-machine conversation is necessary, where use is made of the ability of man to recognize complex patterns and to not neglect the feedback of learning, and the computer or logical devices grant fail safe behavior.

The approach of experience railway authorities is shown in Figures 8 and 9 [19, 22, 64, 14, 41, 39], and they, of course, rely on methods [56, 1, 45, 18] like:

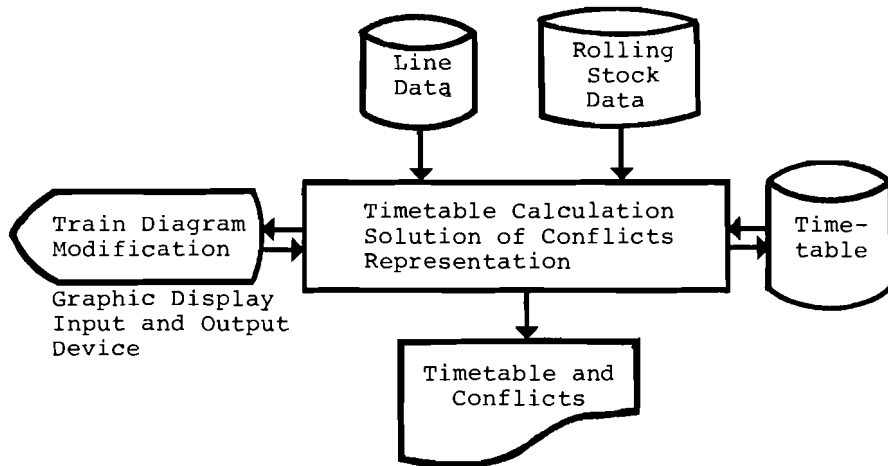
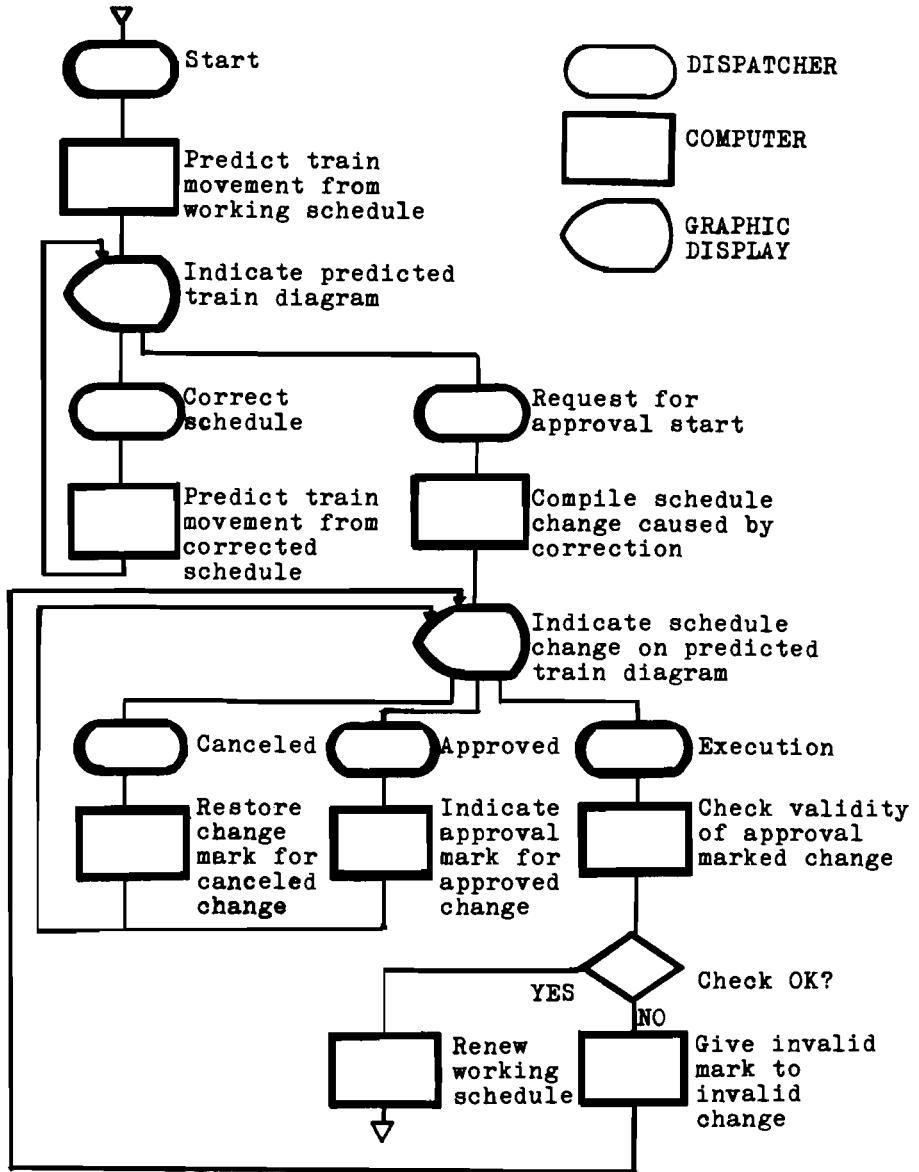


Figure 8. Scheduling system at SNCF [14].



Man-machine dispatching

Program size of COMTRAC [64]:

		Routines	Statements
Route-control	on-line	86	214 700
	off-line	16	124 600
Train dispatching		15	138 400

Figure 9. Scheduling system at JNR [64].



- Linear programming, especially Integer programming [9, 17, 45, 49],
- the Hungarian method [45, 56],
- the Modified distribution method (MODI) [45],
- the Stepping stone [45],
- the Branch and bound [3, 15, 17, 26, 37, 54],
- Dynamic programming [36],
- Permutation methods [4, 5],
- Simulation methods [7, 13, 10, 14, 26, 23] with the use of FORTRAN [19], PL/1 [11], GPSS [12], SIMULA [27, 52], VOPS/SIMDIS [6] and hybrid simulation [54] etc., and
- Methods for optimization of stochastic systems [21, 28, 34, 40, 43].

As long as pure mathematical algorithms are applied, the matrix representation [45, 46] is preferred. With more insight into data manipulation by computers, more suitable representation will be wanted. Considering that timetable calculation is usually done only to some specified depth, and because it is closely connected to graph representation [44, 38, 57, 59], list representation seems suitable [22, 48, 11] (see Figures 10 and 11). Very early on active memories were also considered [30, 31, 54].

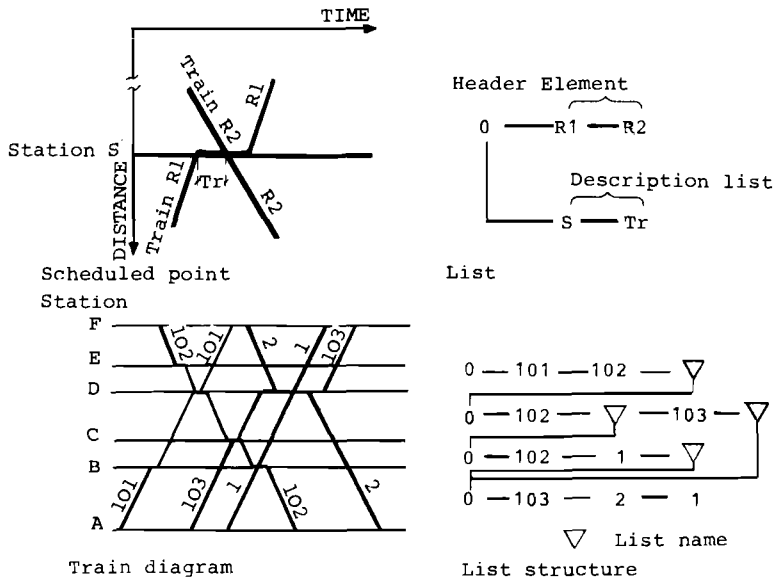
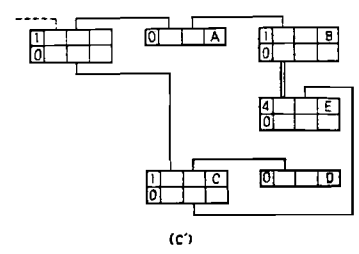
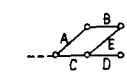
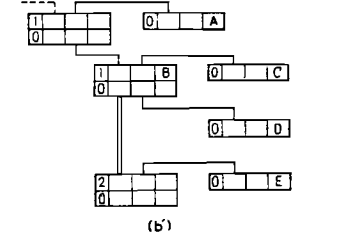
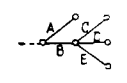
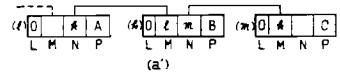
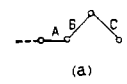
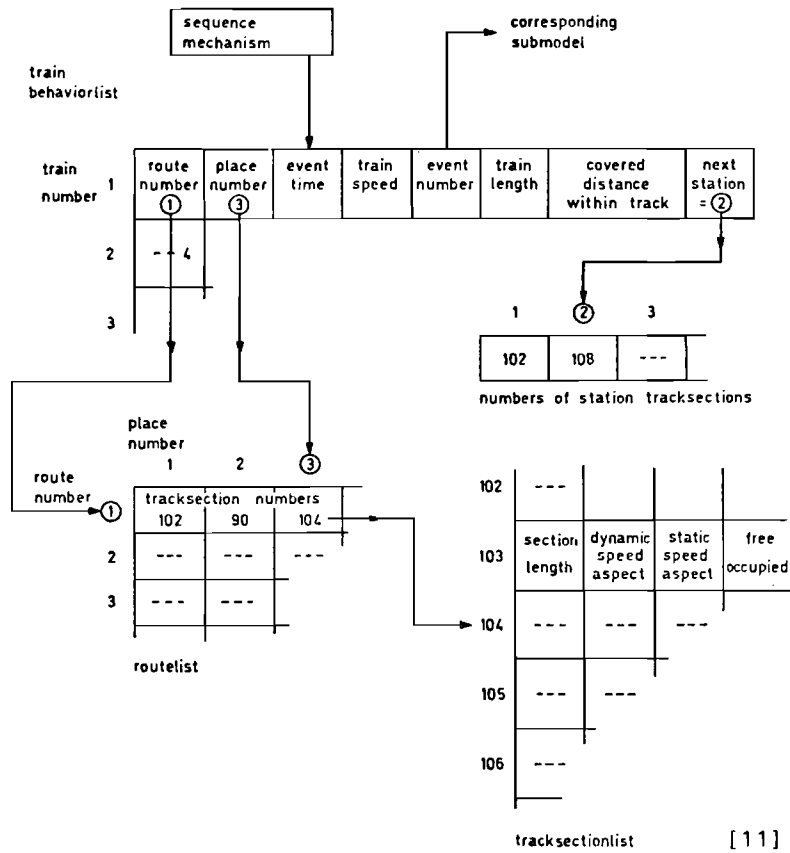


Figure 10. Timetable calculation and list representation [22].



Line segments represented by LIST [48]

Figure 11. List representation.

The decreasing prices of hardware, and the development of new technologies, make it possible to realize associative memories. But also, further development of theory, e.g. colored graphs [38, 57, 59] and the application of pattern recognition and self learning systems [16, 50], will make the theory and automation more practicable (Figure 12). In the case of pattern recognition, not all states need to be stored, because only after the failing of a simple algorithm used does the grid of experience have to be extended. (For a further description of the conditions for timetable optimization and scheduling see [50, 51, 60].)

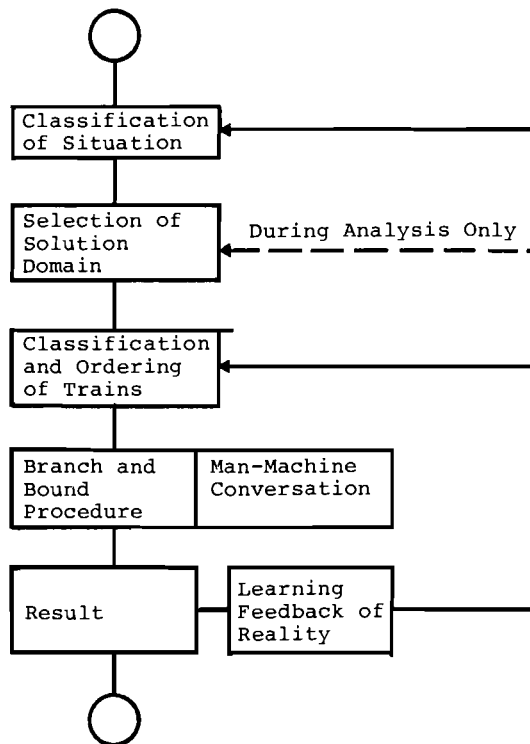


Figure 12. Scheduling procedure with self learning [16].

Besides general timetable problems, the same methods are used for investigating line capacities [8, 10, 13, 15, 40, 43, 44, 47, 65], and for studying the influence of disturbances and the effect or sensitivity of timetables to traffic demand and vice versa [2, 8, 12, 1].

But what is gained by improving scheduling methods? Reliability can be improved, because of man's limited ability to process the quantities of information arising in disturbances or changes and because of his relatively high error rate. Under normal conditions, improvements will generally not be very high, because man has had time to learn in the long life of the railway system. Figure 13 should show the differences of time that can be gained between the worst and best scheduling of trains in a simple case [8].

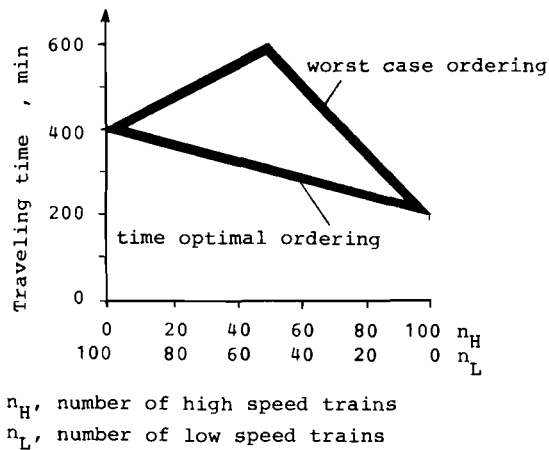


Figure 13. Sequencing of trains with two speed classes [8].

References for General Remarks and Scheduling

- [1] Ajzen, A., and L. Rokeach, Routenwahl in Liniennetzen, *Zeitschrift für Operations Research (Würzburg)*, 18 (1974), B 101 - B 120.
- [2] Annis, A.J., and G.D. Brook, A Real-Time Computer System to Aid Train Regulations at Glasgow Central Station, Paper given at 4th ORE colloquium on Technical Computer Programs, München, 1974.

- [3] Aubert, M., and A. Lemaire, Regulation des circulations en gare de Paris-Nord. Paper given at IFAC/IFIP/IFORS Symposium on Traffic Control and Transportation Systems, Monte Carlo, 1974.
- [4] Auclair, J.P., Débit des lignes équipées en block automatique et exploitées par batteries, Paper given at 1st ORE colloquium on Technical Computer Programs, Derby, 1969.
- [5] Auclair, J.P., Etudes d'implantation et d'exploitation d'installation permanents de centre-sens, Paper given at 4th ORE colloquium on Technical Computer Programs, München, 1974.
- [6] Bechtel, J., et al., Anwendung des VOPS/SIMIDIS auf Verkehrsabläufe in Eisenbahnsystemen, *Wiss. Zeitschrift HFV (Dresden)*, 21, 4 (1974), 620 - 628.
- [7] Beseler, E., Literaturauswahl zur Simulation des Eisenbahnbetriebes unter besonderer Berücksichtigung des Zugbetriebes einer Strecke, *Wiss. Zeitschrift HFV (Dresden)*, 20, 4 (1973), 865 - 870.
- [8] Blind, W., Abhängigkeit der Leistungsfähigkeit und des Verspätungsabbaues auf Engpaßstrecken von der Aufeinanderfolge der Züge verschiedener Geschwindigkeit, *Archiv f. Eisenbahntechnik (Darmstadt)*, 26, Nov. (1971), 45 - 70.
- [9] Bradley, G.H., and P.N. Wahi, An Algorithm for Integer Linear Programming: A Combined Algebraic and Enumeration Approach, *Operat. Research (Baltimore)*, 21, 1 (1973), 45 - 60.
- [10] Brettmann, E., Einsatz eines Digitalrechners zur Konstruktion eines starren Fahrplanes auf eingleisiger Strecke, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 5, 10 (1968), 402 - 422.
- [11] Breur, M.W.K.A., Entwicklung und Anwendung eines Simulationsmodells für Metro-Betriebe, *Schienen der Welt*, 4, 8 (1973), 897 - 907.
- [12] Carsten, J.P., et al., Economic Models for Rail Systems, *IEEE Trans.*, SSC-2, 2 (1966), 128 - 134.
- [13] Cooper, C.P., and A.G. Noordam, Assessing the Capacity of a Railway Line by Simulating Traffic, Paper given at 1st ORE colloquium on Computer Programs, Derby, 1969.
- [14] Drayer, R., L'automatisation dans la confection des horaires, *Rev. Gen. des Chemins de fer*, 94, 6 (1975), 345 - 349.

- [15] Gatez, J.L., Détermination de l'horaire le plus dense sur une ligne quelconque, Paper given at 4th ORE colloquium on Technical Computer Programs, München, 1974.
- [16] Genser, R., *Entwurf eines Verfahrens für die Zug-Verkehrssteuerung*, Department of Applied Mathematics and Physics, Kyoto University, Kyoto, 1971.
- [17] Gondran, M., and J.L. Laurière, Un algorithme pour les problèmes de recouvrement, *Rev. Française d'Inform. et de Rech. Opérat.*, 9, V-2 (1975), 33 - 51.
- [18] Guyer, W., *Optimale Fahrplangestaltung im Reisezugverkehr*, Dissertation der ETH, Zürich, 1969.
- [19] Hawkes, J., General Area Time-based Train Simulator (GATTS), Paper given at 2nd ORE colloquium on Technical Computer Programs, Derby, 1972.
- [20] Hobara, T., et al., System Design and Simulation Test of COMTRAC System for New Lines, Paper given at 1st IFAC/IFIP Symposium on Traffic Control, Versailles, 1970.
- [21] Houpt, P.K., *Decentralized Stochastic Control of Finite-State Systems with Application to Vehicular Traffic Control*, Doctor Thesis, M.I.T. Electronic Systems Lab., Cambridge, Mass., 1974.
- [22] Iida, Y., Preparation of Train Diagram with a Computer, *Quarterly Reports JNR (Tokyo)*, 12, 3 (1971), 158-165.
- [23] Iida, Y., Railway Traffic Planning System, Paper given at 5th IFAC congress, Boston, 1975.
- [24] Inada, N., and Y. Iida, A Program System to Prepare Train Diagram for Operation Planning, Paper given at 1st IFAC/IFIP symposium on traffic control, Versailles, 1970.
- [25] Inada, N., et al., Betriebsleitung und Steuerungssystem der Shinkansen-Linie, *Schienen der Welt*, 6, 5 (1975), 361 - 376.
- [26] Kosemund, M., Zur Simulation eines prozeßrechnergesteuerten Zugbetriebes, *Wiss. Zeitschrift HFV (Dresden)*, 18, 4 (1971), 805 - 815.
- [27] König, H., and S. Stähli, Simulation des Betriebsablaufes an kleinen Eisenbahnknoten, *Schienen der Welt*, 2, 12 (1971), 1015 - 1021
- [28] Kuo, B.C., and D.W. Peterson, Optimal Discretization of Continuous-Data Control System, *Automatica*, 9, 1 (1973), 125 - 129.

- [29] Lazak, D., A Heuristic Approach to Algorithms Intended for the Solution of Timetable-Problem, *Computing (Wien)*, 4, 4 (1969), 359 - 377.
- [30] Lemaire, A., and J. Floquet, Kybernetische Systeme, mit denen die Leistungen der Strecke und Knotenpunkte erhöht und flüssiger gestaltet werden können, *Schienen der Welt*, 1, 9 (1970), 642 - 647.
- [31] Lemaire, A., Le simulateur électronique dénommé "mémoire active", *Rev. Gen. de Chemins de fer*, 90, 10 (1971), 608 - 618.
- [32] Levitt, K.N., and W.H. Kautz, Cellular Arrays for the Solution of Graph Problems, *Comm. of the ACM*, 15, 9 (1972), 789 - 801.
- [33] Mallinson, H. Systemanalysen für die Anwendung von Computern bei der Eisenbahnplanung, *Schienen der Welt*, 1, 6 (1970), 419 - 423.
- [34] Melzer, S.M., and B.C. Kuo, Optimal Regulation of Systems Described by a Countably Infinite Number of Objects, *Automatica*, 7, 3 (1971), 359 - 366.
- [35] Merle, G., Näherungsverfahren zur Lösung eines Transportproblemes mit nichtlinearen Transportkoeffizienten, *Angewandte Informatik*, 9 (1972), 430 - 432.
- [36] Morlok, E.K., and H.L. Vandersypen, Schedule Planning and Timetable Construction for Commuter Railroad Operations, *Transportation Eng. Journal (New York)*, 99, 3 (1973), 627 - 636.
- [37] Michel, P., et al., Fluidification par ordinateur du trafic d'un noeud ferroviaire, *Automatisme*, 15, 5 (1970), 203 - 209.
- [38] Neufeld, G.A., and J. Tartar, Graph Coloring Conditions for the Existence of Solutions to the Timetable Problem, *Comm. of the ACM*, 17, 8 (1974), 450 - 453.
- [39] Ozeki, M., et al., Railway Traffic Control System, Paper given at IFIP congress, Stockholm, 1974.
- [40] Prahl, J., Ermittlung der Leistungsfähigkeit komplexer Fahrstraßenknoten, *Eisenbahntechn. Rundschau*, 22, 6 (1973), 245 - 250.
- [41] Petrov, A.P., Einsatz von Elektronenrechnern, Datenübertragungseinrichtungen und kybernetischen Verfahren für die Lenkung des Güterverkehrs, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 3, 5 (1966), 1207 - 1255.

- [42] Plietz, K.-D., et al., Anwendungsmöglichkeiten der EDV für die Zugfahrplankonstruktion, *Die Eisenbahntechnik (Berlin)*, 23, 11 (1975), 492 - 494.
- [43] Potthoff, G., *Verkehrsstömungslehre*, Transpress VEB Verl. Verkehrswesen, Berlin, 1965
- [44] Potthoff, G., Fahrstraßenverträglichkeit, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 2, 3 (1965), 151 - 153.
- [45] Richter, K.-J., *Ökonomisch-mathematische Entscheidungsmodelle im Verkehrswesen*, Transpress VEB Verl. Verkehrswesen, Berlin, 1972.
- [46] Richter, K.-J., *Die Transportmatrix*, Transpress VEB Verl. Verkehrswesen, Berlin, 1974.
- [47] Roth, P., Aufstellung von Betriebsfahrplänen für zeitweise eingleisigen Betrieb. Paper given at 4th ORE colloquium on Technical Computer Programs, München, 1974.
- [48] Sakai, T., and M. Nagao, Storing and Processing Traffic Maps in Digital Computer (Jap.), *The Journ. of the Inst. of Electr. Co., Eng. of Japan*, 50, 4 (1967), 629 - 636.
- [49] Savage, M.J., Junction Optimization Technique, *The Computer Journal*, 12 (1969), 268 - 272.
- [50] Schlegel, D., Mathematische Verfahren bei der operativen Betriebsplanung der Eisenbahnen, *Die Eisenbahntechnik (Berlin)*, 22, 8 (1974), 367 - 368.
- [51] Steinfeld, H., and W. Geber, Bezirkssteuerzentrale für den Raum Saarbrücken, *Eisenbahntechnische Rundschau*, 21, 1/2 (1972), 31 - 40.
- [52] Sitzmann, E., Probleme der Betriebsführung, Fahrplangestaltung und betrieblichen Organisationen zur Bewältigung der Transportaufgaben im Ballungsverkehr, *Schienen der Welt*, 4, 2 (1973), 235 - 249.
- [53] Stähli, S., Vergleich von Entscheidungsalgorithmen zur Lösung betrieblicher Konfliktsituationen mittels Simulation, Paper given at 4th ORE colloquium on Technical Computer Programs, München, 1974.
- [54] Tessier, M., L'informatique au service de la recherche, *Rev. Gen. des Chemins de fer*, 91, 11 (1972), 730 - 734.
- [55] Thiele, E., Ein Simulationsmodell für die Erstellung nachfragegesteuerter Wagenlaufpläne einer U-Bahn-Linie, *Verkehr und Technik (Bielefeld)*, (1974), 1, 13 - 16, 25 - 26, 28 - 30.



- [56] Uebe, G., *Optimale Fahrpläne*, Springer-Verlag, Berlin, 1970.
- [57] Wang, Ch.C., An Algorithm for the Chromatic Number of a Graph, *Journal of the ACM*, 21, 3 (1974), 385 - 391.
- [58] Weinberg, F., and C.A. Zehnder, *Heuristische Planungsmethoden*, Springer-Verlag, Berlin, 1969.
- [59] Werra, D.de, Some Results in Chromatic Scheduling, *Zeitschrift f. Operations Research (Würzburg)*, 18 (1974), 167 - 175.
- [60] Wunderlich, W., and E. Kockelkorn, Bessere Disposition durch schnellere Information, *Eisenbahntechnische Rundschau*, 24, 7/8 (1975), 283 - 294.
- [61] Yaged, B., Traffic Flow Information for Minimum Cost Routing Procedures, *Automatica*, 5, 2 (1969), 167 - 173.
- [62] Yamada, H., Aus acht Untersystemen bestehendes einheitliches Zugbetriebssystem, *Schienen der Welt*, 1, 7/8 (1970), 512 - 517.
- [63] Yamamoto, I., Analytical Methods for the Train-Traffic Control Problem, Paper given at 1st IFAC/IFIP Symposium on Traffic Control, Versailles, 1970.
- [64] Yamamoto, I., et al., Basic Study on COMTRAC System for OKAYAMA Extension of SHIN-KANSEN, *Quarterly Reports JNR (Tokyo)*, 15, 3 (1974), 144 - 155.
- [65] Zavjalov, B.A., Ispytanija sistemy "ucastkov yi avto-dispetcer" v aytomaticaskoj, *Vestnik VNII ZT*, 30, 2 (1971), 60 - 62.

#### ROLLING STOCK ROSTERING

The vast investment needed for freight cars because of the long time interval between use and reuse, and the capacity problem of crowded marshalling yards, stimulated railway management all over the world to improve the rostering problem of empty freight cars and rolling stock at a very early stage by means of electronic data processing (EDP) [1, 2, 5, 17, 20, 34, 35, 7, 8, 9, 10, 11, 12, 13, 28, 18, 19, 22, 24, 29, 30, 37].

The algorithms used are similar to those for the scheduling problem to some extent:

Linear programming [4, 7, 16, 18, 19, 27, 29, 37],  
Out of kilter [12],  
Algorithm of Dijkstra [4],  
Ford-Fulkerson [35],

Branch and bound [3, 32],  
Dynamic programming [4, 15],  
Heuristics [1, 13], etc.,

and Figure 14 gives a pictorial representation of the problem.

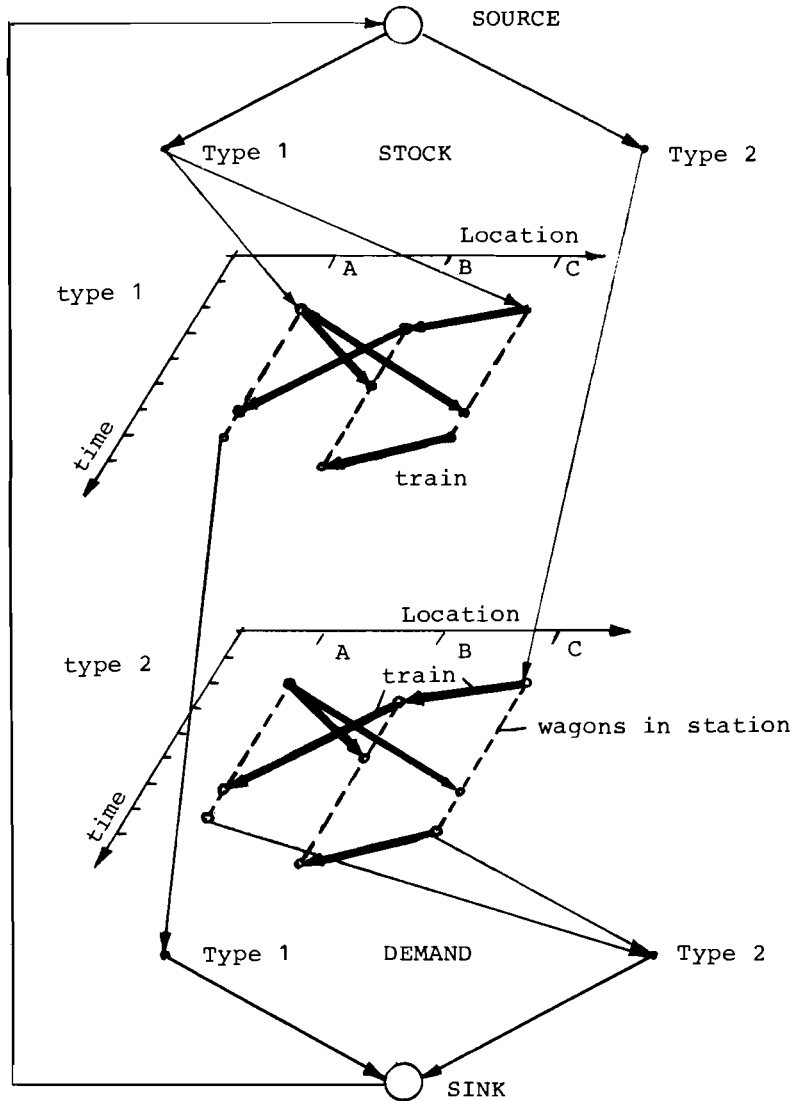


Figure 14. Distribution of two types of freight cars [10].

The planning of shunting in marshalling yards, and the composition of freight trains, can be stated as a subproblem [3, 15, 21, 23, 25, 26, 27, 28, 33]. Figure 15 gives an example of this approach [28].

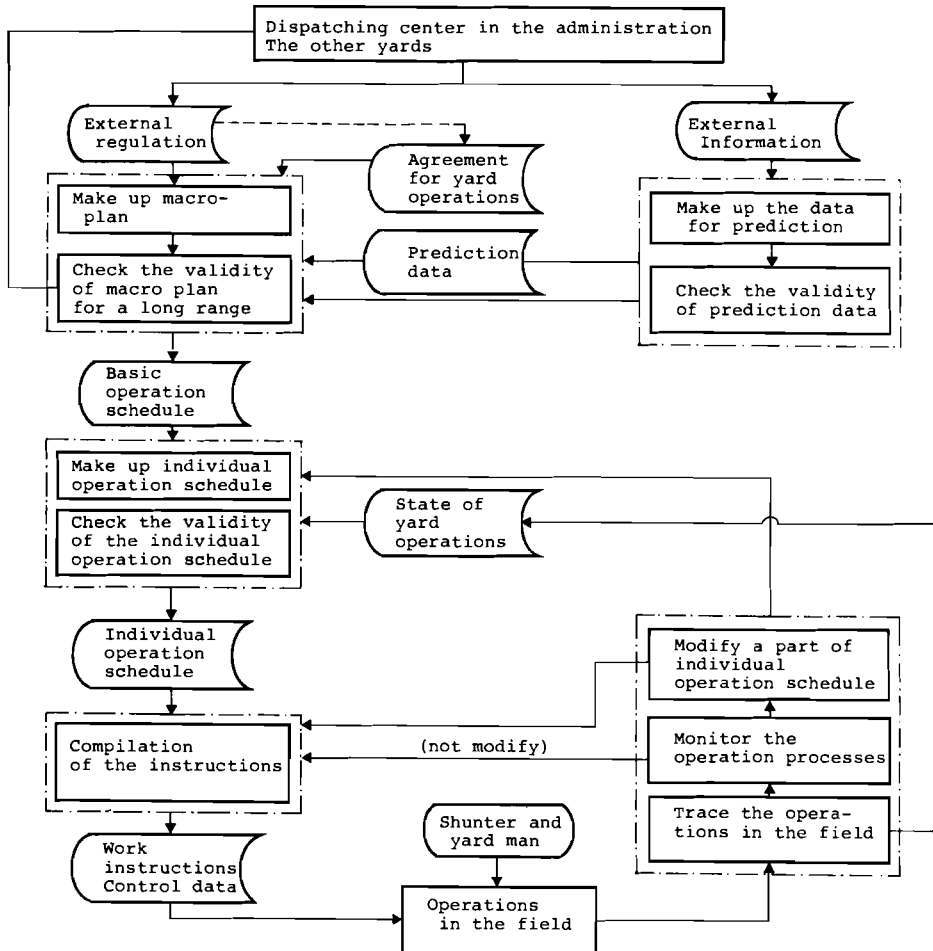


Figure 15. Planning system for marshalling yards at JNR [28].

The solutions for freight cars rostering can be adapted for solving the problem of passenger cars allocation, of course [4, 6, 14, 31]. Passenger cars are not very sensitive to traffic demand, however, because of low pass filtering behavior. This is why some railway managements are interested in container shipment, because if freight trains could be operated like a conveyor belt, sensitivity to traffic demand would be reduced.

References for Rolling Stock Rostering

- [1] Anonym, *Etude UIC-METRA pour la gestion centralisée du trafic marchandises (1967 - 1972)*, Etude UIC-METRA, Paris, 1973.
- [2] Anonym, *Modell "Operativer Leerwagenausgleich"*, Zentrales Forschungsinstitut des Verkehrswesens der DDR, Berlin, 1974.
- [3] Achermann, J., *Modell und Rechenprogramm zur optimalen Bildung von Ferngüterzügen*, I.E.K.V.: *Kybernetik und Elektronik bei den Eisenbahnen*, 6, 5 (1969), 187 - 231.
- [4] Ajzen H., et al., *Grundlagen für die Erstellung nachfragegesteuerter Wagenumlaufpläne*, *Nahverkehrspraxis*, 20, 5 (1973), 178 - 185, 6 (1973), 248 - 252.
- [5] Bomberault, A.M., and W.W. White, *Networks and Transportation*, IBM Report No. 320-2952, IBM, New York, 1968.
- [6] Brettmann, E., *Umlaufpläne für Sonderzüge im Ferien- und Festverkehr*, *Die Bundesbahn (Darmstadt)*, 18, 6 (1971), 285 - 290.
- [7] Christof, C.A., *Operativer Leerwagenausgleich mit Hilfe von EDVA bei den Bulgarischen Staatsbahnen*, *Zeitschrift OSShD (Warschau)*, 15, 3/4 (1972), 15 - 18.
- [8] Dalaban, N., *Anwendung der elektronischen Datenverarbeitung für die Planung des Gütertransportes bei den CFR*, *Zeitschrift OSShD (Warschau)*, 15, 1/2 (1972).
- [9] Frischknecht, J., *Der Systemansatz beim Repartierungsproblem von Güterwagen*, Diplomarbeit Hochschule St. Gallen, St. Gallen, Schweiz, 1974.
- [10] Gigon, J.-M., and F.S. Hiaepfer, *Die Leerwagenverteilung auf einer elektronischen Rechenanlage*, I.E.K.V.: *Kybernetik und Elektronik bei den Eisenbahnen*, 4, 7 (1967), 303 - 324.
- [11] Gottfried, G., *Die Verteilung leerer Transporteinheiten im Netz der Eisenbahnen*, I.E.K.V.: *Kybernetik und Elektronik bei den Eisenbahnen*, 5, 11 (1968), 423 - 441.
- [12] Herren, H., *Die Leerwagenverteilung mit EDV*, *Schienen der Welt*, 4, 10 (1973), 1039 - 1044.
- [13] Iida, Y., and S. Koga, *Study of Rolling Stock Rostering*, Paper given at 2nd IFAC/IFIP/IFORS symposium on Traffic Control and Transportation Systems, Monte Carlo, 1974.

- [14] Jakubski, S., Fahr- und Wagenlaufplanung für den städtischen Nahverkehr als mathematische Optimierungsaufgabe, *DDR-Verkehr (Berlin)*, 7, 8 (1974), 330 - 335.
- [15] König, H., and P. Schaltegger, Optimale Simultanformation von Nahgüterzügen in Rangierbahnhöfen, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 4, 1 (1967), 1- 18.
- [16] Kutukova, G.A., and L.K. Markow, Zugbildungsplan für ein Eisenbahnnetz, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 1, 5 (1964), 183 - 191.
- [17] Leddon, C.D., and E. Wrathall, Programmierung der Leerwagenverteilung der Eisenbahn von Louisville und Nashville, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 5, 1 (1968), 23 - 27.
- [18] Martens, A., Mathematische Modelle zur Lösung des Problems der Beförderung der Güterwagen zwischen den Zugbildungsbahnhöfen, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 4, 4 (1967), 151 - 162.
- [19] Misra, S.C., Leerwagenverteilung mit Verfahren der linearen Optimierung, *Schienen der Welt*, 3, 4 (1972), 283 - 291.
- [20] Norman, A.R.D., and M.J. Dowling, *Railroad Car Inventory*, IBM Report No. 320-2926, IBM, New York, 1968.
- [21] Notari, M., Die Anwendung der Kybernetik und Elektronik bei den Italienischen Staatsbahnen auf das Problem der zentralisierten Lenkung des Güterverkehrs, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 3, 5 (1966), 209 - 269,
- [22] Olejniczak, R., Ausarbeitung eines technologischen und mathematischen Modells für die Leitung des Transportprozesses bei den PKP, *Zeitschrift OSShD (Warschau)*, 16, 3 (1973), 1 - 6.
- [23] Plietz, K.-D., Anwendung der Graphentheorie am Beispiel der Güterzugsbildungsplanung, *Deutsche Eisenbahntechnik (Berlin)*, 20, 11 (1972), 518.
- [24] Revol, B., and M. Gochet, Optimierung der Lenkung des Einzelgüterwagenverkehrs Studie UIC-METRA, *Schienen der Welt*, 4, 9 (1973), 939 - 946.
- [25] Römmermann, K.-H., Theorie und Praxis der Berechnung des Güterzugsbildungsplans in der Sowjetunion, *Wiss. Zeitschrift HFV (Dresden)*, 18, 2 (1971), 295 - 306.
- [26] Römmermann, K.-H., Ein Beitrag zur Berechnung von Güterzugsbildungsplänen, *Wiss. Zeitschrift HFV (Dresden)*, 20, 4 (1973), 841 - 847.

- [27] Rusnak, O., Das Bilden von Güterzügen durch ganzzahlige Programmierung, *Deutsche Eisenbahntechnik (Berlin)*, 19, 11 (1971), 520 - 522.
- [28] Sato, A., et al., Information Processing System of Marshaling Yard, *Quarterly Reports JNR (Tokyo)*, 16, 2 (1975), 79 - 87.
- [29] Truskolaski, A., and W. Grabowski, Problem der Wagenverteilung, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 1, 5 (1964), 192 - 197.
- [30] Truskolaski, A., Anwendung von Digitalrechnern zur Optimierung von Güterzugbildungsplänen, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 5, 5 (1968), 231 - 248.
- [31] Waddell, M.C., et al., *Disposition of Empty Vehicles in a Personal Rapid Transit System*, No. UMTA-MD-06-0018-71-1, US-Department of Commerce - NTIS, Springfield, 1974.
- [32] Weinberg, F., *Einführung in die Methode Branch and Bound*, Springer-Verlag, Berlin, 1968.
- [33] Wöhner, W., Güterzugbildung als kombinatorische Optimierungsaufgabe, *Wiss. Zeitschrift HFV (Dresden)*, 20, 4 (1973), 831 - 840.
- [34] White, W.W., and A.M. Bomberault, A Network Algorithm for Empty Freight Car Allocation, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 6, 11 (1969), 488 - 510.
- [35] White, W.W., and E. Wrathall, *A System for Railroad Traffic Scheduling*, IBM Report No. 320-2993, IBM, Philadelphia, 1970.
- [36] Wunderlich, W.M., and R. Wiedenbein, Die Simulation betrieblicher Vorgänge in Zugbildungsbahnhöfen am Beispiel eines Rangierbahnhofs, *Schienen der Welt*, 3, 5 (1972), 310 - 322.
- [37] Zahariev, E., and S. Ovcharov, Ziele, Methoden und Möglichkeiten der langfristigen, mittelfristigen und betrieblichen Lenkung des Eisenbahn-Güterverkehrs, *Schienen der Welt*, 5, 1 (1974), 9 - 12.

### TRACTIVE UNIT ALLOCATION AND MANPOWER PLANNING

Generally, tractive unit allocation is separated from the car rostering problem [1, 2, 6, 7, 3, 4, 9, 10, 11, 12, 13, 15]. But for tractive unit allocation as well, a man-machine conversation will be convenient for general use.

Manpower planning is considered to be closely related to tractive unit allocation [5, 8, 9, 14], but, because of more difficult boundary conditions, in most of the cases the solutions were not satisfactory. Manpower planning is more restrictive in railway systems than in bus systems. Safety and reliability standards, and the extension of railway line configurations, are the reason.

### References for Tractive Units Allocation and Manpower Planning

- [1] Carencio, G., Experimentation d'une automatisaion de l'equipement des trains en engins moteurs, Paper given at 4th ORE colloquium on Technical Computer Programs, München, 1974.
- [2] Csikos, M., Aufstellung des Lokumlaufplans mit Hilfe der elektronischen Datenverarbeitung, *Zeitschrift OSShD (Warschau)*, 16, 4 (1973), 9 - 11.
- [3] Genser, R., *Entwurf eines Verfahrens zur Ermittlung der Triebfahrzeugumlaufpläne*, Interne Studie OBB/EDV, Wien, 1972.
- [4] Glück, H., Die Optimierung von Triebfahrzeugumlaufplänen bei großen Zugzahlen mit Hilfe elektronischer Datenverarbeitungsanlagen, *Eisenbahntechnische Rundschau*, 22, 9 (1973), 354 - 363.
- [5] Greven, G., and A. Ter Marten, Beschreibung eines Automatisierungsprojektes für das Aufstellen von Diensten und Dienstplänen für Zugschaffner bei der Niederländischen Eisenbahn, *Schienen der Welt*, 3, 6 (1972), 375 - 387.
- [6] Hein, O., Ein Näherungsverfahren zur Ermittlung von Fahrzeugumlaufplänen, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 6, 11 (1969), 457 - 470.
- [7] Holt, J., Erstellung von Triebfahrzeugumlaufplänen mit Computer - "BASHPEAK", *Schienen der Welt*, 4, 10 (1973), 1090 - 1093.
- [8] Konya, H., et al., A Study on Computer-Aided Crew Scheduling, *Quarterly Reports JNR (Tokyo)*, 122, 2 (1971), 91 - 97.

- [9] Kregeloh, H., et al., Bildung von Fahr-, Umlauf- und Dienstplänen nach heuristischen Verfahren mit Hilfe der EDV. *UITP Revue (Brüssel)*, 24, 1 (1975), 87 - 92.
- [10] McGaughey, R.S., et al., Das Planen der Verteilung von Lokomotiven und Packwagen, *Schienen der Welt*, 4, 11/12 (1973), 1269 - 1275.
- [11] Passow, H.-J., Zur Lösung von Umlaufproblemen, *Rechentechnik/Datenverarbeitung (Berlin)*, 10, 2 (1973), 32 - 36.
- [12] Potthoff, G., Der optimale Lokeinsatz, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 4, 3 (1967), 116 - 125.
- [13] Surmont, J., Die Unternehmensforschung ermöglicht eine Optimierung der Umläufe, *I.E.K.V.: Kybernetik und Elektronik bei den Eisenbahnen*, 2, 8/9 (1965), 359 - 371.
- [14] Unger, C., Ein Programmsystem zur maschinellen Erstellung von Dienstplänen und Dienstreihenfolgenplänen, *UITP Revue (Brüssel)*, 24, 1 (1975), 68 - 72.
- [15] Zalewski, R., Optimierung der Triebfahrzeugumläufe mit Hilfe mathematischer Methoden und EDVA, *Zeitschrift OSShD (Warschau)*, 14, 1 (1975), 17 - 18.

#### MAINTENANCE PLANNING

In many production plants, project management planning and job scheduling are carried out by computers with well known software packages [4, 5, 6, 7, 8]. The goals are:

- the shortest time spent for the job in the workshop,
- an accurate delivery date, and
- the optimal use of available capacity.

Even if the breakdown of a car could be predicted, private repair shops would not know when a customer would arrive with his car. The arrivals of cars for repair can only be forecast by analyzing past events. Railways can use information on the behavior of equipment to formulate suitable maintenance plans to control throughput in railway workshops. The goal of optimal workshop control has to be the smallest investment for workshop and rolling stock, and the optimal strategy is [42]:

- to have an information system with highly efficient data processing for the most reliable railway system [1, 3, 4] - there should not be the slightest objective of controlling the efficiency of man;



- to use a suitable maintenance plan; and
- to analyze deviations from the plan with the purpose of improving the system.

Of course, common software packages like PERT, etc. are also used in railways in track construction, etc., or in sub-systems, but an automated on-line control in workshops, just smooths over bad states. The real trouble spots, for example an unsuitable organization, will not be investigated. Man only produces data and cannot recognize the effect. He feels he is being controlled by a machine: this is not at all stimulating.

#### References for Maintenance Planning

- [1] Genser, R., and A. Grassl, State of Reliability Effort of European Railways, *IEEE Trans.*, R-23, 3 (1974), 156 - 161.
- [2] Genser, R., *Auftragsablaufplanung und Steuerung bei den Hauptwerkstätten der ÖBB*, Interne Diskussionsunterlage ÖBB/EDV, Wien, 1973.
- [3] Harada, R., Control of Rolling Stock Maintenance, *Japanese Railway Engineering*, 10, 4 (1969), 8 - 11.
- [4] Ishii, K., and H. Toko, Outline of SHINKANSEN HAKATA Rolling Stock Maintenance Base, *Japanese Railway Engineering*, 15, 3/4 (1975), 15 - 19.
- [5] Lübke, D., Die Anwendung der Netzplantechnik bei der Pflege und Unterhaltung moderner elektrischer Triebfahrzeuge in den Bahnbetriebswerken, *Elektr. Bahnen (München)*, 42, 3 (1971), 50 - 55.
- [6] White, J.C., and E.F. Paul, Ein Rechner für die Fertigungskontrolle in Eisenbahnwerkstätten, Paper given at 3rd Int. Symp. Kybernetik im Eisenbahnwesen, Tokyo, 1970, 219 - 222.
- [7] Wöhl, G., and W. Kandler, Netzwerkgesteuerte Produktion bei der Instandhaltung von Diesellokomotiven, *Schiene-fahrzeug (Berlin)*, 17, 8 (1973), 259 - 262.
- [8] Zirn, A., Angewandte Kybernetik bei der Planung und Steuerung im Arbeitsbereich eines Bundesbahn-Ausbesserungswerkes, *Archiv für Eisenbahntechnik*, 28, Dezember (1973), 28 - 50.

CONCLUSION

What can be done to improve the present state?

- Information processing has to be improved. This includes data collection, transmission, filtering, identification, and pattern recognition. A good example would be in the approach to electrical power distribution systems.
- New technologies have to be used for man-machine interaction. Cathode ray tubes have not been satisfactory.
- Self-learning systems should be implemented, and pattern recognition should be used for considering new technologies.

It must be recognized that this is a task for engineers! Simple algorithms should be preferred that give insight into the problem and that present a collection of solutions rather than just one abstract value.

But the most important task will be to educate and inform the management of less advanced railways that present day possibilities are used for the advantage of the passengers, the community, the government, and last but not least the operators.

## Optimal Planning of Shipment Volumes

O.I. Aven, S.E. Lovetsky, and G.E. Moiseenko

In terms of the number of ships and total tonnage, the Soviet merchant fleet is one of the largest in the world. In a year, Soviet ships visit more than 1000 ports of 105 countries (Voronkov and Klementiev, 1972).

The large distances involved, the number of ships and ports, and the distinctly probabilistic nature of operations, which may be disturbed by wars, strikes, random cargo and tonnage in the freight market, and many other factors, make the merchant fleet an industry quite unlike all others. Planning for such a large and complicated system necessitates the receipt and processing of a great amount of information and is made especially complicated by the fact that the industry operates both in countries with socialist planned economies and in the free market. The Merchant Marine Ministry charters Soviet ships out and freights foreign ships.

Performance requirements of transportation increase with the number of ships, cargo carrying capacity and scope of the fleet activities. To improve the effectiveness of expanding fleet operations, a Management Information System (MIS) for the merchant fleet has now been developed, where computers, and economic and mathematical methods are widely used to solve planning and management problems.

The most important of various planning problems is transportation, because the quality of a plan as a whole depends above all on how many and in what way the ordered cargoes are allocated to the tonnage available. Determination of the amounts to be transported makes it easy to define all secondary parameters, such as gains, costs, and cargo reloading capacity.

In the Institute of Control Sciences, a set of optimal transportation planning models is developed in compliance with the following major principles. Because planning is carried out at all levels of the management system, the models should simulate all levels of the system's hierarchy (Figure 1). The merchant fleet is a branch of the national economy and its plans should interact with those of the whole national economy. The models should comply with the existing classification of plans into quintennial, annual, quarterly, and monthly. Uncertainty of initial data and specificity of planned indices is higher at a higher level of management.

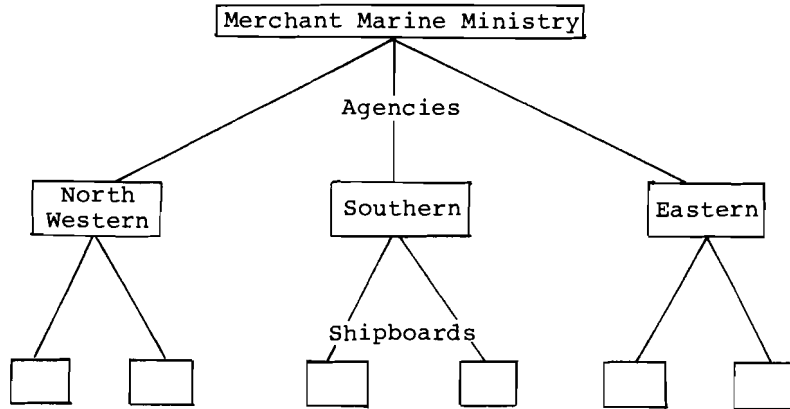


Figure 1.

Consequently, planning should be discrete-time, or over specified time intervals, such as five years, a year, a quarter, or a month; but since the environment is highly variable, dynamic planning should be continuous in that plans should be updated in response to significant changes in the initial data. Thus, monthly schedules are monitored and updated in the existing system every ten days, quarterly plans every month, and annual plans every quarter.

Plans should be coordinated both for the hierarchy levels, in that the plans of lower echelons result from those of higher echelons with the use of more detailed data, and for planning the hierarchy, in that an annual plan is more specific than a quintennial plan: a quarterly plan is more specific than an annual plan.

It is necessary to formulate the problems in the model so that they are both equivalent to actual planning problems and solvable by available computers.

In developing a set of models for annual and quarterly planning, the dimensionality is reduced in two ways - by decomposition of the overall problem into subproblems and by data aggregation. An equivalent system of smaller problems is substituted for the original problem. The subproblems are entrusted to parts of the management systems. The subproblems of different levels differ in the degree of data aggregation. Solutions obtained at different levels are coordinated by iterative updating of parameters for models of each level in compliance with solutions obtained at other levels.

Without limiting the generality of our reasoning, the functioning of annual and quarterly planning models can be studied with two interacting two-level models as an example (Figure 2).

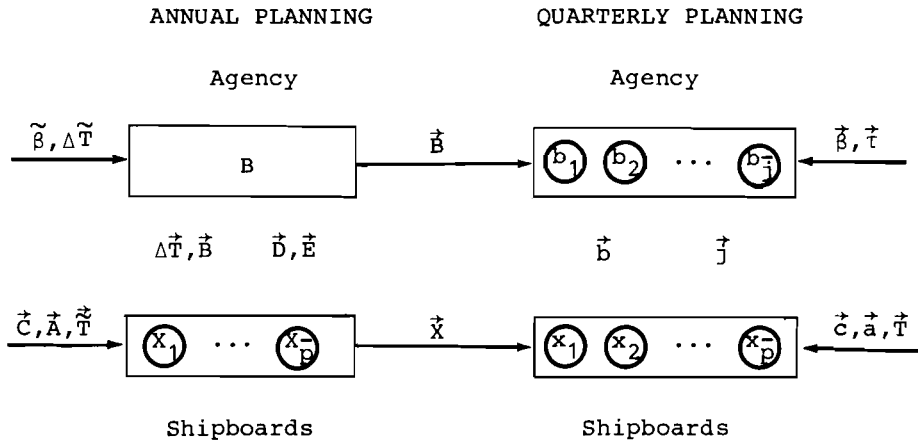


Figure 2. Interaction of the models of annual and quarterly planning.

A specific feature of this particular system of planning models is that plans are specified both for levels of the planning system (problems for the upper and lower levels are identified) and for the planning itself (annual and quarterly planning).

#### MODELS OF ANNUAL PLANNING

The overall problem of annual planning can be written

$$\sum_p \sum_m \sum_t \left( \sum_j C_{pjt}^M x_{pjt}^M - E_{pt}^M \Delta T_{pt}^M \right) \rightarrow \max, \quad (1)$$

$$\sum_p \sum_M x_{pjt}^M \leq \tilde{\beta}_{jt}, \quad j = \overline{1, \bar{j}}, \quad t = \overline{1, \bar{t}}, \quad (2)$$

$$\sum_j A_{pjt}^M x_{pjt}^M + \Delta T_{pt}^M \leq \tilde{T}_{pt}^M, \quad M = \overline{1, \bar{M}}, \quad p = \overline{1, \bar{p}}, \quad t = \overline{1, \bar{t}}, \quad (3)$$

$$\sum_t \Delta T_{pt}^M = \Delta \tilde{T}_p^M, \quad M = \overline{1, \bar{M}}, \quad p = \overline{1, \bar{p}}, \quad (4)$$

$$x_{pjt}^M, \Delta T_{pt}^M \geq 0. \quad (5)$$

For each group,  $j$  ( $j = \overline{1, \bar{j}}$ ), of routes and for each plan period (quarter),  $t$  ( $t = \overline{1, \bar{t}}$ ), the amount of cargo,  $\tilde{\beta}_{jt}$ , to be transported and the profit rates,  $C_{pjt}^M$ , for transportation by ships of the  $M$ th type belonging to the  $p$ th shipboard are assumed to be given. To avoid introduction of one more subscript it may also be assumed without loss of generality that only one kind of cargo is transported along each route. The labor consumptions,  $A_{pjt}^M$ , (specific expenditures of time budget), total time budgets,  $\tilde{T}_{pt}^M$ , (product of the overall carrying capacities of the ships of  $M$ th type by the length of the exploitation period), shares of the overall time budget,  $\Delta T_p^M$ , that are allotted for scheduled repairing of ships, and specific losses,  $E_{pt}^M$ , from non-use of ships during repair are also known.

After decomposition of the problem (1)-(5) with due regard for the management structure, a two-level system was obtained with one problem at the upper level and  $\bar{p}$  problems at the lower:

The problem of the agency (problem B):

$$\sum_j \sum_p \sum_t D_{pjt} B_{pjt} - \sum_p \sum_m \sum_t E_{pt}^M \Delta T_{pt}^M \rightarrow \max, \quad (6)$$

$$\sum_p B_{pjt} \leq \tilde{\beta}_{jt}, \quad j = \overline{1, \bar{j}}, \quad t = \overline{1, \bar{t}}, \quad (7)$$

$$\sum_t \Delta T_{pt}^M = \tilde{\Delta T}_p^M, \quad M = \overline{1, \bar{M}}, \quad p = \overline{1, \bar{p}}, \quad (8)$$

$$B_{pjt}, \Delta T_{pt}^M \geq 0. \quad (9)$$

The problem of the  $p$ th subunit of the lower level (problem  $X_p$ ):

$$\sum_M \sum_j \sum_t C_{pjt}^M X_{pjt}^M \rightarrow \max, \quad (10)$$

$$\sum_M X_{pjt}^M \leq B_{pjt}, \quad j = \overline{1, \bar{j}}, \quad t = \overline{1, \bar{t}}, \quad (11)$$

$$\sum_j A_{pjt}^M X_{pjt}^M \leq \tilde{T}_{pt}^M - \Delta T_{pt}^M, \quad M = \overline{1, \bar{M}}, \quad t = \overline{1, \bar{t}}, \quad (12)$$

$$X_{pjt}^M \geq 0. \quad (13)$$

The agency distributes among the subunits  $p$  ( $p = \overline{1, \bar{p}}$ ) the cargoes,  $\beta_{jt}$ , that should be transported along a group of routes over a  $t$ th period (quarter). The cargoes and time for scheduled repair are distributed so as to maximize the profit minus losses due to the non-use of ships during repair (the profit,  $D_{pjt}$ , from shipping a unit of cargo by the  $p$ th subunit and losses,  $E_{pt}^M$ , are assumed to be known).

At the lower level, the problem is decomposed according to the shipboards to which the ships belong; such decomposition minimizes the number of interrelations between the subproblems.

The subunit  $p$  defines the optimal amounts of transportations with due regard to the amounts of cargoes,  $B_{pjt}$ , and shares of time budget,  $\Delta T_{pt}^M$ , allotted to scheduled repair of ships specified by the higher level unit.

#### MODELS OF QUARTERLY PLANNING

The problem of the  $j$ th agency for the  $t$ th quarter (problem  $b_j$ ):

$$\sum_{\ell \in L_j} \sum_p d_{p\ell t} b_{p\ell t} + \sum_p \left( \psi_{pjt}^+ \Delta_{pjt}^+ - \psi_{pjt}^- \Delta_{pjt}^- \right) \rightarrow \max,$$

$$\sum_p b_{p\ell t} \leq \beta_{\ell t}, \quad \ell \in L_j,$$

$$\sum_{\ell \in L_j \cap L_h} \sum_p b_{p\ell t} \leq \tau_h, \quad h = \overline{1, \bar{h}},$$

$$\sum_{\ell \in L_j} b_{p\ell t} + \Delta_{pjt}^- - \Delta_{pjt}^+ = B_{pjt}, \quad p = \overline{1, \bar{p}},$$

$$b_{p\ell t}, \Delta_{pjt}^\pm \geq 0.$$

The problem of the pth subunit for the tth quarter (problem  $x_p$ ):

$$\sum_m \sum_{\ell} c_{p\ell t}^m x_{p\ell t}^m + \sum_M \sum_j \left( \phi_{pjt}^{M+} \delta_{pjt}^{M+} - \phi_{pjt}^{M-} \delta_{pjt}^{M-} \right) \rightarrow \max ,$$

$$\sum_m x_{p\ell t}^m \leq b_{p\ell t} , \quad \ell = \overline{1, \bar{\ell}} ,$$

$$\sum_{m \in \mu_M} \sum_{\ell} a_{p\ell t}^m x_{p\ell t}^m \leq \sum_{m \in \mu_M} T_{pt}^M - \Delta T_{pt}^M , \quad M = \overline{1, \bar{M}} ,$$

$$\sum_{\ell} a_{p\ell t}^m x_{p\ell t}^m \leq T_{pt} , \quad m \in \mu_M , \quad M = \overline{1, \bar{M}} ,$$

$$\sum_{m \in \mu_M} \sum_{\ell \in L_j} x_{p\ell t}^m + \delta_{pjt}^{M-} - \delta_{pjt}^{M+} = x_{pjt}^M , \quad M = \overline{1, \bar{M}} , \quad j = \overline{1, \bar{j}} ,$$

$$x_{p\ell t}^m, \delta_{pjt}^{M\pm} \geq 0 .$$

Models of quarterly planning differ from those of annual planning in that they include specific directions,  $\ell = \overline{1, \bar{\ell}}$ , a more detailed description of the ship types,  $m \in \mu_M$ , a higher number of constraints (such as the throughput,  $\tau_h$ , of the ports,  $h = \overline{1, \bar{h}}$ ). To coordinate it with an annual plan, the model recognizes that the overall amount of specified cargoes,  $\sum_{\ell \in L_j} b_{p\ell t}$ , and the overall transportation amount,  $\sum_{m \in \mu_M} \sum_{\ell \in L_j} b_{p\ell t}$ , should be equal to the corresponding figures of the annual plan for the upper and lower level,  $B_{pjt}$  and  $x_{pjt}^M$ , plus the expected over implementation,  $\Delta_{pjt}^+, \delta_{pjt}^{M+}$ , or under implementation,  $\Delta_{pjt}^-, \delta_{pjt}^{M-}$ , of the plan. To provide an incentive for meeting the planned targets, the criteria of the problems should include, in addition to profits, the bonuses,  $\{\psi_{pjt}^+ \Delta_{pjt}^+\}$ ,  $\{\phi_{pjt}^{M+} \delta_{pjt}^{M+}\}$ , for over implementation and costs,



$\{\psi_{pjt}^- \Delta_{pjt}^-\}$ ,  $\{\phi_{pjt}^{M-} \delta_{pjt}^{M-}\}$ , for underloading the ships.

## FUNCTIONING OF THE SYSTEM OF MODELS

### Annual Planning

Cargoes and time for repairing are initially distributed by higher level units in compliance with statistical and forecast data on profits,  $D_{pjt}$ , and losses,  $E_{pt}^M$ .

At each iteration,  $k$  subunits of the lower level find such transportation amounts,  $\hat{X}_{pjt}^M(k)$  (for cargoes specified by the upper level), that maximal profit is obtained with the constraints (11)-(13) observed. The profits,  $\hat{D}_{pjt}(k)$ , and losses,  $E_{pt}^M(k)$ , are found simultaneously, and may be dual estimates of the constraints (11) and (12) or may be computed as in Aven, et al. (1974).

Because planning is done with incomplete data the decision made at each iteration is, generally speaking, non-optimal; therefore the iterative search for an optimum may oscillate widely. The process is stabilized by damping the changes in planned targets so as to smoothen the approach to the optimum. A parameter  $y(k)$  is damped (e.g.  $\Delta T(k)$ ,  $B(k)$ ,  $D(k)$ , and  $X(k)$ ) by the formula

$$y(k) = (1 - \alpha_k)y(k - 1) + \alpha_k \hat{y}(k) \quad ,$$

where  $\alpha_k$  is a certain selected sequence such that

$$\alpha_k > 0, \quad \alpha_k \rightarrow 0 \text{ as } k \rightarrow \infty, \quad \sum_k \alpha_k = \infty.$$

Consequently, the planned targets for the  $k$ th iteration are mean weighted values of parameters in the solutions available at this iteration rather than the parameters,  $\hat{y}(k)$ , themselves. With more iterations the corrections,  $\alpha_k \hat{y}(k)$ , decrease and may be made arbitrarily small so that the problem can be solved with a specified accuracy. The effectiveness of the damping system largely depends on the parameters and the damping law, or the sequence  $\alpha_k$ .

The upper level units respond to the new profits,  $D(k)$ , and losses,  $E(k)$ , reported by the lower level by finding a new distribution,  $B(k)$ , of the cargoes. The resultant solution is also damped. An annual plan is complete when changes from one iteration to another are negligible.

By using the results of iterative methods in Game Theory and Programming, it may be shown that the proposed iterative process converges to the optimum if  $D(k)$  and  $E(k)$  are dual estimates of the constraints (11) and (12).

### Quarterly Planning

Once an annual plan is found, a more detailed, quarterly planning begins which has to meet the year's targets and to respond to the updated information on demand and capabilities. Plans of shipments are made by an algorithm which is identical to that of making an annual plan.

### EXTENSIONS OF THE SYSTEM OF MODELS

One of the extensions is to provide for updating the annual plan at the beginning of each new period,  $t'$ , in response to recent data (the problem of optimal replanning). Thus the model of the system of annual planning is rearranged.

The problem of the upper level for the period  $t'$ :

$$\sum_{t \geq t'} \sum_j \sum_p \left[ D_{pjt}(t') B_{pjt}(t') - \theta_{pjt}^+ \xi_{pjt}^+(t') - \theta_{pjt}^- \xi_{pjt}^-(t') \right] - \sum_{t \geq t'} \sum_p \sum_M \left[ E_{pt}^M(t') \Delta T_{pt}^M(t') + \gamma_{pt}^{M+} \zeta_{pt}^{M+}(t') + \gamma_{pt}^{M-} \zeta_{pt}^{M-}(t') \right]$$

$\rightarrow \max$  ,

$$\sum_p B_{pjt}(t') \leq \sum_{\ell \in L_j} \tilde{\beta}_{\ell t}(t') \quad , \quad j = \overline{1, \bar{j}} \quad , \quad t = \overline{t', \bar{t}} \quad ,$$

$$\sum_{t \geq t'} \Delta T_{pt}^M(t') = \tilde{\Delta T}_p^M - \sum_{t < t'} \Delta T_{pt}^M \quad , \quad M = \overline{1, \bar{M}} \quad , \quad p = \overline{1, \bar{p}}$$

$$B_{pjt}(t') - \xi_{pjt}^+(t') + \xi_{pjt}^-(t') = B_{pjt}(t'-1) \quad ,$$

$$j = \overline{1, \bar{j}} \quad , \quad p = \overline{1, \bar{p}} \quad , \quad t = \overline{t', \bar{t}} \quad ,$$

$$\Delta T_{pt}^M(t') + \zeta_{pt}^{M-}(t') - \zeta_{pt}^{M+}(t') = \Delta T_{pt}^M(t'-1) \quad ,$$

$$p = \overline{1, \bar{p}} \quad , \quad M = \overline{1, \bar{M}} \quad , \quad t = \overline{t', \bar{t}} \quad ,$$

$$B_{pjt}(t'), \Delta T_{pt}^M(t'), \zeta_{pt}^{M\pm}(t'), \xi_{pjt}^{\pm}(t') \geq 0 .$$

$t'$  : The problem of subunit  $p$  of the lower level for the period

$$\sum_M \sum_j \sum_t \left[ c_{pjt}^M(t') x_{pjt}^M(t') - \eta_{pjt}^{M+} \kappa_{pjt}^{M+}(t') - \eta_{pjt}^{M-} \kappa_{pjt}^{M-}(t') \right] \rightarrow \max ,$$

$$\sum_M x_{pjt}^M(t') \leq B_{pjt}(t') , \quad j = \overline{1, \bar{j}} , \quad t = \overline{t', \bar{t}} ,$$

$$\sum_j A_{pjt}^M(t') x_{pjt}^M(t') \leq \tilde{T}_{pt}^M(t') - \Delta T_{pt}^M(t') , \quad M = \overline{1, \bar{M}} , \quad t = \overline{t', \bar{t}} ,$$

$$x_{pjt}^M(t') + \kappa_{pjt}^{M-}(t') - \kappa_{pjt}^{M+}(t') = x_{pjt}^M(t'-1) ,$$

$$M = \overline{1, \bar{M}} , \quad j = \overline{1, \bar{j}} , \quad t = \overline{t', \bar{t}} ,$$

$$x_{pjt}^M(t'), \kappa_{pjt}^{M\pm}(t') \geq 0 .$$

This model's criteria include the costs of updating

$\left( \theta_{pjt}^{\pm} \xi_{pjt}^{\pm} , \gamma_{pt}^{M\pm} \zeta_{pt}^{M\pm} , \text{ and } \eta_{pjt}^{M\pm} \kappa_{pjt}^{M\pm} \right)$ , as well as the profit from shipments and losses from the non-use of the ship during repair. Because the form of these problems is the same as the above ones, all the reasoning used in this paper is applicable to them too.

A quarterly plan can be updated in a similar way.

The model should recognise the possibility of chartering a certain number of ships, and is applicable to the planning of transportation by other kinds of vehicles.

The method would also work in a more general, dynamic model in which unshipped cargoes can be handled at later periods, and, although developed for two-level models in a linear programming form, the method is valid for other linear models in which the same kind of interaction is used, such as network models, and can be extended to multi-level systems.

REFERENCES

- Voronkov, A.V., and Yu. V. Klementiev (1972), *The Soviet Merchant Fleet* (in Russian), Transport, Moscow.
- Aven, O.I., et al. (1974), A Multilevel Model of Shipment Planning, in *Optimization Methods - Applied Aspects*, Preprints of IFAC/IFORS Symposium, Varna, Bulgaria.
- Iterative Methods in Game Theory and Programming* (in Russian), Nauka, Moscow, 1974.

Optimal Scheduling of Trucks - A Routing Problem  
With Tight Due Times for Delivery

O.B.G. Madsen

INTRODUCTION

The problem is part of a larger problem concerning the design of an integrated computerized distribution control system for one of the largest newspaper and magazine distribution companies in Denmark (Madsen, 1972).

THE PROBLEM

A certain area contains  $n$  points. A set of geographical coordinates and a due time are attached to each point. Also given is the desired number of routes,  $q$ , the point from which the routes start, and the related starting times. Furthermore, it is assumed that it is unnecessary to take the truck capacity into consideration.

Determine  $q$  routes that will minimize the total distance traveled when visiting the  $n$  points before their due time.

THEORETICAL SOLUTION

No Due Times

We then have a  $q$ -dimensional traveling salesman problem called "the clover leaf problem". This problem can be solved by branch and bound methods (Lilbaek and Winther, 1970). Numerical experiments indicate that the computing time increases very rapidly with the number of points, and decreases with the number of routes. For the time being, it is not realistic to solve problems including more than 70-80 points by an optimum finding algorithm (Hansen and Krarup, 1974).

One Route With Due Times

It is possible to solve this problem also by branch and bound methods (Lilbaek and Winther, 1970), but it is very time consuming, especially if the number of points exceeds 20.

Several Routes With Due Times

This problem, again, can probably be solved by branch and bound methods, but not without enormous consumption of computer time. Furthermore, it would be difficult to include different starting times in the model. It is, therefore, advisable to use a heuristic method to obtain an applicable procedure.

HEURISTIC METHOD No. 1

The first heuristic to be constructed was based on a combination of draw-functions. The values of these draw-functions are calculated for each truck and for each point not yet visited, and they depend on:

- (a) The actual time.
- (b) The due time for the point.
- (c) The distance between the truck and the point.
- (d) The average distance between the truck and the points not yet visited.
- (e) The earliest due time for the points not yet visited.

The contributions from these five elements are combined to form a single draw-value

$$\alpha_1 a + \alpha_2 b + \alpha_3 c + \alpha_4 d + \alpha_5 e .$$

Hence the heuristic method can be expressed in the following way:

- (1)  $t$  (time) = 0.
- (2) The trucks are placed at the starting point and they are now released at the given starting times.
- (3)  $t = t + 1$ .
- (4) Are any trucks ready? (not occupied).
  - (a) No: go to 3.
  - (b) More than one: choose one and go to 5.
  - (c) One: go to 5.
- (5) Calculate for this specific truck the draw-functions in connection with each point not yet visited by any of the trucks.

- (6) The truck is going to visit the point with the largest value of the draw-function. The point is marked as a point visited and the truck is marked occupied until the point is actually visited.
- (7) Have all points been visited?
  - (a) No: go to 3.
  - (b) Yes: go to 8.
- (8) Stop.

This type of heuristic method gave good results in earlier practical cases. In a way it is similar to the Lagrange multiplier technique, where you incorporate some of the constraints in the objective function. The first tests gave promising results. Unfortunately, later application of the method was discouraging. A case with 5 trucks starting from the same point and 14 points with due times, took 1.5 min on an IBM 360/40 computer generating one solution (corresponding to about 1.5 s on an IBM 360/75 or about 0.5 s on an IBM 370/165). After a variation of the weights, we got 30-40 solutions, each of which was calculated in about 1.5 min, but none of the solutions were feasible because of the very tight time constraints. In addition, the practical experiments with the algorithm showed that it was not possible to give a reasonable decision rule to determine the values of  $\alpha_1, \dots, \alpha_5$ .

#### HEURISTIC METHOD No. 2

After these discouraging results, a very simple algorithm based on the Monte Carlo technique was made. The reason for choosing the Monte Carlo method was that if you have to choose  $\alpha_1, \dots, \alpha_5$  in an arbitrary way, it is of no use to combine the method with a rather complicated calculation principle.

The Monte Carlo method was extended by a very simple routine to move a point from one route to another. The method can briefly be described as follows:

- (1)  $m = 1$ .
- (2) All the points are allocated arbitrarily to the  $q$  routes.
- (3) The points belonging to a route are placed in an arbitrary sequence.
- (4) The points belonging to a route are now, one by one, placed on the route, their due times considered in the sequence order from 3 and in such a way that the latest departure time is maximized.

- (5) Is the solution "better" than the preceding one?
  - (a) Yes: go to 7.
  - (b) No:  $m = m + 1$ .
- (6) Is  $m \geq 6$ ?
  - (a) Yes: go to 1 (or 9).
  - (b) No: go to 7.
- (7) Find the critical point of the most critical route and move this point to the least critical route.
- (8) Go to 3.
- (9) Stop.

It is assumed that because of the tight due times for delivery, one truck can only be used for one route. Remembering the discouraging results from heuristic method 1, the objectives have now changed slightly: we are now satisfied if we can find feasible solutions.

In order to make the formulation of the problem a little more realistic (and to facilitate the application of heuristic methods), it is assumed that newspapers are distributed in standard packages all containing the same number. The newspapers arrive unpacked at the starting point at time  $t_p$ , and are then packed at a constant rate. The packing time also includes time for loading the trucks. The trucks drive at constant speed. The delivery points are characterized by their geographical coordinates, their demand measured in standard packages, the time necessary for unloading the packages, and the latest time for delivery,  $l_{ij}$  (point  $i$  on route  $j$ ).

Point 4 is carried out by a special procedure (a sort of traveling salesman routine). When route  $j$  is created the starting time  $s_j$ , from the packing center can be calculated assuming that all the delivery points,  $i$ , must be visited before their due times,  $l_{ij}$ . From this, the actual delivery time,  $t_{ij}$ , for point  $i$  can be determined, and for at least one of the points we know that  $l_{ij} = t_{ij}$ . This point  $i_j^c$  is called the critical point on route  $j$ .

Having determined all the routes, we can now draw Figure 1. For each route-starting time  $s_j$ , the number of ready packages needed so far is shown. Knowing the constant packing speed, we can determine the latest possible packing start time,  $t_p^c$ . At the



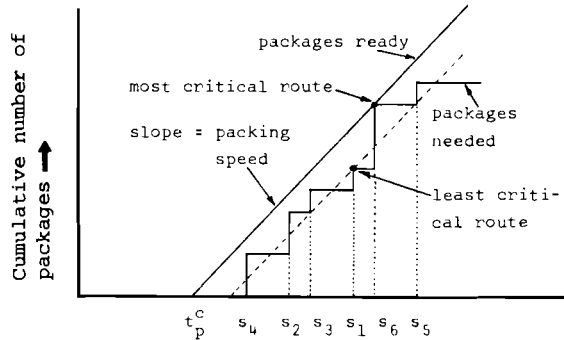


Figure 1. Cumulative number of packages.

same time we obtain the most critical route,  $j^c$ , which is route 6 in Figure 1. If the packing process starts later than  $t_p^c$ , at least the critical point  $i_j^c$  on the critical route  $j^c$  will receive the packages too late. In a similar way, we can determine the least critical route  $j^u$  (route 1 in Figure 1).

If  $t_p^c \geq t_p$  (the newspaper arrival time), the solution is called feasible, and the solution is stored for later listing. By a 'better' solution (see point 5), we mean that  $t_p^c$  is later than in the previous solution. In the example in Figure 1, point 7 of the heuristic method moves the critical point on route 6 to route 1.

For feasible solutions, we have a nonnegative slack,  $t_p^c - t_p$ . If a solution with maximum slack is preferred, then the objective function mentioned in the description of the problem will be satisfied automatically.

The computational results show that for  $n = 14$  and  $q = 5$ , it is now possible to generate about 200 routes in 1.5 min on an IBM 360/40 computer. On average, only two of these routes are feasible, the number being very sensitive to  $q$ , the number of trucks. If  $q = 6$  (in the example mentioned above), it was very easy to find a feasible solution, and if  $q$  was decreased to 4 it was almost impossible to find any feasible solution at all on account of the due-time constraints.

The test example is the same as used for heuristic 1.

ACTUAL CASE

Figure 2 shows the results from a real life case containing 15 points. The routes are generated by heuristic 2, and for a fixed number of trucks the routes are chosen to maximize the slack between the newspaper delivery time,  $t_p$ , and the latest time for the start of packing,  $t_p^c$ . It was impossible to generate a feasible solution with 3 trucks, and if you had more than 6 trucks, the extras would not be needed.

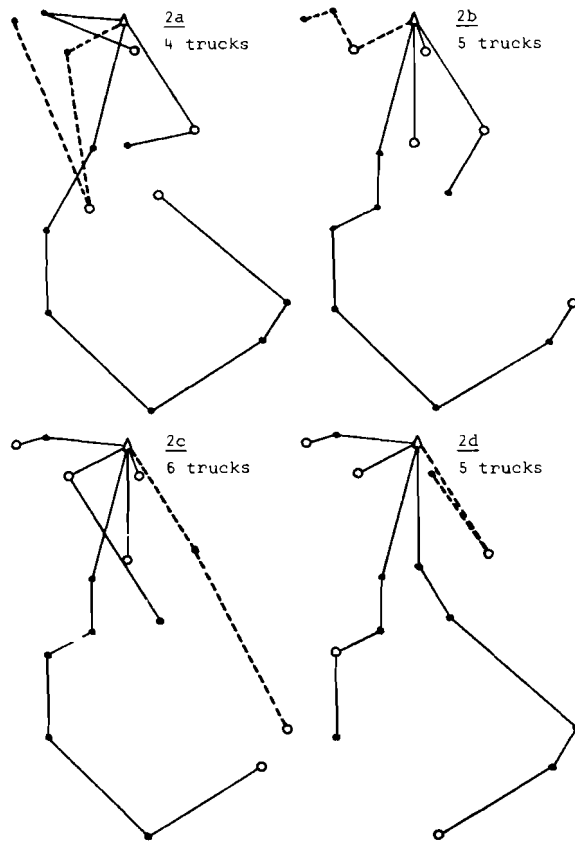


Figure 2. Routes in a real life case.  $\Delta$ , packing center;  $\circ$ , critical point; ---, most critical route; —, routes.

<u>no. of trucks</u>	<u><math>t_p</math></u>	<u><math>t_p^C</math></u>	<u>slack (min)</u>
3	165	-	neg.
4	165	165	0
5	165	173	8
6	165	178	13
7	165	178	13

Figure 2d shows the actual solution used by the company, which needed 5 trucks and gave a negative slack of 1 min!

#### UNCERTAINTY AND STABILITY

The newspaper delivery time  $t_p$ , however, varies in a stochastic way from day to day. In order not to have to make daily changes in the routes, we wanted to construct routes which are stable to changes in  $t_p$ , and which on average give the smallest total cost. In the case mentioned, we have a very detailed description of the variations in  $t_p$ . The newspapers arrive only 20% of the days in due time, while there are delays of at least 30 min on 40% of the days. If the delays are smaller than the slack, it will still be possible to deliver all the packages in due time, but if the delays are greater, then additional costs must be taken into consideration. These costs are mainly due to overtime payment to the persons who distribute the newspapers to people at the final end of the distribution chain. In addition to this, you have unsatisfied customers, but this is very difficult to quantify.

In the actual case you will have minimum expected costs if you choose 5 trucks and take the cost of the trucks and the delay cost into consideration.

Heuristic 2 generates feasible solutions that can be compared with respect to the quality of the routes, the costs, and the sensitivity to delays. It is very easy to perform additional analysis with the program, for example, to analyze changes in packing speed, truck speed, or latest due times for delivery.

#### EXTENSIONS OF THE MODEL

The actual case had 5-10 packing centers, 100-150 delivery points, and around 1500 packages. Hence it was necessary to divide the delivery points into districts belonging to each packing center in order to use heuristic 2.

ACKNOWLEDGEMENTS

I am indebted to Michael K. Knudsen, M.Sc., former operations research specialist at a large newspaper distribution company, for his assistance. He has implemented heuristic 2, and has constructed and compared some of the computer programs.

REFERENCES

- Hansen, K.H., and J. Krarup (1974), Improvements of the Held-Karp Algorithm for the Symmetric Travelling Salesman Problem, *Mathematical Programming*, 7, 87-96.
- Lilbaek, F., and E. Winther (1970), *Bladkompagniet Ltd., Progress Report no. 1* (in Danish), IMSOR report, Sept., Technical University of Denmark, Lyngby.
- Madsen, O.B.G. (1972), *Introduction to an integrated computerized distribution control system* (in Danish), IMSOR report, March, Technical University of Denmark, Lyngby.

## Optimal Bus Routing: A Case Study

C. Mandl

### GENERAL REMARKS

Transportation has become one of the urgent problems of urban areas. In many cities, the gap between the need for, and the possibilities of, transportation seems still to be growing, resulting in increasing travel times for citizens. Air pollution, noise, and accidents are quite unwanted side effects of this development. Therefore new solutions and planning in urban transportation systems are needed. Although, on the political scene, it is still an open question if public transportation systems should be given priority over individual car traffic, it has become quite clear that individual transportation systems, like the present day car, will not be able to solve the urban transportation problem. Since at present there exists no alternative to cars other than buses, trams, and railways (above and below the ground), strong efforts to improve these mass transportation systems should be undertaken.

Planning transportation systems cannot easily be separated into various subproblems, because all the aspects of such a system depend on each other.

Transportation planning tries to fulfill the forecast demand, but rarely takes into account the fact that future demand also depends on the results of today's transportation planning. People choose their jobs and homes depending on transportation facilities, and firms, shops and offices try to find good locations that also depend on transportation facilities. Therefore a good transportation system tends to create new transportation demands. The long term effect of this important phenomenon can hardly be forecast yet, as too little is known about people's behavior in this respect.

Given a certain transportation demand, public and individual transportation are in a competitive situation. Although it is not too difficult to evaluate the number of people that use one or other transportation possibility, only weak models exist for forecasting the splitting of demand between the two for a changed transportation system, e.g. if a better road, or new bus or underground line is built. Forecast models of this type, called modal-split models, are presented in H. Hensel & P. Mäcke (1975). So far modal-split models are only of the forecast-type and do not try to optimize certain criteria of transportation planning.

Quantitative models that do optimize or suboptimize come into consideration if one is willing to neglect:

- long term effects - that means accepting the assumption of independence between forecast demand and transportation planning, and
- the competitive situation between public and individual transportation implying that the demand for a public transportation system does not change if that system is changed - it is assumed that people either go by car or by a public transportation system, but do not move from one to the other.

We are quite aware of these assumptions and have to state that for an optimization problem, we do not yet see a way out of this dilemma. So far, two case studies of this type (see Lampkin and Saalmans, 1967, and Silman et al., 1974) have been described. The case study we are going to present has certain properties that differ from these two and that have made it possible to develop a more effective but less general heuristic algorithm for optimization (see also Hein and Hoidn, 1975).

## THE CASE STUDY

### The Problem Formulation

The case study presented was done for Aarau, a smallish town in Switzerland. There, the public transportation network, which was all buses, had to be completely reorganized. The problem was to find reasonable routes.

For this purpose, Aarau was divided into zones small enough that traffic within each zone could be ignored; people within a zone would all use the same public transportation facility (a zone, divided by a river without a bridge could not be allowed, for example, because people in such a zone obviously could not use the same bus route). The important data to be found out was the demand for public transportation between all pairs of zones, so it was therefore necessary that these zones were also created such that statistics on transportation within Aarau that already existed could be applied. Only the rush hour demands were represented, because the bus lines would stay the same during the rest of the day. The demands were symmetric, i.e. the same demand between zone  $i$  and  $j$  as between  $j$  and  $i$ .

Given the zones and the fixed transportation demands between them, all possible streets connecting these zones were stated and the travel time on such a street (assumed to be constant) was given. The map of Aarau with the zones and the streets can be seen in Figure 1. All these data lead to the formulation of a graph where the nodes represent the zones, the edges represent streets, and the length of an edge is the distance (see Figure 2 - some edges are directional because of one way streets). Certain



Figure 1. Map of Aarau.

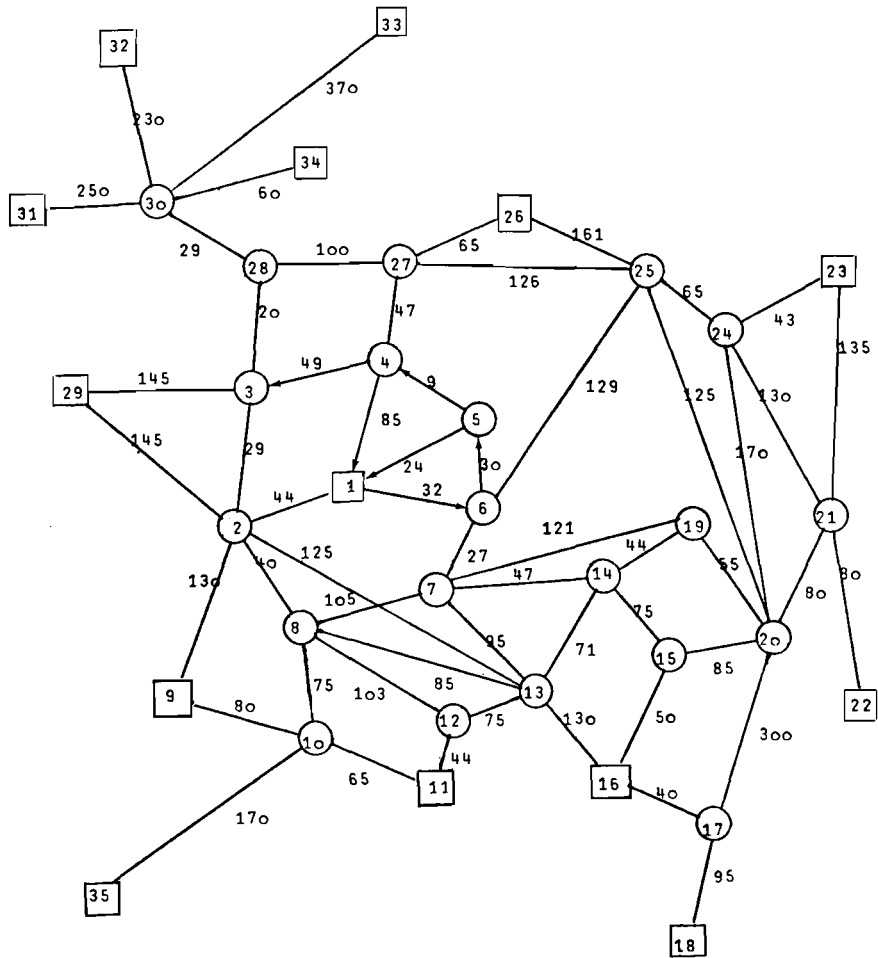


Figure 2. Directed graph of the street network.  
Distances are given in 10 m units.  
□, start or end points for bus routes.

restrictions were imposed by the local authority of Aarau that reduced the complexity of the problem. All extreme nodes (9, 11, 18, 22, 23, 26, 29, 31, 32, 33, 34 and 35), as well as nodes 1 and 16, have to be the starting or end points of one and only one bus line (see Figure 2). This fixes the number of bus lines to 7, as there are 14 such end nodes. Each node has to be served by at least one bus line, i.e. all zones of Aarau should be included in the public transportation system. The most important restriction is that all bus lines have to pass through node 1, the railway station, because a high but unknown number of people come to work, to school or for shopping from outside Aarau by train. This results in a high transportation demand from the railway station to all zones of Aarau.



Under these restrictions, the goal is to find 7 bus lines that would minimize the total transportation time of all passengers. For a given set,  $S$ , of 7 such lines and a given time,  $t$ , that one needs to change buses, the shortest route between a pair of zones  $(i,j)$  can be computed. If the travel time on this shortest route is called  $T_{S,t}(i,j)$ , then the total transportation time  $TR(S,t)$  is

$$TR(S,t) = \sum_{\substack{i,j \\ i \neq j}} d(i,j)T_{S,t}(i,j) \quad (1)$$

where  $d(i,j)$  is the demand between the pair  $(i,j)$ . The parameter  $t$  depends, of course, on the frequency the buses serve the nodes. Thus the problem was to find a set of 7 bus lines that fulfill the restrictions and minimize  $TR(S,t)$  for a given  $t$ . The local authority of Aarau did not want to find an optimal  $t$ , but rather preferred a set of solutions for different  $t$  (especially  $t = 5, 10, 15$ , and  $20$  min).

As the objective,  $TR(S,t)$ , does not seem to be directly applicable for optimization, we reformulate the objective. The function  $T_{S,t}(i,j)$  consists of the travel time and the waiting time, if a change of bus is needed. Now, if the waiting time,  $t$ , is large compared with the travel time (which is true in our case for  $t \geq 10$  min), the latter can be ignored. But, if the waiting time is considered as the only objective, the restriction that all bus lines are near to the shortest paths of the graph given in Figure 2 has to be included to ensure that the travel time remains short. Because all buses pass the railway station, at most one change of bus will be necessary to go from zone  $i$  to zone  $j$ . The original objective function (1) is replaced by

$$TW(S,t) = \sum_{\substack{i,j \\ i \neq j}} d(i,j)W_{S,t}(i,j) \quad , \quad (2)$$

where  $TW(S,t)$  is the total waiting time,  
and

$$W_{S,t}(i,j) = \begin{cases} t & \text{if zone } i \text{ and } j \text{ are not served by the} \\ & \text{same bus line (bus change necessary),} \\ 0 & \text{if zone } i \text{ and } j \text{ are served by the same} \\ & \text{bus line (bus change not needed).} \end{cases}$$

The Heuristic Algorithm

As it does not seem possible to solve the given problem exactly, primarily because a standard formulation (like integer programming) of the problem cannot be given, we construct a heuristic algorithm, the idea of which, as usual, is to break the original optimization problem into smaller ones, which are solved sequentially, in the hope that the outcome is a reasonable suboptimal solution of the problem.

Step 1:

As all routes should be near shortest paths within the given graph and as all routes should pass by node 1, the first step is to find the shortest routes from and to node 1 to and from all other nodes. This results in the graph shown in Figure 3.

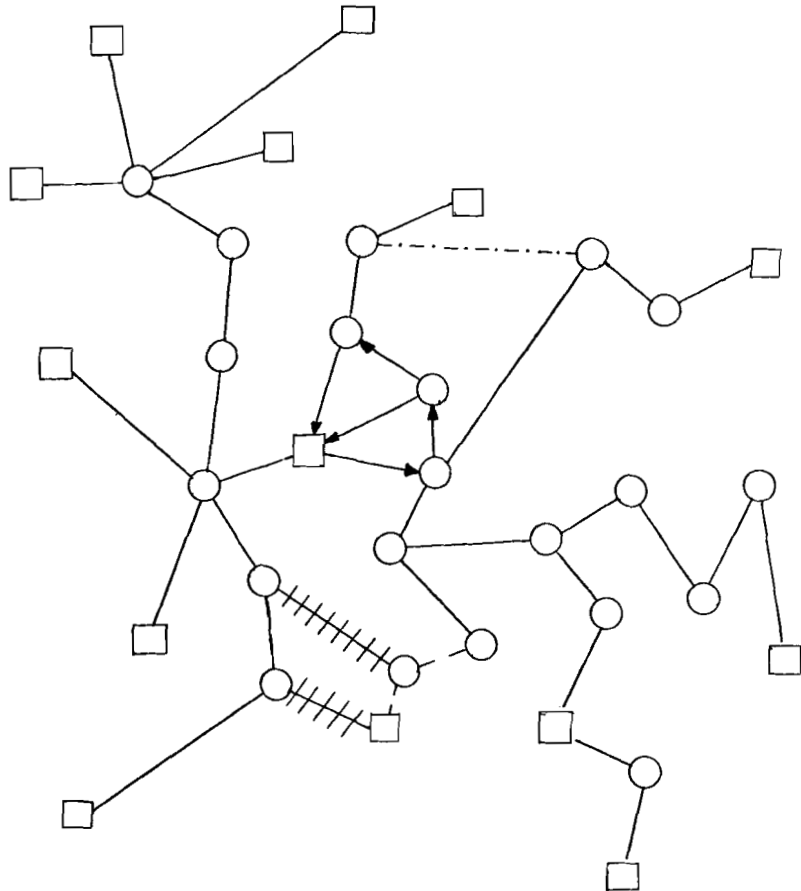


Figure 3. Shortest routes from node 1 to all other nodes and vice versa.

Except for nodes 12 and 13, all end nodes are also end points of the bus routes. As node 12 and 13 could only be combined with node 11 or node 16, a revised graph with dotted lines was chosen as shown in Figure 3. So, step 1 results in defining (almost) the shortest routes from all given end nodes to node 1, where all nodes of the original graph of Figure 2 are included.

Step 2:

Part 2 of the algorithm is to find all feasible routes from the end nodes to node 1 such that all nodes belonging to a route still belong to this route, i.e. only new nodes may be included but none excluded.

A route is feasible if

$$\text{length of route from end node to node 1} \leq a \times \text{shortest path from end node to node 1 (given in Figure 3)}.$$

(3)

This constant  $a$  was chosen to have a value of 1.3, but some sensitivity analysis was done on  $a$  at the end of the study.

The algorithm by which all feasible routes were found is the following:

- (a) For a pair of nodes  $(i,j)$  that are directly connected by an edge 1 in Figure 3, exclude this edge 1 from the graph in Figure 2.
- (b) Find shortest path in both directions between this pair of nodes  $(i,j)$ . If none exists or if this path includes a node that already belongs to this bus route, choose next pair of nodes and go to (a), or, if all pairs already chosen go to (e).
- (c) For this new path prove whether (3) still holds. If so, go to (d); if not, find next pair of nodes and go to (a); otherwise go to (e).
- (d) Save this path, find next pair of nodes and go to (a), otherwise go to (e).
- (e) Find feasible combinations of already found paths and save them as well.

The result of this algorithm for  $a = 1.3$  was to produce only one more possible route shown in Figure 3 as ---.

Step 3:

Finally the best routes, combining always pairs of end nodes to each route, have to be found.

This problem can be solved in two ways. Although we chose the second method, both of them are presented here.

(a) Linear programming formulation

Let  $(i, \ell)$  define the  $\ell$ th feasible path from end node  $i$  ( $i = 2, \dots, 14$ ) to node 1. For the moment, one-way streets are treated as two-way streets. From each node there are  $L_i$  possible paths. In our problem, for all  $i$   $L_i = 1$ , except  $L_2 = 2$ .

Let

$$x_{ij\ell k} = \begin{cases} 0 & \text{if path } \ell \text{ of line } i \text{ is not connected with} \\ & \text{path } k \text{ of line } j \text{ ,} \\ 1 & \text{if path } \ell \text{ of line } i \text{ is connected with path} \\ & \text{k of line } j \text{ .} \end{cases}$$

Each line must be connected with exactly one other line but two paths of one line may not be connected with each other.

That means, where clearly  $x_{ij\ell k} = x_{jik\ell}$  ,

$$x_{i\ell k} = 0 \text{ , } i = 1, \dots, 2N \text{ ,} \\ \ell = 1, \dots, L_i \text{ , } k = 1, \dots, L_j \text{ ,}$$

where  $N$  is the number of bus routes and  $2N$  the number of end nodes.

$$\sum_{j=1}^{2N} \sum_{\ell=1}^{L_i} \sum_{k=1}^{L_j} x_{ij\ell k} = 1 \text{ , } i = 1, \dots, 2N \text{ . } (4)$$

Let  $a_{ij\ell k}$  be the number of passengers who can travel without a bus change on a route combining nodes  $i$  and  $j$  via path  $(i, \ell)$  and  $(j, k)$ . Because of the symmetry of the transportation demands it holds that

$$a_{ij\ell k} = a_{jik\ell} \text{ .}$$

Then the objective is

$$\sum_{i=1}^{2N} \sum_{j=1}^{2N} \sum_{\ell=1}^{L_i} \sum_{k=1}^{L_j} a_{ij\ell k} x_{ij\ell k} \rightarrow \max .$$

As we know, this problem solved as a linear program, with restriction (4), and  $x_{ij\ell k} \geq 0$ , always gives an optimal solution where

$$x_{ij\ell k} \in \{0,1\} , \quad \forall i,j .$$

This formulation has some disadvantages, because of the way the  $a_{ij}$ 's are defined, where some transportation flows are counted more than once, because the demands between pairs of nodes served by, for example, two lines are counted twice. But one can argue that those demands fulfilled by more than one bus route are fulfilled better, and should therefore count more.

- (b) To avoid the problem of double counting, the actual algorithm chosen is sequential:

As the first bus route  $(i,j)$ , the one is taken that has the highest number of passengers using it without a change of bus. Then the matrix of transportation demands is reduced by setting the already fulfilled demands to zero. Then the next best route is chosen and so on. To find the best route, one just has to choose  $\bar{i}, \bar{j}, \bar{k}, \bar{\ell}$ , such that

$$\max_{i,j,k,\ell} a_{ijk\ell} = a_{\bar{i}\bar{j}\bar{k}\bar{\ell}}$$

Step 4:

Finally, whether some routes pass along one way streets in the wrong direction is checked, and if so, for this section the best alternative route is chosen.

Results and Sensitivity Analysis

As a result of the algorithm presented here, the bus routes were chosen as shown in Figure 4.

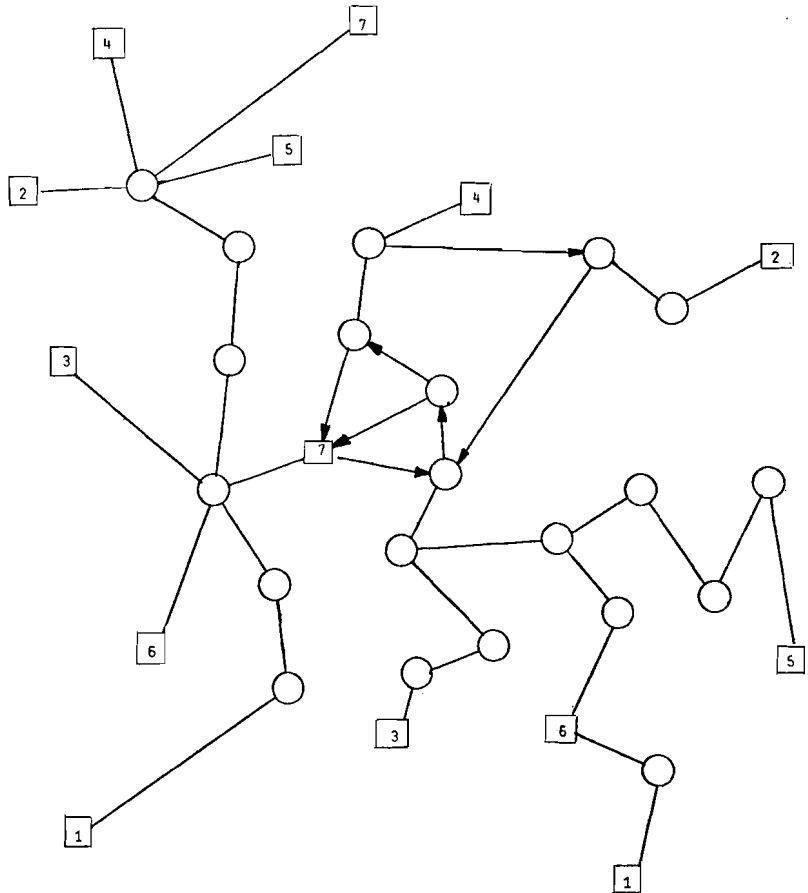


Figure 4. Final bus routes (lines with equal number are combined into one route).

This transportation network was then measured by various other criteria:

- The percentage of passengers who could use this network without changing buses. This turned out to be 57% of the total.
- The number of passengers between each pair of nodes. This is found by determining the routes of all passengers, assuming that all passengers change buses not more than once and travel along shortest routes within the given transportation network of Figure 4.
- The total number of passengers for each bus route, assuming that along edges where more than one bus route

passes, the number of passengers is split equally between these bus routes.

- The total transportation time for a given average waiting time,  $t$ .

Sensitivity analysis was done on the value  $a$  of restriction (3), and on the given data, especially the transportation demands. For  $a \geq 1.3$ , there was no change until  $a = 1.5$ . Higher values were not analyzed, because they did not seem to be meaningful. For some pair of zones the demands were changed (proposed by the local authority), but the result remained stable.

#### REFERENCES

- Hein, W., and H.P. Hoidn (1975), *Planung des Busnetzes von Aarau*, Semesterarbeit, Institut für OR, ETH-Zürich.
- Hensel, H., and P. Mäcke (1975), *Arbeitsmethode der städtischen Verkehrsplanung*, Bauverlag GmbH, Wiesbaden.
- Lampkin, W., and P.D. Saalmans (1967), The design of routes, service frequencies and schedules for a municipal bus undertaking: a case study, *OR Quarterly*, 18, 4, 375 - 397.
- Silman, L.A., et al. (1974), Planning the route system for urban buses, *Comput. and OR*, 1, 201 - 211.





**CONTROL AND GUIDANCE  
OF TRANSPORTATION SYSTEMS**



### Summary of the Panel Discussion\*

In this session the panel discussion dealt mainly with the following three questions:

- what advanced methods of optimal control and identification theory are already used or could be used in transportation systems?
- what similarities and differences exist in the application of this methodology to different modes of transportation, i.e. railway systems, urban street and freeway traffic, automated guideway transit, etc.?
- what are promising areas for future applied and fundamental research?

The results of the discussion can be summarized as follows.

#### Potentials and Limitations of Optimal Control Theory

The papers presented in this session of the workshop as well as at other IFAC meetings during the last three years or so show that transportation systems are fruitful application areas for modern control theory. This is especially true for the analysis of computerized systems for control and surveillance of:

- urban, freeway, and street traffic, and
- the new demand-responsive automated guideway transit systems (AGT).

For these traffic systems remarkable developments in the application of advanced methods of:

- static and dynamic optimal control, e.g. by using Pontrjagin's maximum principles and Ricatti equations, etc.;
- state and parameter estimation by using extended Kalman filters; and
- principles of decentralized control of large scale systems,

---

\*Prepared by H. Strobel

have taken place during the last five years. Nevertheless, most of the work considers that the practical implementation is at an initial stage. A strong impulse for a more comprehensive usage of modern control theory will be given by the introduction of the new generation of process computers, i.e. microprocessors and microcomputers, into transportation systems.

#### Similarities and Differences Between Different Modes of Transportation

The control problems occurring in different transportation systems may be classified at three levels:

- route guidance,
- traffic flow control, and
- vehicle movement control.

The highest level contains similar problems from a methodological point of view for urban, freeway, and street traffic, for the new AGT systems, and for demand-responsive public transportation systems like the Dial-a-Ride system, but it does not have any significance for other modes of public transport. Similarities exist too for the two lower levels between freeway traffic and AGT systems especially with the so-called merging control problem and the distance regulation problem for a line of moving vehicles. These problems are obviously quite different from other transportation modes, e.g. rail or bus transport.

#### Promising Future Research Directions

The following three research directions were identified as promising from a methodological (control theory) point of view:

- development of decentralized traffic control systems on the basis of spatially distributed microcomputers, i.e. of microcomputer networks, especially with respect to urban, freeway, and street traffic control, and control and surveillance of new AGT systems;
- safety and reliability of these decentralized control systems and of computerized traffic control systems in general; and
- more extensive use of advanced identification methodology for the creation of traffic surveillance systems, e.g. the implementation of automatic incident detection systems for freeways, tunnels, etc.

### Abstracts of Papers

*A Review of Research and Development Programs of the U.S. Department of Transportation Concerned With Automation, Control and Optimal Planning*  
J.J. Fearnside

(Abstract only)

The Department's research and development programs were reviewed with special attention given to the similarities that exist among these programs, and to those aspects of them that seem most amenable to the application of modern control theory, operations research, and optimization methods. Three on-going projects in the Advanced Research Program, which is conducted in the Office of the Secretary of Transportation, were described in detail.

The first of these concerns the application of modern control and estimation theories to the optimal control of arterial street signals, ramp meters, and directional signs in an urban freeway corridor. Some of the questions addressed are:

- What are the cost and performance trade-offs between traffic control systems with a central control computer and those systems with many decentralized controllers?
- What advances in modern control theory are needed to assure realistic application to traffic control systems?
- What kinds of information should be shared among adjacent decentralized controllers and how does this affect the on-line estimation of the state of the system?
- How many traffic sensors are required and where should they be located for minimum cost?

The second project concerns the errors and biases imposed on the analysis of large-scale transportation networks by aggregations of nodes, links, or other parameters. The problem of disaggregating results obtained by aggregation was also discussed, with the principal focus on the traffic assignment problem.

The third project considers the comparative advantages of applying decomposition techniques for the analysis of transportation networks, the emphasis here being on the network design

problem and on the application of modern mathematical programming techniques and data handling methods to very large networks.

Finally, some recommendations for future research were made.

*On the Application of Optimal Control Theory to Traffic Control Systems Analysis*  
H. Strobel

(Already published by IIASA as part of Research Report RR-77-12)

This paper presented a survey of the application of modern control theory - i.e. of methods for modelling, identification, and control - to urban, street, and freeway traffic. For this purpose the question was considered of what methods of modern control theory - i.e. of dynamic optimal control, and parameter and state estimation - are already used or are expected to be used for the design of traffic and transportation control and surveillance systems. It was shown that the following hierarchy of control tasks deals with the question well:

- route guidance systems,
- traffic flow control, and
- vehicle movement control.

The paper analyzed the control and identification problems occurring in that hierarchy for urban, freeway, and street traffic.

*Optimization and Evaluation Techniques for Street Traffic Control*  
J.L. Schlaefli

(This volume, pp. 161-172)

A definition of the basic problems of the use of optimization techniques for street traffic control is given. Analytical techniques have been applied extensively in this field but most practical implementations have been based on empirical results. Simulation has emerged as a good "middle ground" for attacking urban traffic control problems. The advent of modern large-scale computers has fostered the initial successes of simulation approaches to optimizing and evaluating new concepts of street traffic control.

Three computer based simulation models are discussed. The first involves a method for determining optimal signal timing for arterial streets. This method is rigorous mathematically, and has enjoyed considerable success in the United States. The TRANSYT simulation developed in England enjoys an international reputation for being the best available tool for developing

optimal signal timing in a network of traffic signals. The TRANSYT technique has been applied extensively. The Dynamic Highway Transportation Model (DHTM) attacks a new problem in urban traffic control optimization in that it can treat dynamic variables (e.g. driver routing). This simulation technique is just beginning to be recognized for its unique capabilities.

Traffic management presents the most significant challenge to the research community. This and the other research needs defined should be given emphasis by research organizations throughout the world.

*Optimal Decongestion Techniques for Urban Traffic Networks*  
R. Camus, A. D'Amore, and P. Sipala

(This volume, pp. 173-186)

Most methods currently used for the optimal synchronization of traffic signals are based on the assumption that no arc is saturated by its traffic flow. A notable exception to this is provided by the system THESEE, developed by SETRA, where saturation is permitted, and an attempt is made, during the optimization phase, to eliminate or reduce it.

In this paper we present an extension of the THESEE methods, which permits complete removal of all remaining cases of saturation from the interior arcs of the network, by means of a suitable reduction of the traffic entering it. To this purpose, a technique drawn from graph theory is first adopted, to evaluate the transfer factors correlating internal traffic flows to the volumes entering the network. A linear programming algorithm is then applied to distribute the required traffic reduction among the input arcs in an optimal way to remove saturation. In the new condition, the optimum synchronization procedure is then applied, and the whole process repeated until no saturation occurs within the network.

The optimization technique and the functional to be minimized have also been considerably modified to take into account the possibility of exceeding maximum queue lengths, and the presence of preferential flows or directions.

*Design of Control Systems in Automated Transport Systems*  
L.D. Burrow

(This volume, pp. 187-193)

This contribution discusses the two classes of longitudinal control strategies in automated transport systems.

### Deterministic Systems

These require complete knowledge of every vehicle's present and future positions, implying prearranged journeys scheduled to remove conflicts with other vehicles.

An example of deterministic control is the synchronous slot (marker following) system. Control and communication costs have been much reduced by making each vehicle follow the same velocity position profile. Line capacity is therefore constant at all speeds.

### Stochastic Systems

These require only limited knowledge of vehicle positions. Vehicles run asynchronously with local resolution of conflicts. Hierarchical control structures are particularly appropriate with such schemes.

Vehicle follower methods are stochastic. Line capacity is a function of speed and frequently trade-offs can be made between the two. Intervehicle ranging is difficult and expensive. Vehicle platoons running under vehicle following conditions are hard to stabilize and the close packing of vehicles through a speed change, necessary to trade capacity and speed, is a complex operation.

The ability of vehicle following systems to respond locally to traffic conditions reduces the effects of failures, as opposed to marker follower schemes in which a fault can easily halt the entire system.

The shortcomings of these established control schemes demonstrates the need to design other control strategies more capable of running the system well under both normal and faulty conditions. The design process involved must optimize over the whole system, performance indices such as cost, delay and reliability, while satisfying safety criteria and performance constraints. Clearly to aim for the global optimization of such complex systems is unrealistic. At best only local optima can be found and combined to form an overall 'good' system. Optimization can proceed at three levels:

- Structure optimization,
- Maneuver subsystem optimization, and
- Parameter optimization.

Often the only structure optimization that can be achieved is the identification of required characteristics in similar existing systems, although some theoretical work on the structure of systems does exist.



The optimization of maneuver subsystems is well suited to a simulation approach. Within the main control structure it is helpful to isolate subsystems and study their main characteristics independently. Such an approach builds up a repertoire of expected behavior patterns leading to a better comprehension of the overall system often enabling better structures to be evolved.

Parameter optimization can frequently be approached mathematically, although in transport control the many parameter constraints and nonlinearities prevent general solutions from being found and recourse has to be made to iterative techniques.

The contribution considers this systematic approach to the design of complex control structures and uses as an example a hierarchical asynchronous system of vehicle control that has been developed.

*Optimal Control of Railway Marshalling Yard Operations*  
J. Sokolowski

(This volume, pp. 195-209)

This paper presents a model-building technique enabling one to construct an electronic data processing (EDP) system for automatic production of optimal or suboptimal patterns of conducting traffic operations on a marshalling yard.

These patterns perform two functions:

- they constitute a basis for short-term planning of traffic operations at a marshalling yard;
- they are a guide to the evaluation of marshalling yard operations.

The method is a compilation of operational research, statistical inference and system theory methods.

The paper outlines:

- the technique for the observation of a real object (as well as a formal systems approach to the description of observation) is given for the system and subsystems, the aims and criteria of the system and subsystem's functioning, and also all the valid system elements and the processes taking place within them;
- the method for formally describing elementary processes and examples of ways of determining certain transformations describing specific elementary processes;
- the structure of constructed models; and

- the idea of an EDP system devoted to medium-term planning of traffic operations at a marshalling yard.

*Optimal Air Traffic Control At Airports*  
Y. Istefanopoulos

(Abstract only)

Air-traffic congestion problems have been troubling major airports all over the world. In the present study, several analytical models for air-traffic in the general air terminal area were constructed, with the purpose of exploring some questions about air-traffic congestion and determining the optimal operational policies for various runway configurations.

Attention was first focused on a single runway used exclusively for landings, and the distribution for interarrival times of aircraft was obtained under capacity conditions. The inputs to the model are the velocity distribution of incoming aircraft, the error distribution in spacing the aircraft and the minimum separation requirements, both in the air and on the ground, as specified by the Federal Aviation Authority.

Realizing that parallel runways offer the important operational option of simultaneous landings under congestion conditions, we investigated the statistics of the service time for the landing operation on two independent parallel runways and discussed the safety problems involved with simultaneous approaches.

In the case of two dependent parallel runways we constructed an original scheme of sequencing aircraft, that allowed us to model the landing process as a single server queue.

The dual-lane runway configuration was analyzed next, and two operational modes considered. The first mode assumed alternating arrivals and departures for the dual-lane runway. Several statistics of interest associated with this situation were computed, and a numerical evaluation of the resulting delays and system capacity performed. The second mode assumed that departures can be released from the dual runway between successive landings if the gap is long enough. The probability of interleaving departures in the landing stream was calculated, as well as the expected number of departures per unit time dispatched under these conditions.

Finally, a model of segregating aircraft according to their approach speeds into two classes, with each class using a separate runway, was discussed in relation to minimizing waiting times.

*Optimal Vehicle Control from the Point of View of Energy Conservation*

L. Skyva

(This volume, pp. 211-224)

In this paper we try to find the optimal vehicle control from the point of view of energy consumption. We are especially concerned with train control as this is the most important kind of transportation in our country and besides, the route and its characteristics are fixed for computation.

Of the several methods for finding minimum-energy control, the best known and most widely used are Pontrjagin's maximum principle and Bellman's method of dynamic programming. The former is quite convenient for solving simplified problems, but for the real movement of a train, which is complicated, it is necessary to use the more ingenious numerical methods or Bellman's method. These need a comparatively large computer which cannot be placed on board the vehicle. Solving the problem in a computation center takes quite a time, but this is acceptable.

For future minimum-energy vehicle control we must find a convenient algorithm that uses a combination of the above methods, and that is fast and simple enough for using in a real time control system.



Optimization and Evaluation Techniques  
For Street Traffic Control

J.L. Schlaefli

BACKGROUND

Street traffic control has presented a perplexing problem to researchers and practicing traffic engineers throughout the world. Within the last ten years, dynamic changes in the technology available for street traffic control and evaluation have evolved, primarily through the development and application of digital computers. This paper attempts to summarize some of the important optimization and evaluation techniques that are currently being used, and identifies important unresolved problems. First, it is important to define some aspects of the problem of street traffic control. In order to make the scope of this paper manageable, some restrictive definitions had to be made. The work discussed herein applies to the basic control of vehicles (i.e. cars, buses, and trucks) in congested urban street networks. Further, the principal method being used today is by traffic signals installed at major intersecting streets.

Much research has focused on the problem of controlling traffic signals in an optimum way and evaluating the results. Obviously, a strong interaction between the traffic signal timing parameters and the dynamic structure of traffic flow exists. Understanding these interactions has a direct impact on the implementation of an effective traffic control policy. To approach the problem, one must deal with a set of rather restrictive static variables that can be generally categorized as follows:

- Location of vehicle origins and/or destinations.
- Physical network design, including lengths of blocks, street width, channelization, and physical barriers.
- Characteristics of the driver population.

On the other hand, the most important dynamic variables include:

- Traffic demand variations over space and time.
- Stochastic fluctuations of traffic flow.
- Vehicle - vehicle interactions.
- Vehicle - pedestrian interactions.

- Discharge headways.
- Truck and bus operations.
- Motorist behavior.
- Dynamic routing.

To present an optimal solution to traffic control problems in an urban environment involves treating the above static and dynamic variables completely, and presents a significant challenge. It is extremely important to keep a systematic approach in dealing with these problems. In general an approach should involve the following four steps:

- Develop a set of goals or criteria against which different solutions to the problem can be compared.
- Provide a method for generating potential solutions.
- Develop a method for measuring the effectiveness of each proposed solution against the goals and evaluation criteria that were established above.
- Compare the results and conclusions with real-world data to ensure that the best solution is implementable.

Optimization theory and evaluation techniques are principally applicable to the last two steps mentioned. The first two steps can be difficult to carry out when a specific problem is being considered. This paper deals with techniques used for the last two steps (i.e. for measuring the effectiveness of potential solutions to an urban traffic control problem).

Much of the work that has been done on the development of improved urban traffic control methods has not been implemented. The principal reason for this is not the lack of technology, but the inability of the research and development community to transmit the results in an implementable form to the urban decision-maker. Significant problems in communication exist. Often, multiple jurisdictions are involved, and the resulting political constraints invoked make it almost impossible to employ an "optimal" solution. Since traffic congestion is the subject of continuous attention from the public, urban traffic control problems gain political prominence. On the other hand, only recently have major financial resources been applied directly to solving traffic control problems. With the above general introduction in mind, the succeeding sections treat some optimization examples and attempt to point out some important research areas.

## OPTIMIZATION OF TRAFFIC SYSTEMS

The literature contains many examples of the application of optimization theory to traffic control. Classically, these examples involve simplifying the problem of single or multiple intersection control and applying an optimization analysis to come up with a proposed new technique. While there are many examples of this in the academic literature, few of these techniques have actually been put in practice, the major reason for which seems to be that the practicing engineers responsible for the operation of traffic control systems are operating under a set of constraints that allow little flexibility for trying new approaches. Traffic engineering practice is largely empirical; that is to say, if a new optimal control technique has not been thoroughly tested in a field study, then it very likely will not be adopted.

By far the most successful optimization approach in the traffic control field has been the use of simulation. Properly applied, traffic simulation is an optimization tool for comparing various control alternatives. One can see why this technique has emerged as most practical. On the one hand, analytical techniques, while being available at relatively low cost, require so many simplifying assumptions in the traffic control environment, that the results often are not real. The purely analytical approach is satisfying in that, once the results have been obtained, they are almost always reproducible. On the other hand, a purely empirical approach to developing and evaluating new control strategies using field test procedures, involves very high costs. While the outcome is very real since major assumptions are not necessary, the overall results of a field test are very often not reproducible. Simulation of traffic control policies forms a middle ground between analytical and empirical approaches. The present state-of-the-art indicates that the simulation approach, when supported properly by analytical techniques and field studies, can be the basis for practical developments of optimal traffic control policy. Some of the simulation techniques being used and their success will now be considered. One amazing fact in the field of street traffic control is that nearly 90% of the traffic engineers responsible for operating traffic signals in urban areas throughout the world use simple manual techniques for determining traffic signal parameters (i.e. cycle length, split and offset). With the increasing availability of computers, computer assisted signal timing methods are becoming more popular. Direct mathematical techniques and simulation models for optimizing traffic signal parameters are just emerging as practical tools.

### Maximal Bandwidth

One approach to the simple problem of optimizing signal timing along major streets is called the maximal bandwidth method and was developed by Morgan and Little of MIT (1964). It is

designed to determine the signal offsets for a major arterial street which results in maximizing the green bandwidth. The system can be designed to result in equal bandwidths in the two directions, or the synchronization can be adjusted to increase one bandwidth to some specified feasible value and maintain the other as large as is then possible. The method also permits the user to specify speeds by direction between any two adjacent signals, and apportion bandwidth between directions on the basis of platoon size. Later the method was expanded into a mixed-integer linear programming formulation (see Little, 1966). The new formulation offered no advantages and many disadvantages to the objectives mentioned above. However, the linear programming format opens up possibilities for solving more general problems, for introducing new decision variables, and may be extended to the problem of synchronizing signals on a network basis.

For efficient execution of this algorithm, a computer program was developed which arranges bandwidths so that vehicular platoons traveling in both directions along an arterial road fit into their respective green progression band. The program maximizes progression bandwidths subject to the following conditions:

- If the platoons in both directions are equal, maximum equal bandwidths are provided for each direction of travel.
- If the sum of the two bandwidths is greater than the sum of the two platoon lengths (in units of time), the individual bandwidths are made proportional to platoon lengths as far as possible.
- If the sum of the two bandwidths is less than the sum of the two platoon lengths, the larger platoon is first accommodated, if possible, and then as much bandwidth as can be arranged is given to the direction with the smaller platoon.

The maximal bandwidth model optimizes progression bandwidth, a geometrical quantity on the time-space diagram which does not necessarily relate to any actual traffic characteristics. In other words, the model does not necessarily minimize travel time, delay or stops. The effectiveness of the model is, therefore, dependent on the type of traffic conditions existing in the system and on how well the bandwidth is utilized. Where signals are closely spaced and vehicle platoons remain intact throughout the system, the model will provide efficient signal operation. The same is true if the traffic is light and the bandwidth is wide enough to allow for platoon dispersion. In cases where the bandwidth is barely adequate for the platoon at the start of the progression, platoon dispersion and interruption along the route will cause vehicles at the end of the platoon to stop at the critical intersections.

Despite its weaknesses, use of the model results in reasonably efficient arterial signal operation, provided care is taken



in its application and realistic input data are used. This has been demonstrated by recent field studies that ranked the maximal bandwidth method high among a number of alternative procedures for improving the operation of an urban arterial signal system (Wagner, et al., 1969).

#### Traffic Signal Network Study Tool

The most conventional approach to network signal timing is to provide preferential treatment for one or more arterial streets in the system. After favorable signal timings are assigned on the preferred streets, the remaining signals are adjusted to conform to the network. In effect, this approach involves reducing the network problem to one of a series of arterial progressions. The manual work involved in designing timings for a network is quite cumbersome and, in many urban areas, not done. Again, the advent of the computer and optimization techniques have led to the development of computer programs that have great potential for increasing the efficiency of a traffic signal network. The Traffic Signal Network Study tool (TRANSYT) is an excellent example of such an approach (Robertson, 1969). In essence TRANSYT is an optimization technique for computing signal offsets and splits for minimum delay and stops in a network. The technique has two main elements: (1) the simulation model, which is used to calculate the performance index of the network for a given set of signal timings, and (2) a hill-climbing optimization process which leads toward optimal signal phasing and offsets.

Generally speaking, TRANSYT is based on the following assumptions:

- All major intersections in the network are signalized.
- All signals have a common cycle length or a cycle length half the common value.
- Traffic enters the boundary signal at a constant specified rate.
- The right-turn and left-turn ratios are constant throughout the cycle. (No re-routing)

As stated above, the computer program consists of the computation of traffic flow patterns on all network links and a hill-climbing process for optimizing offsets and splits. To find the optimum signal timing, a system performance index,  $P$ , is used as a means of evaluation:

$$P = \sum_{i=1}^n (d_i + d_i' + KS_i)$$

where  $i$  is the link number,  
 $n$ , the number of links in network,  
 $d_i$ , the average uniform delay along the  $i$ th link,  
 $d'_i$ , the average random delay along the  $i$ th link,  
 $S_i$ , the average number of stops along  $i$ th link, and  
 $K$ , a stop penalty factor to convert the number of stops into equivalent seconds of delay.

To calculate the performance index, traffic flow information on each link is required. In the computation routine, the cycle is divided into a number of equal units of time. The flow rate entering a link during each interval is assumed to be some fraction of the flow leaving the upstream links. To obtain the arrival rates at the downstream signal, the flow entering the link is exponentially smoothed by the use of a platoon dispersion model. The departure rate leaving the link is assumed to be equal to the saturation flow when a queue exists at the signal approach, or equal to the arrival rate if no queue is present. Then the average uniform vehicle delay along the  $i$ th link for an average cycle is given by:

$$d_i = \frac{\Delta t}{A_c} \left\{ \sum_{t=1}^n (A-D) \right\} ,$$

where  $t$  is the interval number,  
 $n$ , the number of computational steps,  
 $A$ , the cumulative arrivals,  
 $D$ , the cumulative departures,  
 $A_c$ , the total arrivals over the entire cycle, and  
 $\Delta t$ , the size of the computational step ( $1/n$ ) .

The above equation is valid only if the queue at the approach is completely dissipated at some time during the green.

To allow for random fluctuations of traffic flow from cycle to cycle, an average random delay component is computed as follows:

$$d_i' = \frac{x^2}{4(1-x)},$$

where  $x$  is the degree of saturation or flow/saturation flow.

The average number of stops ( $S_i$ ) on the  $i$ th link is given by the following general expression:

$$S_i = \frac{1}{C} \sum_{t=t_r}^{t_0} A, \quad ,$$

where  $A$  is the cumulative arrivals,

$C$ , the cycle length,

$t$ , the number of intervals,

$t_r$ , the interval at the beginning of red, and

$t_0$ , the interval during green at which the queue becomes zero.

In the optimization logic of TRANSYT, a hill-climbing iteration process is used to obtain a set of optimum signal settings that will minimize the performance index. The first step of the method is to calculate the performance index for an initial set of signal timings. The next stage is to alter the offset at one of the signals by a predetermined increment of cycle intervals and to recalculate the network performance index. This is repeated until a local minimum value of the index is reached. The same procedure is then applied to each of the other signals in the network. The entire offset optimization procedure for the network is in turn repeated for a variety of cycle interval increments to obtain the final optimum settings. TRANSYT also has the capability of optimizing the splits at each signal by reallocating the green time among the various intersection approaches if the reallocation would reduce the performance index for the network. Special features are contained in the latest version of TRANSYT for treating the following cases:

- Multi-phase signals.
- The use of half cycles for minor intersections.
- Grouping of signals for maintaining constant offset differences between them.
- Multiple links at a common stop-line.

- Effects of bus progression and stoppages.

TRANSYT has been demonstrated to be reliable and effective both as a design and as an evaluation tool. In a theoretical study, the hill-climbing optimization process in TRANSYT was found to be very effective in obtaining optimum offsets. The process was very stable since a standard deviation of only about 1% was found to exist among delay results for optimum signal settings obtained from different sets of initial settings.

#### Dynamic Highway Transportation Model

One major dynamic variable that is not treated at all by either of the above techniques is that of re-routing of vehicles in a congested traffic network. The Dynamic Highway Transportation Model (DHTM) has been developed and applied to this problem (Boehne and Finkelstein, 1967, Hutchins, et al., 1969, Sandys and Hutchins, 1975). DHTM is a simulation model for use in evaluating urban traffic management techniques. The approach is unique in that it can respond to changed driver routing patterns and changed traffic demands dynamically. The simulation specifies a network in terms of numbered intersections, links, and origin-destination (O-D) zones. The maximum-size network that can currently be treated consists of 240 links, 75 intersections, 80 origin zones, and 80 destination zones. The interconnections between O-D zones, links and intersections must be specified. The demand volumes between each O-D pair must also be specified as a function of time. The traffic engineer can study the response of the network to changes in demand by altering the demand volumes used as inputs. Similarly, other inputs can be changed to study the effects of such changes. For instance, each link is described by its number of lanes, length, speed limit, and capacity characteristics; intersections are described by their signal controller types and timing, number of lanes, turn channels, lane width, detector distances, and turning radii, so the model can provide a traffic engineer with considerable flexibility in choosing those changes he wishes to evaluate.

The DHTM operates in two logical steps. In the first, alternative routes are selected between each O-D pair. There are several alternative routes from an origin to a destination, and a driver may select different routes, based on travel time, the weather, current construction projects, dangerous stretches, or aesthetics. It is not currently possible to model the effects of all these considerations. The DHTM has assumed minimum distances, volume-independent travel time, and volume-dependent travel time to be the factors affecting the selection of alternate routes.

A dynamic process is used by the model to assign vehicles to the different routes between O-D pairs. In this procedure, which is iterative, vehicles are assigned to the network; vehicle delays are then computed at the intersections and links of the

network, based on the initial assignment; then vehicles are reassigned, based on the computed delays. Assignments are made, during this procedure, by fine time period. Thus, peak demands of traffic during a fine time period will cause different traffic assignments to take place during that time period (and perhaps subsequent ones). Congestion, in the form of queues at intersections, is kept track of from time period to time period, and it influences the assignments.

The model input data that are usually the most difficult to obtain are the numbers of vehicles traveling between O-D pairs. The O-D data used to calibrate the DHTM were gathered in a cordon-line field survey, where vehicles entering the network were stopped to determine their origin and destination. More than 35% of vehicles entering a network were sampled in the calibration of DHTM. This data requirement has been overcome by modeling a network based only on field volumes on each link of the network, or by generating O-D data based on such field volumes. Where only field volumes are available, vehicles are not re-routed around congested areas. Where O-D data are generated from field volumes the O-D zones must be local or adjacent to the network to derive meaningful results.

The output from the program is a function of urban mobility, currently measured as total travel time of all users of the network. The components of this measure--the travel times for each O-D pair as a function of time of day--are available. Stopped time and time in motion can also be separated.

Other intermediate output is available to allow analysis of the causes of delay. For example, the number of vehicles allocated to each of the available routes for each time interval is recorded, together with initial demand (demand on the first iteration) for each intersection and link component. These quantities show the proportion of the population being re-routed and why, and help to identify improvements that will be the most effective in reducing congestion. Alternatively, travel times for O-D pairs and component demands as a function of time indicate where and for whom the network is failing.

#### SOME IMPORTANT UNRESOLVED PROBLEMS

In reviewing the literature and discussing the problem with well known traffic control researchers, a number of major research needs for unanswered problems have been pointed out. By far the most important of these is the need to find better methods to put the results of research already carried out into the hands of the practicing engineer. Perhaps research on this subject is needed. It is a matter for the research engineer, psychologist and political scientist to seek methods of motivation, implementation and control.

Beyond this, several other important research problems have been identified and are now briefly summarized.

### Urban Traffic Management

A great deal of research needs to be done on classifying urban traffic management problems. Currently, one tends to look at the elements of an urban traffic system individually rather than in total. Questions like: how can central responsibility for traffic management be established?; what is the form of an urban traffic management objective function?; and what methodology has or should be used?; all need to be answered.

### Centralized Versus Distributed Control

A thorough study of potential effectiveness of distributed control for traffic needs to be done. The capabilities of traffic signal controllers and small computers are increasing at an enormous rate. New traffic control philosophies and optimization techniques can now be considered both technically and economically feasible.

### Route Control

The dynamic routing of vehicles in an urban network has shown significant potential. New techniques for developing route control strategies and evaluating the results are sorely needed. The possible use of new signal displays with arrows controlling turning movements need to be considered.

### REFERENCES

- Allsop, R.E. (1968), Selection of Offsets to Minimize Delay to Traffic in a Network Controlled by Fixed-Time Signals, *Transportation Science*, 2, No. 1.
- Boehne, R.C., and R.P. Finkelstein (1967), *A Rational Approach to Urban Highway Congestion*, SRI Final Report (Project IRD 188531-150), Stanford Research Institute, Menlo Park, California.
- Brooks, W.D., *Designing Arterial Progressions Using a Digital Computer*, IBM Corporation, New York.
- Chamberlain, R.N., *Traffic Progression Program*, IBM Corporation, New York.
- Dunne, M.C., and R.B. Potts (1965), Analysis of Computer Control of an Isolated Intersection, *Vehicular Traffic Science* (Third International Symposium on the Theory of Traffic Flow), New York.
- Gartner, N. (1972), *Platoon Profiles and Link Delay Functions for Optimal Coordination of Traffic Signals on Arterial Streets*, Research Report No. 9, University of Toronto-York University Joint Program in Transportation, Toronto.

- Gazis, D.C. (1964), Optimum Control of a System of Oversaturated Intersection, *Operations Research*, 12.
- Holroyd, J., and J.A. Hillier (1971), *The Glasgow Experiment: PLIDENT and After*, Report No. LR 384 Road Research Laboratory, HMSO, London.
- Hutchins, S.H., et al. (1969), *A Low Cost Computerized Intersection Model*, SRI Project IRD 436531-808, Technical Note, Stanford Research Institute, Menlo Park, California.
- Little, J.D.C., et al. (1966), Synchronizing Signals for Maximal Bandwidth, *Highway Research Record*, No. 118.
- Little, J.D.C. (1966), The Synchronization of Traffic Signals by Mixed-Integer Linear Programming, *Operations Research* 14, No. 4.
- Miller, A.J. (1969), Practical Capacity of Intersections, in N. Clark and R.L. Pretty, eds., *Traffic Engineering Practice*, Department of Civil Engineering, University of Melbourne, Melbourne.
- Morgan, D., and D. Little (1964), Synchronizing Traffic Signals for Maximal Bandwidth, *Operations Research*, 12, No. 6.
- Organization for Economic Cooperation and Development (OECD) (1972), *Area Traffic Control Systems*, Final Report of Road Research Group T1, Paris.
- Pignataro, L.J., et al. (1973), *Traffic Control in Oversaturated Street Networks*, Final Report for NCHRP Project 3-18 (2), Polytechnic Institute of Brooklyn, New York.
- Robertson, D.I. (1969), *TRANSYT: A Traffic Network Study Tool*, Report No. LR 253 Road Research Laboratory, HMSO, London.
- Ross, D.W., et al. (1971), A Computer Control Scheme for Critical-Intersection Control in an Urban Network, *Transportation Science*, 5, 2.
- Sandys, R.C., and S.H. Hutchins (1975), *The Dynamic Highway Transportation Model (DHTM) Users Manual*, Applied Transportation Systems, Inc., Palo Alto, California.
- Schlaefli, J.L., et al. (1971), *Chicago CBD Traffic Control System, Volume 1: Systems Analysis and Design*, Stanford Research Institute, Stanford, California.
- Schlaefli, J.L. (1972), Computerized Traffic Signal Systems-A Future, *Traffic Engineering*, 42, 9.
- Traffic Control in Oversaturated Street Networks* (1973), NCHRP Report No. 3-18(2), Polytechnic Institute of Brooklyn, New York.

- Wagner, D., et al. (1969), *Improved Criteria for Traffic Signal Systems on Urban Arterials*, NCHRP Report 73, Polytechnic Institute of Brooklyn, New York.
- Wagner, D., et al. (1971), *Improved Criteria for Traffic Signal Systems in Urban Networks*, NCHRP Report 124, Polytechnic Institute of Brooklyn, New York.
- Webster, F.V. (1958), *Traffic Signal Settings*, British Road Research Laboratory Technical Paper No. 39, HMSO, London.
- Yardeni, L.A. (1965), Algorithms for Traffic-Signal Control, *IBM Systems Journal*, 4, 2.



Optimal Decongestion Techniques for  
Urban Traffic Networks\*

R. Camus, A. D'Amore, and P. Sipala

INTRODUCTION

Digital computers are now widely used for the centralized control of traffic in urban areas. They allow the adoption of sophisticated strategies for synchronizing the operation of traffic lights, as required particularly to avoid congestion during rush hours.

There are two main ways to deal with congestion in a traffic network.

In the first approach, the aim is simply to control the extreme effects of traffic saturation, i.e. the build-up of unstable queues, propagating from an intersection to the ones adjacent to it (Longley, 1968).

In the second, the objective is to remove completely all congestion from the controlled area by means of a suitable reduction of the traffic flows entering it. In this way, congestion is temporarily transferred outside the central area, where it can be more easily tolerated and absorbed (Todi, 1974).

This paper presents a method for off-line calculation of traffic light plans in congested networks, based on the second of the two approaches above.

The evaluation of signal settings is obtained in two stages. In the first stage, based on the THESEE system developed by SETRA (Le Cocq, 1973), we try to reduce as much as possible saturation within the network. In the second stage, all remaining congestion is transferred from the interior area to its outskirts by a suitable reduction of the green times for the links entering the network from outside (input links), where unstable queues are therefore temporarily produced.

The reduction of traffic entering through the input links is evenly distributed among them, and is calculated so that the overall queue length on those links is minimized. In this way, at the end of the rush period, the queues on input links will be absorbed by the network in as short a time as possible.

---

\*This work has been supported in part by C.N.R. Rome under contract no. 74 00791 07 115 5013.

MODEL DESCRIPTION

Let us recall first the main features of the traffic model used in the THESEE system.

The network is represented as a directed graph. It is assumed that cycle times, flows, saturation flows, turning ratios, journey times and maximum queue lengths are fixed in advance.

Closed traffic paths and saturated links are allowed. The optimization procedure minimizes the following total performance index:

$$\phi = \sum_1^r (d_i + 8s_i + c_i) = \sum_1^r \phi_i$$

where:

- r is the number of input and internal links,
- $d_i$ , the average delay for the  $i$ th link,
- $s_i$ , the number of stops,
- $c_i$ , an additional weighting factor depending on the queue length (it is the integral of the difference between the effective and the maximum queue length, evaluated during the time when the difference is positive), and
- $\phi_i$ , the performance index of the  $i$ th arc.

The minimum of  $\phi$  is obtained by an iterative method, divided into two stages.

In the first stage, offsets are varied in order to obtain a local minimum of

$$\phi_j = \sum_1^k \phi_i$$

for every node  $j$ , considering all performance indices  $\phi_i$  of the links entering or leaving it.

In the second stage, a further reduction of  $\phi$  is attempted by changing the green times.

In order to remove saturation from as many links as possible, the performance index of a saturated arc is expressed as the product of its saturation factor  $\tau_{sat}$  by a large constant value,  $K$ , and therefore is given a very large weight in the formation of  $\phi$ .

MODEL EXTENSIONS

Performance Index Modifications

The original THESEE model, described in the previous section, has been extended in the following way.

- (a) The weighting factor for the number of stops is no longer considered as constant, but as a parameter  $k_s$  assigned with the input data, as done by several other authors (Robertson, 1963).
- (b) The queue weighting factor  $c_i$ , as previously defined, has a negligible effect on the performance index. We consider that the effect of exceeding the maximum queue length  $q_{max}$  should be heavier and comparable to the effect of saturation.

To this purpose the performance index of each link is evaluated as follows:

- $\phi_i = d_i + k_s s_i$  for the ordinary case,
  - $\phi_i = K\tau_{sat}$  for saturated links,
  - $\phi_i = Kq/q_{max}$  for links where the queue length,  $q$ , exceeds  $q_{max}$ , and
  - $\phi_i = K\tau_{sat}q/q_{max}$  when both conditions of the last two cases are present.
- (c) Another problem is represented by the existence of preferential flow patterns. These are taken into account in the model by introducing, for each link  $i$ , weighting factors  $p_i$ , which can be used:

- to evaluate the total performance index:

$$\phi = \sum_{i=1}^r p_i \phi_i$$

allowing a green time setting more suited to the importance of each link in the network;

- to modify the inflow pattern of flow  $f_j$  (only for the calculation of  $\phi_i$ ), giving more importance to

the contributions of feeding flows  $f_i$  (Figure 1) coming from links with larger weights  $p_i$ .

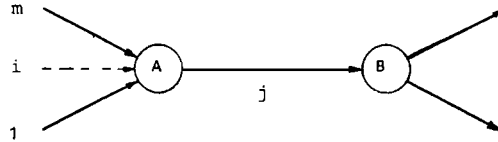


Figure 1.

For this purpose, the  $f_i$  are multiplied by coefficient  $p_i N$ , where

$$N = \frac{\sum_k^n \alpha_{kj} f_k}{\sum_k^n p_k \alpha_{kj} f_k}$$

is a normalization factor, which enables the total volume in the cycle for link  $j$  to stay unchanged. In the expression above,  $\alpha_{kj}$  is the turning ratio from link  $k$  to link  $j$ .

Evaluation of Transfer Factors

To remove all cases of saturation in the interior arcs of the network remaining after the first stage of optimization, it is convenient to evaluate the transfer factors relating internal traffic flows to the volumes entering the network.

Thus we designate, for any cycle  $c$ :

- the flows on the  $m$  input links by the vector

$$V(c) = \| v_1(c), \dots, v_m(c) \|$$

- the flows on the  $n$  internal links by the vector

$$W(c) = \| w_1(c), \dots, w_n(c) \|$$

- the turning ratio matrices by

$$A = \left\| \begin{array}{cccc} \alpha_{11} & \dots & \dots & \alpha_{1n} \\ \dots & \dots & \dots & \dots \\ \alpha_{m1} & \dots & \dots & \alpha_{mn} \end{array} \right\| \quad B = \left\| \begin{array}{cccc} \beta_{11} & \dots & \dots & \beta_{1n} \\ \dots & \dots & \dots & \dots \\ \beta_{n1} & \dots & \dots & \beta_{nn} \end{array} \right\| ,$$

where  $\alpha_{ij}$  is the turning ratio between the input  $i$  and the internal link  $j$ , and  $\beta_{ij}$  is the same ratio for internal arcs  $i$  and  $j$ .

We also assign, for every internal link  $i$ , a reduction factor  $0 < r_i \leq 1$ , such that  $1 - r_i$  indicates the fraction of incoming flow lost inside the arc for various reasons (parking, etc.); we similarly indicate by  $z_i$  the traffic volume generated at every cycle inside the arc itself (moving off from parking etc.). These factors are collected in the matrix

$$R = \begin{pmatrix} r_1 & & & \phi \\ & \cdot & & \\ & & r_2 & \\ & & & \cdot \\ & & & & \cdot \\ & & & & & \cdot \\ \phi & & & & & & r_n \end{pmatrix} ,$$

and in the vector  $Z = \| z_1, \dots, z_n \|$  .

We consider the traffic flow evolution in all internal arcs as taking place in discrete steps at every cycle. Then we have

$$w_j(c + 1) = \left[ \sum_{i=1}^m \alpha_{ij} v_i(c) + \sum_{k=1}^n \beta_{kj} w_k(c) \right] r_j + z_j ,$$

$$(j = 1, \dots, n) ,$$

or in matrix form

$$W(c + 1) = [V(c)A + W(c)B]R + Z .$$

The step response of the network, with the assumption that  $W(0) = \| 0, \dots, 0 \|$  and  $V(c) = V = \| v_1, \dots, v_m \|$  constant for all  $c \geq 0$ , is therefore given by

$$W(c + 1) = (VAR + Z) \sum_{i=0}^c (BR)^i .$$

We now observe that, for matrix B, we have

$$\beta_{ij} \geq 0, i, j = 1, \dots, n; \sum_{j=1}^n \beta_{ij} \leq 1, i = 1, \dots, n,$$

i.e. B is a nonnegative matrix whose lines never add up to a value larger than 1. In fact, for at least one line, the sum is strictly smaller than 1, since we assume that there are arcs in the network leading outside the network (i.e. exit arcs). Moreover, we shall assume that B is irreducible, in other words, we suppose that the turning ratios contained in B express the possibility of travels from every interior arc to any other.

With the above hypotheses on B, and observing that R is a positive diagonal matrix with elements not larger than 1, we can show (see Ortega, 1972, pp. 26 and 109) that the Neumann series

$$\lim_{c \rightarrow \infty} \sum_{i=0}^c (BR)^i$$

converges to the nonnegative matrix  $(I - BR)^{-1}$ , where I is the n by n identity matrix.

We therefore have for the steady state flows in the interior links

$$W(\infty) = (VAR + Z)(I - BR)^{-1}.$$

A change  $\Delta V = \|\Delta v_1, \dots, \Delta v_n\|$  in the constant input flows produces a steady state change,  $\Delta W$ , given by  $\Delta W = \Delta V [AR(I - BR)^{-1}]$ , i.e.

$$\Delta w_i = \sum_{j=1}^m d_{ij} \Delta v_j \quad (i = 1, \dots, n),$$

where

$$D = \|d_{ij}\| = [AR(I - BR)^{-1}]^T$$

is the desired matrix of transfer factors.

Desaturation of Internal Links

The transfer factors evaluated in the previous section are used at the end of the first optimization stage if saturated arcs still remain inside the network.

For every such arc,  $i$ , we consider the flow reduction,  $\Delta f$ , required to bring it within normal flow conditions (conventionally associated with a saturation factor  $\tau_{\text{sat}} = 0.99$ ).

To obtain that, the input arc flows must be subjected to reductions,  $\Delta v_j$ , satisfying the constraints

$$\sum_{j=1}^{jm} d_{ij} \Delta v_j \geq \Delta f \quad , \quad (1)$$

for every saturated internal arc  $i$ .

To avoid excessive input flow reductions we also set upper limits on the values of  $\Delta v_j$ , of the form

$$\Delta v_j \leq K_j v_j \quad , \quad (2)$$

where  $0 \leq K_j \leq 1$  are constants.

Within the linear constraints (1) and (2), we look for values of  $\Delta v_j$  minimizing the queue lengths on the input arcs. More precisely, we observe that if the input link  $j$  contains  $\ell_j$  lanes, its queue length is increased at every cycle by  $\Delta v_j / \ell_j$ . To distribute more evenly the queue lengths over all arcs, we choose to minimize the euclidean norm, i.e. the sum of  $(\Delta v_j / \ell_j)^2$ , taken over all lanes of all input arcs. Our objective function is therefore

$$F = \sum_{j=1}^m (\Delta v_j / \ell_j)^2 \ell_j = \sum_{j=1}^m (\Delta v_j)^2 / \ell_j \quad . \quad (3)$$

Again we might introduce additional importance factors,  $\gamma_j$ , for the input arcs, and therefore change (3) to the form

$$F = \sum_{j=1}^m \gamma_j (\Delta v_j)^2 / \ell_j \quad . \quad (4)$$

We are now confronted with a standard quadratic programming problem for which several solution techniques are available. In our case, the Lemke complementary pivot algorithm (see Ravindran, 1972) has been adopted.

On the basis of the values of  $\Delta v_j$  obtained as above, we evaluate the green period variations,  $\Delta t_j$ , required for each input arc.

We can have two cases:

- The signal phase controlling the input arc does not affect any other traffic flow. In this case we simply assign the green time removed from the arc to the other phases.
- The phase controlling the input arc also affects other traffic streams. In this case the reduction,  $\Delta t_j$ , is applied only to the signal group controlling the input link, leaving unchanged the other groups of the same phase (Figure 2).

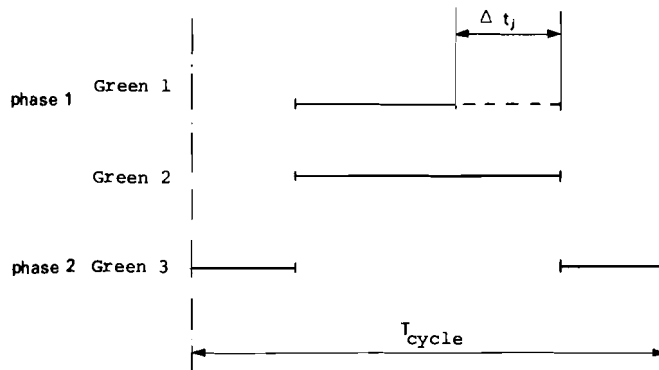


Figure 2.

Using the values of  $\Delta v_j$  previously obtained, we now evaluate the new values for  $w_i$  ( $i = 1, \dots, n$ ) by means of the transfer factors  $d_{ij}$ .

At this point we restart the optimization procedure for the green time setting. During this phase, the green times of the input links are kept fixed. For the second case above, we also set a maximum for,  $\Delta t_j$ , for the allowed green time reduction in the setting of groups pertaining to the same phase as the input arc.



IMPLEMENTATION

The method described in the previous sections has been implemented in Fortran IV on a minicomputer HP 2100A with a 32 K word memory and a disk unit; the system consists of 14 programs chained according to the scheme of Figure 3. In the present version the system allows the optimization of traffic networks having at most 50 nodes and 100 arcs.

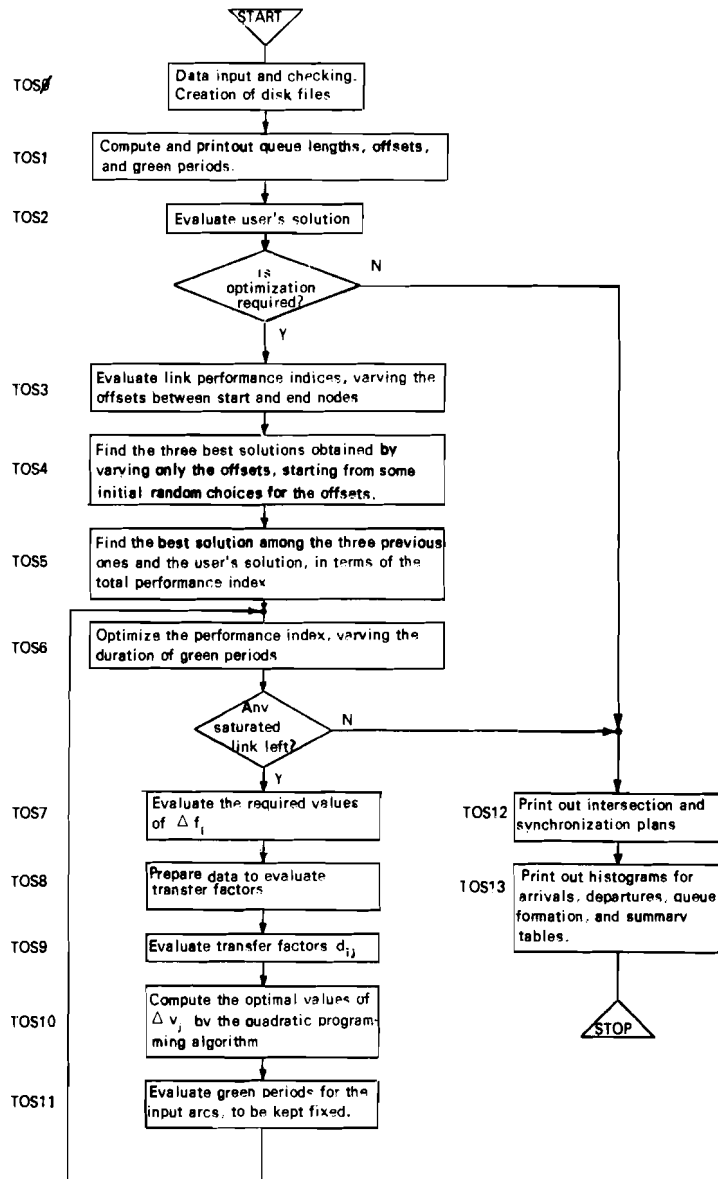


Figure 3. Flow diagram of the optimization programs.

As an example of the application of the method, we considered the network of Figure 4, representing the central part (Theresian Bourg) of the city of Trieste. The graph contains 35 nodes and 98 links.

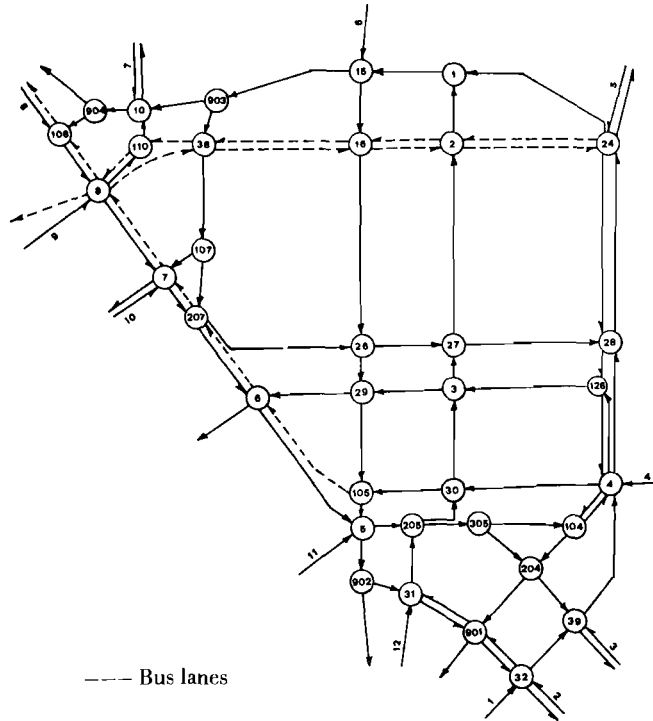


Figure 4. Traffic network of the central part of Trieste.

At the end of the first optimizations phase, the links listed in Table 1 are still oversaturated.

Table 1.

Internal arc	Source node	Destination node	Saturation factor	Flow (vehicles per cycle)	Required flow reduction	Number of lanes	Queue length (m)
30	3	1.02	35.800	1.147	2	2.867	
4	28	1.24	53.575	10.803	2	27.008	
27	28	1.05	27.075	1.480	2	3.700	
39	4	1.11	55.312	6.453	1	32.265	

To remove saturation from internal links the procedure stated earlier is used to modify flows on the input arcs as in Table 2.

Table 2.

Input arc	Saturation factor	Flow (vehicles per cycle)	Flow reduction	Number of lanes	Queue length (m)
11	1.09	30.000	1.322	3	2.20
3	1.35	39.950	8.959	3	14.93
2	1.10	18.950	1.619	1	8.10
8	1.03	33.525	1.430	3	2.38
10	1.09	16.325	1.122	2	2.81
1	1.53	12.500	4.878	2	12.20

In this way for the internal part of the network the value of the functional  $\phi$  is changed from 95,281 to 15,436 while its total value changes from 97,501 to 171,095. In other words, congestion is transferred from the network to its input links.

We observe in particular that, in the original situation, there existed inside the network an unstable queue with a total length of 65,840 m and a maximum value of 32,265 m per cycle. After using the procedure for internal desaturation, unstable queues exist only on the input arcs with total length 42,620 m and maximum value 14,93 m per cycle.

In Figures 5 and 6 we give two samples of the program outputs. The first one presents the signal plans obtained for the nodes controlling the input arcs. In Figure 6 we present an example of signal synchronization on an internal path of the network.

Finally in Table 3 we give the execution times of all programs for the example of Figure 4.

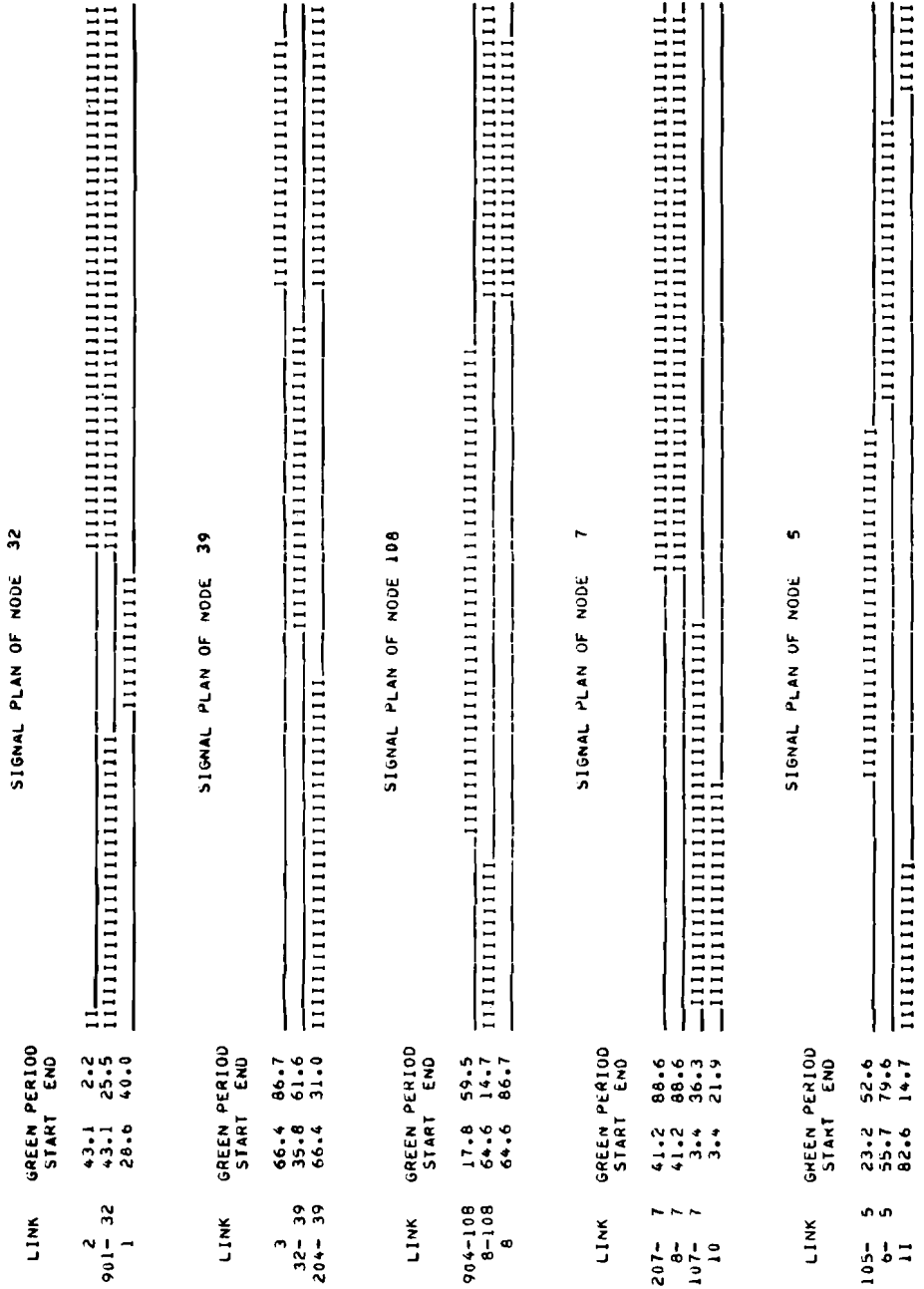


Figure 5. Signal plan for nodes controlling the input arcs.

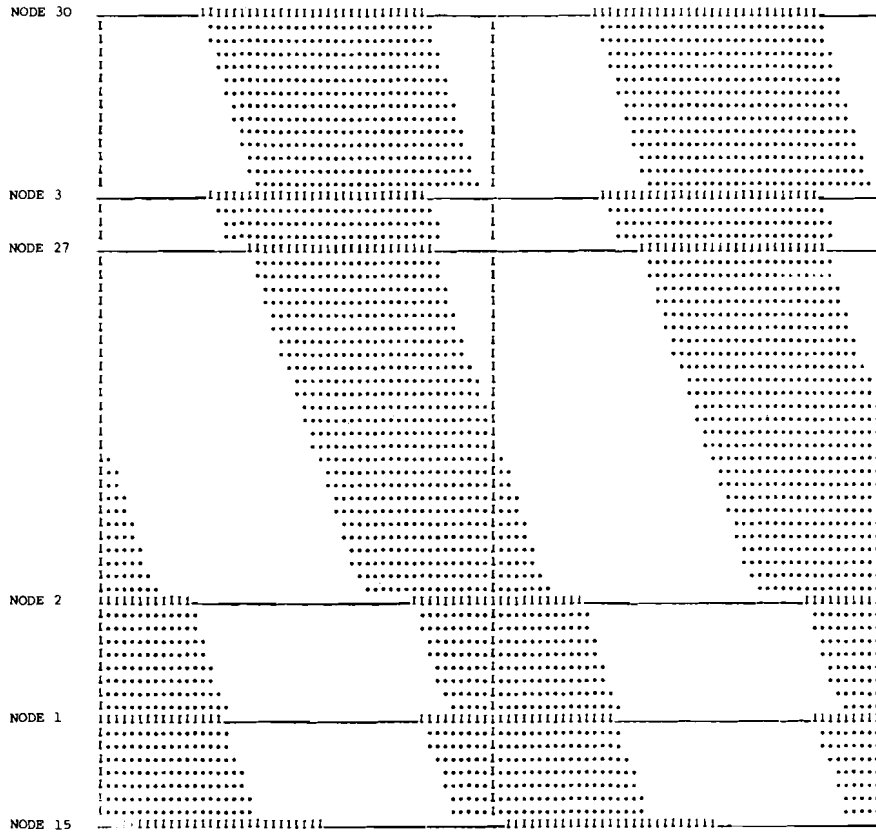


Figure 6. Example of synchronization.

Table 3.

PROGRAM	EXECUTION TIME
TOS0	1'14"
TOS1	6"
TOS2	1'23"
TOS3	2'33"
TOS4	13'14"
TOS5	3'50"
TOS6	23'34"
TOS7	27"
TOS8	3"
TOS9	3'45"
TOS10	30"
TOS11	38"
TOS12	depending on the number of prints required
TOS13	depending on the number of prints required

### CONCLUSIONS

The method proposed here for the decongestion of interior links of a network is essentially based on the evaluation of the transfer factors  $d_{ij}$ , which have been derived from the turning ratios, since these are easily measurable.

Obviously the resulting transfer factors provide a good approximation to the reality if the turning ratios are the same for all streams entering each link. If this is not the case the factors may differ considerably from the theoretical values.

To reduce the amount of error, it should be possible to extend the proposed method to take into account turning ratios measured on several adjacent nodes simultaneously.

### REFERENCES

- Le Cocq, J.P. (1973), *Régulation de la circulation, Revue Générale des Routes et des Aérodrômes*, no.6, June.
- Longley, D. (1968), A Control Strategy for a Congested Computer Controlled Traffic Network, *Transportation Research*, 2.
- Ortega, J.M. (1972), *Numerical Analysis: A Second Course*, Academic Press, New York.
- Ravindran, A. (1972), A Computer Routine for Quadratic and Linear Programming Problems, *Comm. ACM*, 15, 9, 818-820.
- Robertson, D.I. (1963), *Transyt: a Traffic Network Study Tool*, Report LR 253 RRL, HMSO, London.
- Todli, J. (1974), Contribution to the Problem of Urban Road Traffic Control During Rush Hours, Paper presented at IFAC/IFIP/IFORS 2nd Symposium, Monte Carlo.

Design of Control Systems in  
Automated Transport Systems

L.D. Burrow

Control of automated transport systems involves many interacting operations. People have to be informed, guided and regulated. Vehicles have to be maneuvered, directed and dispatched. Failures and faults must be identified and rectified. Safety must be ensured.

Many of these aspects have been extensively studied, often with optimization in mind, yet, when extended to whole system operations, most schemes do not perform well. Either necessary vehicle maneuvers cannot be easily performed, or the system response to fault conditions is inadequate, or unstable modes of operation appear.

Operating schemes are required which will enable the system to operate well under all practical conditions. In complex systems, governed by cost functions embracing qualitative and quantitative economic, social and technical factors, design policies must attempt to find the best operating regions.

To aim for the global optimization of such intricate systems is unrealistic. Even in the event that an accurate mathematical description of the cost function and system could be produced, the complex, highly constrained, non-linear interaction of variables is certain to defy solution. At best only local optima can be found and by careful design, combined to form an overall "good" system.

The benefits which accrue from a judicious design of the control structure far outweigh those that can be achieved by optimization at a detailed level.

This paper will consider a systematic approach to the design process which may help more effective control systems to be evolved.

Good reliability and high safety standards are fundamental factors in any transport control scheme and must figure in any cost function relating to the operation of the whole system. The paper will survey the response of a transport scheme to failures, outlining the requirements for a "fail soft" system and discuss the use of hierarchical structures to achieve such a characteristic.

OPTIMIZATION

The process of optimization is often presented as a highly exact process. Yet, if optimization is taken in the general sense of meaning the systematic approach to a situation with a view to obtaining the best possible outcome with the use of what knowledge is available, then only in a few situations is this true.

To gather the information necessary to decide upon an improvement takes time. Better forecasts require more time. Optimization processes cannot work faster than the systems they are trying to improve. Consequently the evolution of good transport schemes may take several decades, whereas the on line optimization of parameters in a vehicle controller may take only seconds.

DESIGN

The creation of a "good" system is part of an optimization process. The designer, by assembling together his experience, attempts to create a new system whose properties more closely approach the design specification.

An important part of the design process is the accurate specification of the design environment or a definition of all the influences on the system of interest. These include disturbances for which the design must cater, criteria, fixed information and measures of performance (see Figure 1).

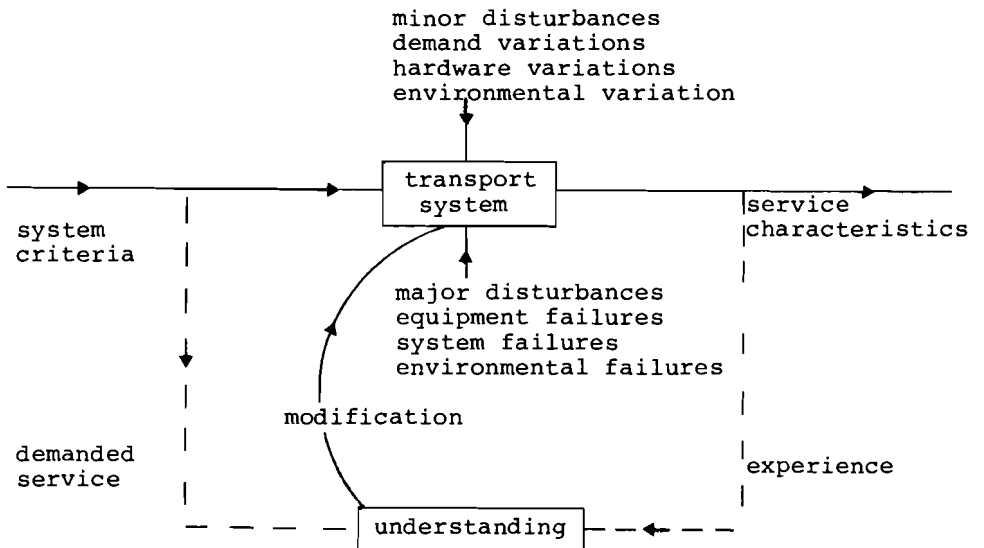


Figure 1.



There are few direct design precedents for automatic transport control systems. The feedback link labeled "experience" is weak. Nevertheless a good design process will make the maximum use possible of what transferable experience exists.

Designs can be evolved at three levels: structure, subsystem, and parameter or equipment.

### Structural Level of Design

A system can be considered as a structure of interconnected subsystems. Potentially every subsystem is itself a system. The original system is a subsystem in a more general system. The design problem is limited by the designer. A control engineer will take as fixed the transport policy determining the particular niche his system will fill. Similarly he will take as fixed the range of components available for use in his circuits. Effectively an upper and lower boundary to the problem has been prescribed.

A system is typically much more complicated than a man can overview - only a piece at a time can be considered - so a set of subsystems has to be defined. The most general level of design defines the system organization. It specifies the most appropriate subsystems and structure to achieve the desired "whole" system properties.

The choice of subsystems in a system is determined by several factors. Some subsystems are immediately apparent as they correspond to necessary functional units in the system. Junction controllers, signalling systems, emergency backup systems are all possible units in a transport control structure.

The choice of subsystems may reflect a degree of complexity, related to the ability of one or more people to fully understand it within a given time. A unit too large to be understood is unlikely to perform well and when it fails will be time-consuming to repair and probably too big to replace.

A subsystem may be chosen because it corresponds closely to an already developed scheme, so reducing the design effort required.

The choice of a structure for a system is less obvious. Some work exists on the theory of structures (Mesarovic, 1970; Langfors, 1971), but generally the choice of an appropriate structure can only be made on the basis of comparison with other systems exhibiting desirable properties. Direct solutions may not be found but the comparison may constitute some demonstration of feasibility.

Likely control structures for an automatic transport system are either centralized or hierarchical (Black, et al., 1975).

In centralized control structures, a central decision-making unit controls all the peripheral subsystems. Information from the subsystems passes to the central office and is available for use in any other subsystem.

Communication costs are high and the centralization of control makes the system very vulnerable to faults.

Well understood centralized control structures can probably offer a better level of control by using all the system information. However the complexity of interactions between subsystems makes the system less easy to understand.

This has two effects:

- The system becomes more prone to software faults. An incomplete specification of subsystem states is more likely and may lead to undefined unsafe conditions. The greater number of subsystem states makes fault monitoring and rectification more difficult and costly.
- The greater number of feedback loops tends to increase the chance of unstable system responses. This forces lower gains to be used and results in a poorer control action.

In a hierarchical structure, control is divided between a number of semi-autonomous levels. Hierarchy decouples elements of a system. Each element in the tree is an autonomously functioning subsystem which uses only limited strategic information from the level above. Frequently this information can be transmitted discontinuously. Communication is in two directions; a command or parameter down the hierarchy specifying what should happen, and a feedback or check up the hierarchy saying what is happening.

Hierarchical organization reduces the number of unwanted feedback loops in the system, so allowing the interaction of subsystems to be more confidently predicted.

Hierarchies show a graduation in properties which are summarized in Figure 2.

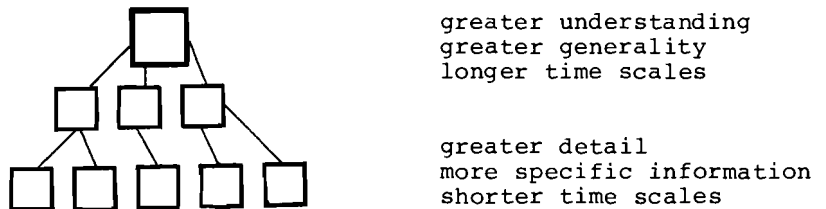


Figure 2.

A hierarchy can be considered as a filter, each layer being concerned only with a range of frequencies. Together the subsystems cater for the entire range of frequencies apparent in the system.

#### Subsystems Level of Design

To a subsystem, the rest of the system is its environment. Where the subsystem is designed in isolation, as often will be the case, its interface or connections with the outer world have to be accurately specified, otherwise incompatibilities will arise.

The designer of a subsystem wants to minimize his own particular cost function. This will generally be achieved at the expense of the outside system. The balancing component from the outside must be made visible to the subsystem so that an overall balance can be achieved, i.e. the subsystems should be given boundary conditions suitable for approximating the total optimization.

The use of simulation is often an appropriate aid to the design of a subsystem. Simulation is a means of modelling approximately the important system interactions at an accelerated time scale. By investigating a large number of specific situations a more complete picture of the process is built up, hopefully enabling better solutions to be found.

Computer simulations have been extensively used in the analysis of transport control systems, particularly for network design studies, vehicle management and operation strategies (Burrow, 1975; York, 1973; and Heap & Thomas, 1974 give a selection of representative work in this field).

#### Parameter Optimization

At the level of the interaction with the real world, parameter optimization can frequently be approached mathematically, although in transport control the many parameter constraints and non-linearities prevent general solutions from being found and recourse has to be made to iterative techniques.

In some circumstances an optimal solution to the problem can be found at the design stage and incorporated into the system hardware. Alternatively the designer can structure the hardware in such a way as to allow optimization to take place "on line". This may lead to a better control but at the cost of added complexity of equipment, measurements and communications.

Dynamic optimal controllers are often proposed for vehicle position and speed controllers while merge control algorithms may well be optimized at the design stage (Athans, et al., 1966; Chu, 1974; and Athans, 1969).

"FAIL SOFT" TRANSPORT CONTROL

In the design of large complex transport systems the quality of service is strongly influenced by the reliability of the system and its ability to cope with faults as they occur. Reliability of individual components can be assured up to some limit, but failures will however still occur. It has been estimated that given reasonable standards of reliability for a medium sized auto taxi system a failure can be expected somewhere every couple of minutes. A system which is very sensitive to faults is going to be at a severe disadvantage. Fail soft systems can be defined as systems which, as failures occur, progressively become degraded in performance, rather than collapse completely. The design of a control structure to have this sort of property is a "black art" for which no systematic approach appears to have been developed.

By the systematic application of the standard techniques of reliability, standby and redundancy, to all levels of the design process, to the structure, subsystems and equipment, it is hoped that the effects of a fault can be minimized, its zone of influence circumscribed and its duration minimized, so leading towards the creation of a fail soft system.

Failures of a transport system cause disruption of service and often create unsafe situations. To ensure the safe running of a system requires two control systems. One the normal running control, the other an independent safety control system. The latter oversees the former and is generally a simple controller activated by the single condition "is the vehicle separation inadequate for the speed of the vehicle?" and issuing one command (e.g. brake at an emergency rate to zero velocity). The safety control system must be independent of the normal control system to ensure that failures in normal control are independent of failures in the safety control system, thus reducing the likelihood of a joint and possibly catastrophic failure.

The disruption caused by a fault is particularly dependent upon the severity of the fault. This severity depends on the area of the system affected, the subsequent propagation of the fault through the system and the time duration of the fault.

All these factors are made less significant by designing subsystems to operate as independently as possible over localized regions of the track.

This independence, necessary also for backup safety systems is an intrinsic property of hierarchical structures. Hierarchy allows structures to expand or contract locally without influencing the remainder of the system. The modular nature of such systems reduces maintenance and repair times by simplifying the detection of faults and their repair.

## CONCLUSIONS

The design of a complex transport control system must integrate all the facets of a transport scheme. Good system availability, safety and fail soft characteristics are especially difficult to design into a system, yet are fundamental to its operation.

Hierarchical structures are more readily broken down and understood. They appear to offer characteristics which allow effective designs to be evolved. Other structures may allow better results to be achieved but probably at the cost of much greater design effort.

## REFERENCES

- Athans, M. (1969), A Unified Approach to the Vehicle Merging Problem, *Transportation Research*, 3, (April) 1.
- Athans, M., et al. (1966), *A System for Optimal and Sub-Optimal Position and Velocity Control of a String of High Speed Vehicles*, Report ESL-R-291 MIT, Cambridge, Mass.
- Black, I.G., et al. (1975), *Advanced Urban Transport*, Saxon House, Lexington, Mass.
- Burrow, L.D. (1975), *The Simulation of Junctions in Automatic Urban Transport Systems Using Interactive Graphic Displays*, paper presented at the Conference on Computer Simulation, United Kingdom Simulation Council, Windermere, 6-8 May.
- Chu, K. (1974), Optimal Decentralized Regulation for a String of Coupled Systems, *IEEE Trans. on Automatic Control*, AC-19, 3.
- Heap, R.C., and T.H. Thomas (1974), Performance of Automatic Transport Systems Under Fixed Schedule Operation, *Proceedings I.E.E. Conference on Control Aspects of New Forms of Guided Land Transport*, London, August.
- Langfors, B. (1971), *Theoretical Analysis of Information Systems Vol. 1.*, Studentlitteratur, Lund, Akademish Forlag, Copenhagen.
- Mesarovic, M.D., et al. (1970), *Theory of Hierarchical, Multi-level Systems*, Academic Press, New York.
- York, H.L. (1973), Simulation of a P.R.T. System Operating Under Quasi-Synchronous Control, *Proceedings of 2nd International Conference on P.R.T.*, University of Minnesota, Minneapolis, Minn.



Optimal Control of Railway  
Marshalling Yard Operations

J. Sokolowski

The problem of traffic planning in marshalling yards (MY) is one of the most difficult in railway transport. These difficulties arise not only from their complexity and size and the many processes taking place in them, but also from the necessity of collecting the vast amount of information required for the planning of traffic operations in the MY. This information must be analyzed and processed so that an object is correctly described, and so that the dynamics of processes which take place in the yard are recognized.

This paper presents a method of building a model for construction of an EDP system which should enable us to produce automatically optimal patterns for conducting traffic operations in the MY.

These patterns will be of dual use:

- as a basis for efficient control of operation in the MY, and
- as a point of reference for the evaluation of MY operations.

The proposed method is based on operational research, statistical inference, and system theory.

OBSERVATION OF OBJECT AND RESULTS

The MY is the lowest level of management in a general pattern of freight transport management. An initial analysis shows that the MY is both an elementary subsystem in a freight transport management system, and a complex system in its own right. Both for observation and for analysis of these types of system, action towards a gradual dismembering of the whole system is required. Many different observation levels and tasks are created, so that each observer at his particular observation level can describe an object, or its element, to a certain degree of accuracy in an appropriate language.

A general picture of MY observation, which includes assumptions and requirements described above, is shown in Figure 1. The tasks for the particular observers are:

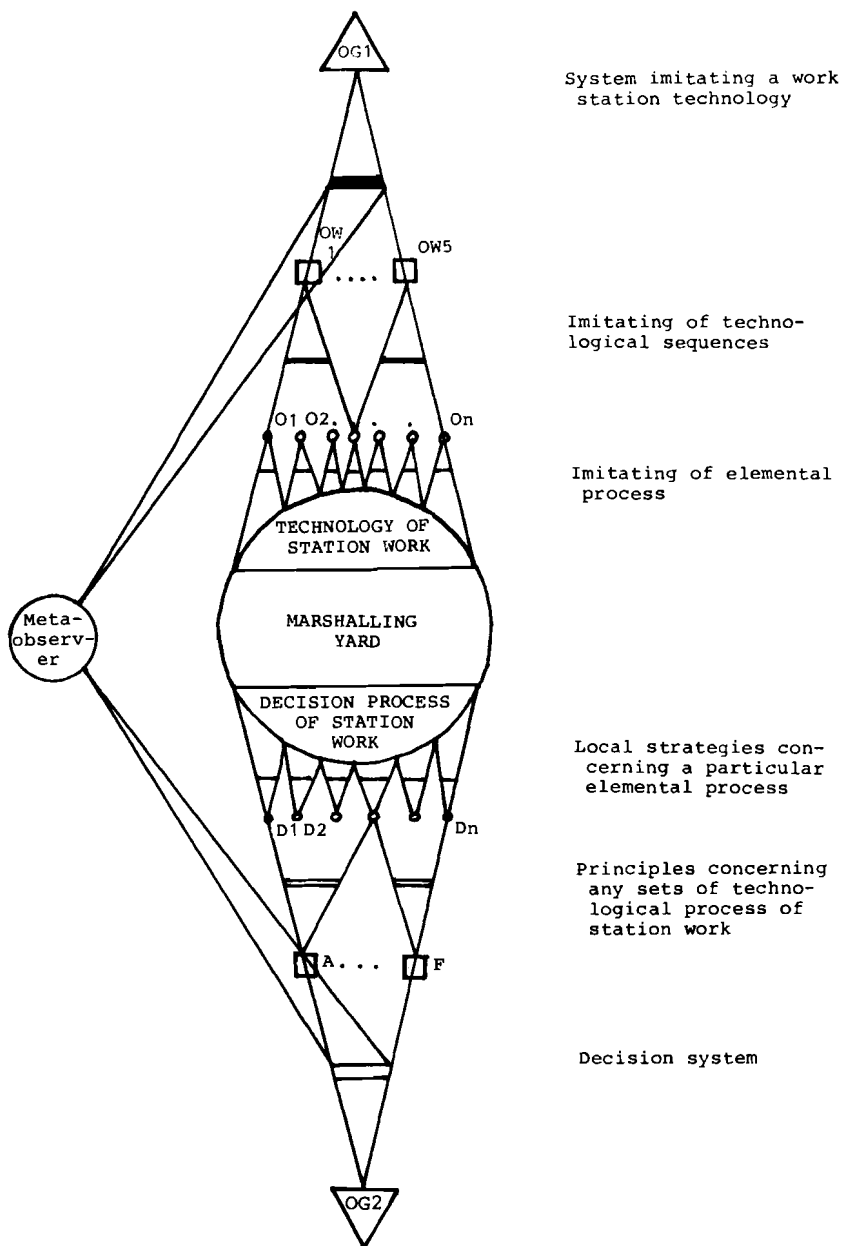


Figure 1. General principles concerning observation in a marshalling yard.



- A metaobserver, watching the whole complex system, selects particular systems as well as interactions and interconnections between them.
- Highest level observers, OG1 and OG2, responsible for the particular systems, select existing technology sequences, logically arrange process sets and decision principles, with the interactions and interconnections existing between them.
- Medium level observers, OW1-OW5 and A-F, who watch the particular technology sequences and decision principles, and select within them the processes, interactions and interconnections between them.
- The lowest level observers, O1-On and D1-Dn, watch the particular elemental processes, and the interactions and interconnections existing between them.

All these observers analyze the material collected and then make a synthesis and describe the investigated object as a system or as its elements. Here, only the observation results of the highest level observers, OG1-OG2, and the metaobserver are presented, and the framework of the model.

Observer OG1 watching and analyzing the technology of the MY activity (see Figure 1), and using the results of the lower level observers, described it as a technology system,  $S_T$ , which is presented as a well-ordered 4-tuple (see Figure 2):

$$S_T = \langle I_T, P_A, P_W, P_B \rangle ,$$

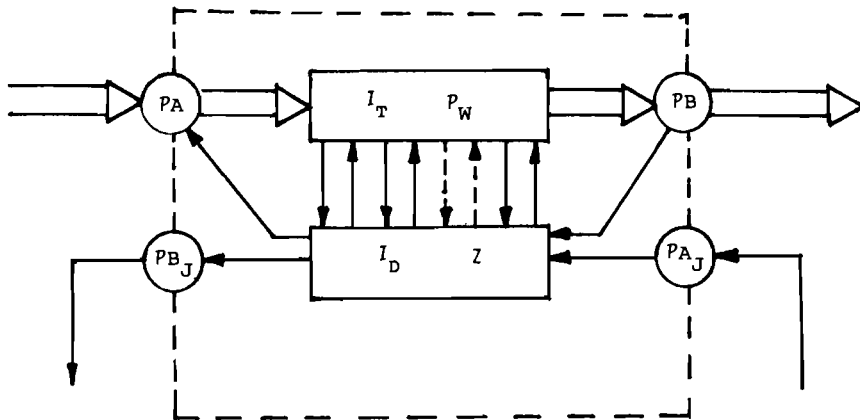


Figure 2.

where

$I_T = \{\langle O^b, A^b \rangle\}$  is the MY technical structure,

while  $b = 1, 2, 3, \dots, B,$

$O^b$  is the name of object  $b$ , and

$A^b$  is a set of attributes which describe object  $b$ .

By a set of objects  $O = \{O_S, O_D\}$  we mean a set consisting of two elements

$O_S$ , a set of objects of a fixed nature, and

$O_D$ , a set of objects of a dynamic nature,

where

$O_S = \{G_T, D_O\}$ , a two-element set consisting of

$G_T$ , a set of tracks, crossovers and crossover sets, and

$D_O$ , a set of service teams with equipment.

Similarly it describes a second element of set  $O$

$O_D = \{P_O, P_R, P_T, P_{ZS}, L_L, S_O, S_R, S_T, S_{ZS}, G_W, O_W\}$  ,

where

$P_O, P_R, P_T, P_{ZS}, L_L$  means: passenger train's set, marshalled goods train's set, transit goods train's set, completed goods train's set, locomotives,

$S_O, S_R, S_T, S_{ZS}$  means: passenger car's set, marshalled goods car's set, transit goods car's set, completed goods car's set,

$G_W$ , set of car groups, and

$O_W$ , set of disconnected cars.

An attribute set A has been described in the same way as an object set  $\theta$ . This attribute set characterizes its particular elements.

$PA = \{PA_1, PA_2, \dots, PA_L\}$  is a process set, describing a stream of trains entering an MY, starting from  $A_1$  and ending at  $A_L$ , with all interactions and interdependences between them,

where L is the number of entries to the MY.

This set determines the influence of objects bordering an MY;

$P_W = \{P_1, P_2, \dots, P_{12}\}$  is a set describing all internal technological processes with influence on all the elements from set  $\theta_D$  inside the MY, and interdependences and interactions existing between them.

Every elemental selected process can form an operation set which may contain one or many elements (see Figure 4).

$PB = \{PB_1, PB_2, \dots, PB_M\}$  is described similarly to the set PA, but concerns an exit stream from the MY, i.e. the MY influences bordering objects.

Technological systems in an MY, described in this way, are realized by the criterion:

$$F_T = \left( \sum_{i=1}^{I^W} \left( \left( \sum_{j=1}^{12} \text{toij} \right) + \left( \sum_{\ell=1}^L \text{toil}^{PA} \right) + \left( \sum_{m=1}^M \text{toim}^{PB} \right) \right) \right) \rightarrow \min ,$$

where

$I^W = \{1, 2, \dots, I^W\}$  is the set of all cars processed by an MY, and

$J = \{1, 2, \dots, 12\}$ , the set of indices determining all internal technological processes in an MY, which belong to set  $P_W$ ,

$\text{toij}$ , the time of operation connected with i car (depending on applied variant in j process), and

$\text{toil}^{PA}$  and  $\text{toim}^{PB}$  are operation times, connected with i car in processes belonging to sets PA and PB (depending on  $\ell$  arrival direction or m dispatch direction and train's category and the kind of traction).

Observer OG2 (see Figure 1) who watches and analyzes a decision area of an MY, utilizing the observations from lower levels, describes his area as a decision system  $S_D$ . This system is presented as a well-ordered 4-tuple (see Figure 2):

$$S_D = \langle I_D, PA_I, Z, PB_I \rangle$$

where

$I_D = \{\langle F^e, C^e \rangle\}$  is the decision structure of the MY,

while  $F^e$ , a name of that which needs deciding,

$C^e$ , an attribute set that describes  $e$ ,

$PA_I$ , an information set that contains the data describing timetable schedule, part IV schedule addition, technical limitations of traffic, transport policy, etc.

These data characterize the work of a decision superstructure for the MY which manages a railway network or part of it.

$Z = \{Z_1, Z_2, Z_3, Z_4\}$  is a set of interior decision processes existing in the MY and realizing the strategic goals of traffic work in the MY. They have been formed by the decision superstructure.

$Z_1$  is a set that determines the interior decision processes connected with the acceptance of trains in the MY (the decision principle is presented in subprogram PD3 shown in Figure 5);

$Z_2$ , a set that determines the interior decision processes on the control of train sets for the operations, and connected with their preparation for marshalling (PD3);

$Z_3$ , a set that determines the interior decision processes concerning the control of train sets for marshalling (PD2);

$Z_4$ , a set that determines the interior decision processes concerning the flow of dynamic entries in technological sequences (e.g. "composing", "transit") and in an interior technological process  $P_{11}$  - "train dispatch from the MY" (PD1).

$PB_I$  is an information set containing data that:

- describe the planned operations in the MY with predetermined entry and exit streams ("MY technological work process", "car flow plan", "car dispatch plan"), and
- describe the proposed changes in an actual schedule for a range of train traffic.

Generally this set characterizes the MY influence on a decision superstructure.

Such an MY decision system realizes its purpose based on the following criterion:

$$F_D = X \uparrow H ,$$

where

$$X = \left( \frac{K}{K_P} \right) \rightarrow 1 , \quad H = \left( \sum_{i=1}^{I^W} t_{pi} \right) \rightarrow \min ,$$

$$0 \leq X \leq 1 .$$

The notation:

X is a coefficient of realization of the established compilation plan for the planned trains for an established planning period;

$H = \sum_{i=1}^{I^W} t_{pi}$  is the total time of interoperation breaks for all cars which stand in an MY and on its approaches in a particular planning period;

$$\sum_{i=1}^{I^W} t_{pi} = \left( \sum_{i=1}^{I^W} \left( \left( \sum_{j=1}^{12} t_{ij} \right) + \left( \sum_{\ell=1}^L t_{i\ell}^{PA} \right) + \left( \sum_{m=1}^M t_{im}^{PB} \right) \right) \right) \rightarrow \min ,$$

where  $t_{ij}$  is the interoperation break time concerning  $i$  car in  $j$  internal technological process (this time contains a time of waiting on a following internal technological process, which will react on any dynamic object, containing  $i$  car) and  $t_i^{PA}$  and  $t_i^{PB}$ , are as  $t_{ij}$  with the difference that these times concern the processes from seats PA, PB;

‡ is an operator denoting that the formula on the left side must be calculated before the formula on the right; the best solution to the right hand side formula is the problem's solution, and is a limitation in further optimization investigations;

$K_p$  is the set's quality of  $N_{zp} = \langle P_z, 1p, P_z, 2p, \dots, P_z, k_p, \dots, P_z, K_p \rangle$ ; these well-ordered sets of trains  $P_z, k_p$ , containing precisely determined attributes in accordance with part IV additions to the schedule, must be sent in the MY in a precisely determined period of time in accordance with the timetable schedule;

$K$  is the set's quality of  $N_z = \langle P_z, 1, P_z, 2, \dots, P_z, k, \dots, P_z, K \rangle$ ; these well-ordered sets of trains with precisely determined attributes can be sent by the MY in a precisely determined period of time in accordance with existing technical-traffic conditions.

In other words, elements of set  $N_z$  must accomplish the following requirement:

$$P_z, k = \begin{cases} P_z, k_p, & \text{when } tk = tkp \pm \Delta tkp, \text{ and } k = kp \\ 0, & \text{w other case} \end{cases}$$

while  $P_z, k_p \in N_{zp}$  ;  $k \in K = \{1, 2, \dots, K\}$  ,  
 $kp \in K_p = \{1, 2, \dots, K_p\}$  ,

where

$tkp$  is the scheduled departure time of  $kp$  train,

$tk$ , the planned departure time of  $k$  train,

$kp$ , the number of trains from set  $N_{zp}$ ,

$k$ , the number of trains from set  $N_z$ , and

$\Delta tkp$ , the admissible departure time deviation of a planned train from the scheduled time.

A metaobserver, analyzing these observations, introduces the following additional factors:

$I = \{I_T, I_D\}$  is a set determining an MY structure;

$P_{WE} = \{PA, PA_I\}$  is a set determining all those processes and information which describe the influence of the environment on the MY - in other words, an entry stream to the system;

$P_{SR} = \{P_W, Z\}$  is a set of interior processes, acting in MY in both technological and decision areas;

$P_{WY} = \{PB, PB_I\}$  is a set determining all those processes and information which describe the MY influence on the environment - in other words, an exit stream from the system which describes the MY as a well-ordered set of four

$$S_{SR} = \langle I, P_{WE}, P_{SR}, P_W \rangle \quad (\text{see Figure 3}) \quad .$$

This will be realizing its purpose of performing the criterion

$$F_{SR} = X \uparrow H \uparrow F_T \quad .$$

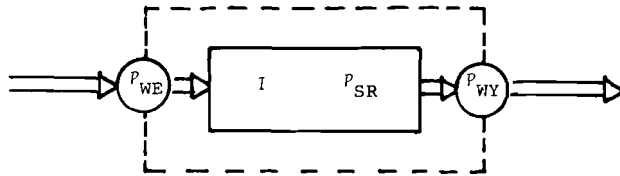


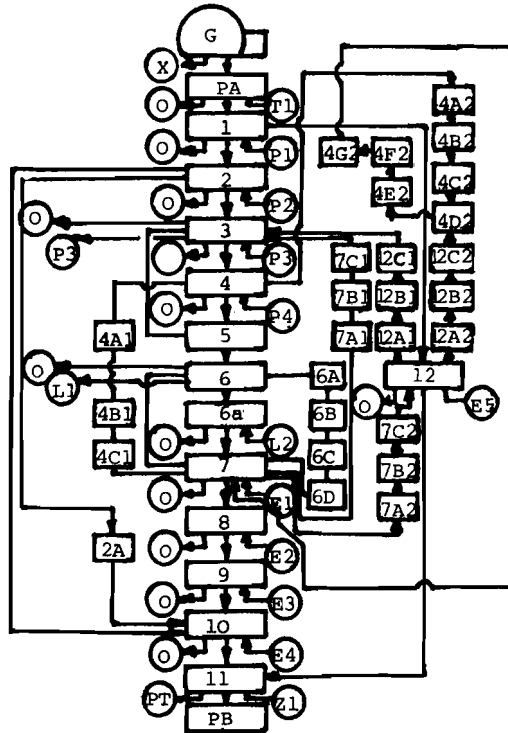
Figure 3.

PRINCIPLES OF DESCRIPTION OF ELEMENTAL PROCESSES AND GENERAL MODEL STRUCTURE

Investigation of the system described above is based on observation of variations of its state, i.e. observation of the process set  $P_{WE}, P_{SR}, P_W$  and its variations in a set  $I$ .

The system of traffic in an MY has been modeled both in the technological area (see Figure 4) and in the decision area (see Figure 5) in order to obtain the required information. The model, overall, can be defined as an MY information set, that permits investigation of its behavior with complex traffic tasks within existing limitations and criteria.

In order to characterize the model more precisely for fixed process sets:

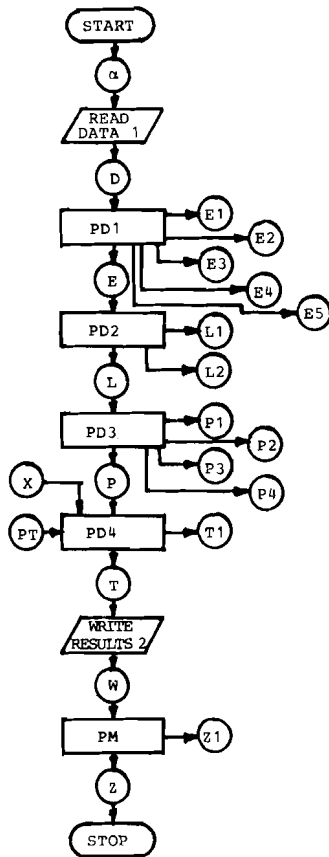


Notation:

- 1, acceptance of a train at an MY
- 2, disconnection of a locomotive
- 3, processing of a car set in an arrival group
- 4, separation from a car set of non-marshalled cars and damaged cars
- 5, pushing a car set up an elevation
- 6, car marshalling
- 6a, pushing of cars in a direction in a group
- 7, cleaning of cars and compilation of car set
- 8, flow of car set or car group
- 9, technical-commercial briefing in a departure group of car set with stationary equipment, enabling realization of brake test
- 10, connecting a locomotive to a car set and checking a brake set, with the use of a locomotive
- 11, departure of a train from an MY
- 12, processing of transit trains in a transit group

Figure 4. Block diagram of the technological model of a marshalling yard.





READ DATA 1, plan of trains and time-table schedule

PD1, subprogram analyzing state of elemental processes from 12 to 7 and the controlling of these elemental processes

PD2, subprogram analyzing state of elemental processes from 6 to 5 and the controlling of these elemental processes

PD3, subprogram analyzing state of elemental processes from 4 to PA and the controlling of these elemental processes

PD4, subprogram describing a moment of simulation end for determining the range of planning and controlling the entry data file

PM, subprogram modifying plan of trains and timetable schedule

WRITE RESULTS 2, plan of train departure (in determined range of planning of MY work)

Figure 5. Block diagram of essential decision processes for a model of technological processes of station work.

(A) Reflecting the technological area of MY traffic, one must determine every process and dependence of an MY, so that a description of variations in state over a period of time can be made.

In modelling analyzed processes these are treated as discrete. Process states are presented in such a way that moment  $t$  and state  $s$  at moment  $t$  can describe moment  $\bar{t}$ . After moment  $\bar{t}$ , the process state changes from  $s$  to  $\bar{s}$ . We have to determine the following transformations:

$$\tau : (t, s) \rightarrow \bar{t} \quad \text{and} \quad \phi : (t, s) \rightarrow \bar{s} .$$

These are determined on a set calculated as the product  $TS$ , where  $T$  is a set of numbers determining moments, and  $S$  is a set of numbers determining possible process states, with the following criterion:

$$t \leq \tau(t,s) = \bar{t} .$$

Transformations determined in this way describe principles of the analyzed process, and lead to precise algorithms simulating this process. As an example, Table 1 shows some described transformations  $\tau$  (principles determining the operation time for some processes from the set  $P_w$ ) determined for a particular MY.

Naturally, the problems of transformation of elemental process mechanisms (determined as  $\tau$  and  $\Phi$ ) into algorithms that simulate them, is complex and has not been included in this paper.

(B) Reflecting the decision area of MY traffic, we will establish a series of every elemental process taking place in every mentioned set.

Then

$$z_{ij} = \langle R_{ij}, w_{ij} \rangle ,$$

where

$R$  is a set of allowed solutions, and

$w$ , a set of methods, of which there are many, enabling a change of optimal or suboptimal solution for:

$i$  set of processes ( $i = 1,2,3,4$ ) and

$j$  decision process ( $j = 1,2,\dots,J$ ) .

THE EDP SYSTEM DEVOTED TO MEDIUM-TERM PLANNING OF TRAFFIC OPERATIONS IN A MARSHALLING YARD (SPO-RS)

If we analyze the tasks of an operations planning system, and information requirements of model construction for MY traffic work, we achieve the sort of block diagram SPO-RS shown in Figure 6.

Table 1. Operation time in an investigated marshalling yard.

Elemental process number	Elemental function	Transformation		Test number					Average from test
				I	II	III	IV	V	
1	2	3	4	5	6	7	8	9	10
1	Arrival of a train at a station	$\tau_1$	Set of values of scheduled arrival time of a train at a station						
2	Disconnection of a locomotive from a train	$\tau_2$	Regular distribution of probability over a period of time of 3-20 min						
3	Technical processing	$\tau_3^1$	average value	31.0	35.0	38.0	31.0	28.0	33.0
			standard value	15.0	16.0	12.0	13.0	13.0	14.0
	Commercial processing	$\tau_3^2$	average value	17.0	18.0	20.0	19.0	20.0	19.0
			standard value	6.0	7.0	6.0	7.0	6.0	7.0
5,6,6a	Marshalling	$\tau_5$	average value	55.0	49.0	-	43.0	51.0	49.0
			standard value	15.0	15.0	-	17.0	20.0	17.0

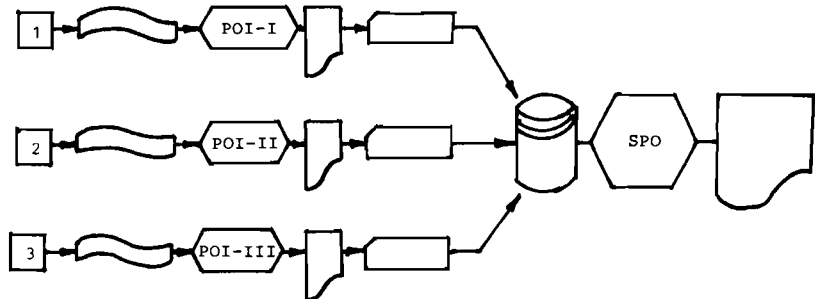


Figure 6.

1. Source information:

- about a set and MY equipment;
- about work organization - courses and other local MY strategies;
- about timetable schedules and part IV additions;

2. Source information: about the time of particular elemental processes and operations in the MY

3. Information:

- about the actual arrival of trains on the MY--(time, marshalling cards);
- about the actual departure of trains from MY--(time, list R-7).

The subsystem presented in Figure 6 is used as follows:

POI-I information processing subsystem processes information from source documents describing MY structure, timetable schedule, and part IV schedule additions, and enables one to form a set  $\theta_s$  and a set of its attributes, then sets  $P_O, P_R, P_T$  and sets of their attributes, and then set  $P_{AI}$ ;

POI-II source information processing subsystem processes information from a traffic documentation process so that it is possible to determine every transformation  $\tau_j$  for every  $j$  process from a set  $P_W$ , and transformations  $\tau_{PA}, \tau_{PB}$  for processes  $P_A$  and  $P_B$ ;

POI-III source information processing subsystem processes information from traffic documents so that the sets  $P_O, P_R, P_T$  and sets of their attributes can be formed (the commonest train deviations from timetable schedules and from part IV additions are thus considered).

SPO operation planning system is used for periodic planning of MY traffic work. This system utilizes information processed by subsystems POI-I and POI-III and the model reflecting MY technological and decision areas, and processes that information to form other information sets. As a result we get a plan for traffic work for the MY investigated. This is (in a classic railway terminology) "a graphic plan of the station's work", a "car flow plan" and a "car departure plan" for a particular schedule and its part IV addition for particular days.

As a result of utilization of information from POI-III we will achieve the same effect with the slight difference that these plans usually contain existing deviations from a schedule and its part IV additions for the MY investigated.



Optimal Vehicle Control from the Point of View  
Of Energy Conservation

L. Skyva

INTRODUCTION

Energy resources have become one of the most serious problems of our time. This is the main reason why we have tried to find the optimal vehicle control from the point of view of energy consumption. The other reason is the introduction of computers for the automatic control of transportation to increase velocity, safety and precision of traffic. Minimal-energy control increases the effectiveness of computer utilization and helps to motivate its application.

Minimal-energy control may be applied to various vehicles. We are especially concerned with train control, as this is the most important kind of transportation in our country. There are various systems and principles of automatic train control under elaboration or being tested in the world at present. Programmed automatic control where the trajectory of motion is predetermined finds wide application. Automated control with a specialized on-board computer ensures better adherence to the time schedule, but is more complicated. It is useful in both cases to determine the algorithm of the optimal control and the optimal trajectory.

There are several methods for finding the minimal-energy control. The best known and most widely used for solving the problem are the Pontrjagin maximum principle and Bellman's method of dynamic programming. Pontrjagin's principle is quite convenient for solving strongly simplified minimal-energy train control (where the route is straight, with only one velocity limit, constant inclination of the track, constant traction force and constant brake, train resistance linearly dependent on velocity). The real running of a train, however, is much more complicated and it is necessary to use more ingenious methods - namely, numerical methods or Bellman's method. These need comparatively large computers which cannot be placed on board the vehicle. Solving the problem in a computing center takes a lot of time, but is acceptable and usable in program-guided automatic train control. Some algorithms must be found in the future for real-time minimal-energy control which would be fast and simple enough.

We start with a short look back at the development of our approach to the solution of the optimal control problem.

DEVELOPMENT OF A TRAIN SIMULATOR AND RELATED EXPERIMENTAL  
HARDWARE

In 1962 we built in the laboratory a full size physical model of an electrical locomotive, with the engine output scale of 1:500, with a real control stand and 3000 V in the collector. The control was semi-automatic, track signals were relayed from the track to the control stand. However, the obtained trajectory was in the range of  $\pm 10\%$  and the results could not be used for making control more accurate.

In 1965 we bought a special electromechanical analog computer made to our specifications from the Swiss firm Amsler. This device, the "Fahrdiagraph", not only plotted the trajectory of train running, but also performed computations of the energy needed for that running. We carried out a variety of runs, looking for the trajectory which had the least consumption of energy. In comparison to intuitive manual control, we achieved savings of energy of about 10%. This device was used to calculate the train running times for the Prague Underground and also for suburban units near Prague.

In the years 1963-67, we constructed an electronic simulator which represented a relatively very accurate model for the running of various trains. This model solves linear differential equations of the second order. Physical factors (speed, acceleration, traction force and brake, specific rolling resistances, voltages, etc.) are modelled by direct-current voltage. There have been 10 such models built in the CSSR and they serve for the energy-economy training for new drivers and also for instruction at our university. The optimal trajectory is computed in advance and the simulator operator uses manual controls to try and run as close as possible to it.

Since 1970 we have been concerned with the minimal-energy control of a train model using a control computer. For this we bought two control computers, the Robotron KRS 4200 (1974) and the Tesla RPP-16 (1975).

THE BASIC APPROACH USED FOR SOLVING THE MINIMAL-ENERGY CONTROL  
PROBLEM

As a rough approximation, we used a simplified train model and the Pontrjagin maximum principle. The train model was simplified by introducing approximations for the rolling resistance, dependence of traction and braking force upon speed and track profile. The rolling resistance was taken to be linearly dependent on the speed. The traction force and brake were both considered constant and independent of speed. While the behavior of the braking force could be kept about constant by a combination of electromagnetic and air brakes, the substitution of a constant for the real behavior of the traction force was only very approximate and did not correspond to reality. The



inclination of the track was considered constant on a given track segment for which the optimal motion trajectory was looked for.

In 1970-71 we used the maximum principle to qualitatively determine the optimal trajectory algorithm which consisted of four stages: run-off with maximal traction force, motion with constant speed, coasting, and braking by applying maximal permissible brake force. This same conclusion was arrived at and published at about the same time by several authors (K. Ichikawa, P. Horn, H. Strobel, U.P. Petrow), and corresponds to operational experience. From the point of view of energy, there are three different modes of running for the definite given distance between stations:

- Four-stage mode with all four above-mentioned stages.
- Three-stage mode lacking motion with constant speed.
- Four-stage mode with constant speed equal to the speed limit of the particular track segment.

The fourth mode amounts to running in the minimum time and is not energy-minimizing. It is possible to determine which mode of motion for the simplified above train model should be selected and also to compute the time instants or track points where there should be a switching from one stage of control to another.

The simplified model as described by all authors differs profoundly from reality, and therefore we started to make it more accurate. The following considerations use results from the analysis of runs with use of the Pontrjagin maximum principle for a simplified train model. Numerical methods were used for the calculations of switching points of the trajectory. This enabled us to make computations of the quadratic dependence of rolling resistance on speed, and for the arbitrary behavior of traction force and brake.

However, the numerical solution is rather time-consuming and this precludes using this method with a small control computer on board the prime mover. The numerical solution is suitable for computations of the optimal trajectory in advance in a computing center, and also for programmed control. The application of this method with exact switching brings energy savings of up to 10% compared with the intuitive control by a driver.

#### ANALYTICAL EXPRESSION FOR THE MINIMAL-ENERGY TRAJECTORY OF TRAIN RUNNING

In spite of these difficulties, we eventually reached a compromise with a solution close enough to reality and easily computable. It is possible to find analytically the instant when coasting begins and the instant when braking begins. Even though the analytic expressions are relatively complex, the computation

is much quicker and easier than numerical computations. The computation for the run-off stage can be done numerically making it possible to calculate the arbitrary dependence of the traction force on velocity. If we choose the run-off period,  $T_1$ , or the maximum velocity at the end of the run-off period,  $V_1$ , we can numerically compute the time behavior of velocity and the trajectory during the run-off stage. Furthermore we can analytically express the beginning of coasting,  $T_2$ , and the beginning of braking,  $T_3$ , in order to determine the energy consumption on the given track segment for the given mode of running. By varying the time,  $T_1$ , or the velocity,  $V_1$ , we can find out the optimal values which minimize energy consumption. In order to determine the optimal values of  $T_1$  and  $V_1$ , we used the method for investigating a monotone function for a minimum. The program for computing the switch points  $T_2$  and  $T_3$  can also be applied for the real-time control. For the minimal-energy train motion one should adhere to the switch points with a high degree of accuracy.

This modified program was used in the simulation model of the train, and with it we could test the influence of parameter variation upon the energy consumption of the train moving on a given segment of the track.

#### MATHEMATICAL FORMULATION AND SOLUTION OF THE PROBLEM

1. There is given a dynamic model which is the mathematical model of train running. The system is described by:

(a) a set of differential equations:

$$\dot{x}(t) = f[x(t), u(t), t] \quad ,$$

where  $x = [x_1, x_2]^T \in X$ ,  $u = [u, 0]^T \in U$  ,

where  $f_i[x(t), u(t), t]$ ;  $\partial f_i[x(t), u(t), t] / x_j$

for  $i = 1, \dots, n$  ,

$j = 1, \dots, n$  ,

are defined and continuous on  $X \times U$ .

(b) The initial state of the system

$$x(0) = x^0 .$$

(c) The object set

$$x(T_k) = x^k .$$

(d) The admissible control  $U$  .

(e) The criterion of optimal control

$$J(u) = \int_0^{T_k} \frac{1}{2}(u + |u|) \cdot x_2 \cdot dt .$$

(f) The Hamiltonian  $H(x, u, \lambda) \rightarrow \max \rightarrow u = \begin{matrix} -F_b \\ 0 \\ p \\ F_t \end{matrix} x_2 .$

The differential equation can have, for example, the form:

$$\dot{x}_1(t) = x_2(t) ,$$

$$\dot{x}_2(t) = u(t) - p(x_2) ,$$

where  $u(t) \in [-F; +F]$  is the control function ,

$p(x_2) = k \cdot x_2$  is the linearized rolling resistance ,

$x_1(t)$  is the trajectory,  $x_1 \in [0, x_k]$  ,

$x_2(t)$  is the velocity,  $x_2 \in [0, x_{2m}]$  ,

$F$  is the traction resp. braking force  $-F_b \leq F \leq F_t$  ,

and  $m$  is the mass of a train.

2. The objective is to find a control  $u(t) = u^*(t)$ ,  $t \in [0, T_k]$ , and a suitable  $u(t) \in U$  and  $x(t)$  satisfying 1(a) - 1(d) and  $J(u) \rightarrow \min$ .

The quantity  $T_k$  can be fixed or arbitrary. The problem thus defined can be solved only for the simplified model of a train.

3. Hence we transform the problem and the new objective is to find the switch points,  $T_i$ ,  $i = 1, 2, 3$ , so that the functional  $J(u)$  is minimal, i.e.

$$J(u) = J(T_1, T_2, T_3) = \int_0^{T_1} F_t x_2 dt + \int_{T_1}^{T_2} p(x_2) \cdot x_2 \cdot dt \rightarrow \min .$$

4. For  $\forall T_1$  fixed  $\exists (T_1, T_2, T_3)$  satisfying the conditions of 1. and for  $\forall T_1 \exists J(T_1)$ , where  $T_{1, \min} \leq T_1 \leq T_{1, \max}$ .  
 $Jf \exists T_{1, \text{opt}} \rightarrow \exists J(T_{1, \text{opt}}) \rightarrow \min$

5. The combined analytical and numerical solution of the optimal control of the non-linear mathematical model has two parts:

- (a) the solution of the linear mathematical model

$$\dot{x}_1(t) = x_2(t) ,$$

$$\dot{x}_2(t) = u(t) - p(x_2) ,$$

which is used as an entry for (b), and

- (b) the solution of the non-linear mathematical model

$$\dot{x}_1(t) = x_2(t) ,$$

$$x_2(t) = u(x_2, k) - p_0(x_2) - s_r - p_b(x_2) ,$$

where  $u(x_2, k)$  is specific traction force, which is a step function,

$$p_0(x_2) = k_0 + k_1 x_2 + k_2 x_2^2 \text{ is the specific rolling resistance,}$$

$s_r$  is the inclination of the track, and

$p_b(x_2)$  is the specific braking force, which is a step function.

First, we solve the linear model 5(a)

(i) In the run-off stage we solve

$$\dot{x}_1(t) = x_2(t) ; \quad \dot{x}_2(t) = u(t) - k \cdot x_2(t)$$

for the initial conditions  $x_1(0) = 0$ ,  $x_2(0) = 0$ ,  
and  $u(t) = F$  .

We obtain

$$x_1^{(1)}(t) = \frac{F}{k} \left[ t - \frac{1}{k} (1 - e^{-kt}) \right] ,$$

$$x_2^{(1)}(t) = \frac{F}{k} (1 - e^{-kt}) , \quad t \in [0, T_1] .$$

(ii) In the run with constant velocity we solve

$$\dot{x}_1(t) = x_2(t); \quad \dot{x}_2(t) = 0 ,$$

for the initial conditions:  $x_1(T_1) = 0$ ,  $x_2(T_1) = V_1$ ,  
and  $u(t) = k \cdot x_2(t)$  .

We obtain

$$x_1^{(2)}(T_2) = V_1(T_2 - T_1) ; \quad x_2^{(2)}(t) = V_1 = \text{const.}$$

(iii) In the coasting stage we solve

$$\dot{x}_1(t) = x_2(t); \quad \dot{x}_2(t) = -k \cdot x_2 ,$$

for the initial conditions:  $x_1(T_2) = 0$ ,  $x_2(T_2) = V_1$ ;  
 $u(t) = 0$ ,  $t \in [T_2, T_3]$  .

We obtain

$$x_1^{(3)}(t) = \frac{V_1}{k} \left( 1 - e^{-k(T_2-t)} \right); \quad x_2^{(3)}(t) = V_1 e^{-k(T_2-t)} .$$

(iv) In the braking stage we solve

$$\dot{x}_1(t) = x_2(t); \quad \dot{x}_2(t) = -F - k \cdot x_2(t) ,$$

for the initial conditions:  $x_1(T_3) = 0$ ,  $x_2(T_3) = V_3$ ;

$$u(t) = -F; \quad t \in [T_3, T_k] .$$

We obtain

$$x_1^{(4)}(t) = \frac{k \cdot V_3 + F}{k} \left( 1 - e^{-k(t-T_3)} \right) + \frac{F}{k} (T_3 - t) ,$$

$$x_2^{(4)}(t) = \frac{k \cdot V_3 + F}{k} e^{-k(t-T_3)} - \frac{F}{k} .$$

In the regular optimal mode we have to satisfy the condition that the sum of track segments  $x_1^i$ ,  $i = 1, 3, 4$  must be equal to the given  $x_k$

$$x_1^{(1)}(T_1) + x_1^{(3)}(T_3) + x_1^{(4)}(T_k) = x_k ,$$

$$\begin{aligned} & - \frac{F}{k^2} \left[ \frac{k}{F} V_1 + \ln \left( 1 - \frac{k}{F} V_1 \right) \right] + \frac{V_1}{k} \left[ 1 - e^{-k(T_2-T_3)} \right] \\ & + \frac{V_3}{k} - \frac{F}{k^2} \ln \left( 1 + \frac{k}{F} V_3 \right) = x_k , \end{aligned}$$

for  $T_2 = T_1 = - \frac{1}{k} \ln \left( 1 - \frac{k}{F} V_1 \right)$ , and  $V_3$

$$= \frac{F V_1}{(F - V_1 k) e^{-k T_k} - k V_1} .$$

We obtain

$$V_1 = \frac{F}{2k} \left\{ \left[ 2 - e^{-\frac{F}{k^2} x_k} \left( 1 + e^{-kT_k} \right) \right] - \sqrt{\left[ e^{-\frac{F}{k^2} x_k} \left( 1 + e^{-kT_k} \right) - 2 \right]^2 - 4 \left( 1 - e^{-\frac{F}{k^2} x_k} \right)} \right\},$$

and for the energy consumption

$$E = -G \frac{F}{k} \left[ \frac{F}{k} \ln \left( 1 - \frac{k}{F} V_1 \right) + V_1 \right].$$

In the singular optimal mode we have, similarly, to satisfy the condition:

$$x_1^{(1)}(T_1) + x_1^{(2)}(T_2) + x_1^{(3)}(T_3) + x_1^{(4)}(T_k) = x_k$$

and, assuming that  $v_3 = 0.5v_1$  ([1]),  $v_1$  equals

$$v_1 = \frac{x_k}{T_k} + \frac{F}{k^2 T_k} \ln \left[ \left( 1 + \frac{k}{2F} v_1 \right) \left( 1 - \frac{k}{F} v_1 \right) + \frac{v_1}{k T_k} \ln 2 \left( 1 + \frac{k}{2F} v_1 \right) / \left( 1 - \frac{k}{F} v_1 \right) \right],$$

and the energy consumption is

$$E = G \left[ -\frac{F^2}{k^2} \ln \left( 1 - \frac{k}{F} v_1 \right) - \frac{F}{k} v_1 + k v_1^2 T_k - v_1^2 \ln \frac{k v_1 + 2F}{F - k v_1} \right].$$

The critical time can be found by using the condition  $T_1 = T_2$ :

$$T_{\text{crit}} = \frac{1}{K} \ln \left( \frac{2F + kV_1}{F - kV_1} \right) .$$

$T_{\text{crit}}$  is a criterion for making a choice between a three-stage or a four-stage mode. If the time,  $T_k$ , prescribed by the timetable, is less than  $T_{\text{crit}}$ , ( $T_k < T_{\text{crit}}$ ), the three-stage mode is optimal, if  $T_k > T_{\text{crit}}$ , the four-stage mode is optimal. By computing  $V_1 = x_2(T_1)$ , we determine the velocity at the end of the run-off stage. This velocity  $x_2(T_1) \leq x_{2,\text{max}}$ .

Applying this procedure, we have found the switch points ( $T_1, T_2$ , and  $T_3$ ) for the linear model. These values can be used for the non-linear model as the first approximation. The non-linear model describes the train movement relatively well. Specific rolling resistance is assumed to have quadratic dependence on velocity, traction force and braking force are arbitrary. The computations are according to the following order of importance of conditions:

- (a) The switching time  $T_3$  is determined by the condition of stopping of the train exactly at the station.
- (b) The switching time  $T_2$  is determined by the condition of exact adherence to the running time,  $T_k$ .
- (c) The switching time  $T_1$  is determined by the condition of minimal energy consumption.

Once the time point  $T_1$  has passed, we cannot influence the optimality of the train movement any more, and by sticking exactly to the time points  $T_2$  and  $T_3$ , we reach the point of destination in exactly the prescribed time and with only the minimum necessary energy consumption. Switching times  $T_2$  and  $T_3$  were set up analytically. Even though the results are very complicated relations, they strictly follow the order of importance of conditions (a) and (b) and, in addition, the calculation time is several times shorter than when numerical methods are used.

The most difficult problem was to set up  $T_1$ . The analytical solution was impossible to use and it was therefore necessary to



apply numerical methods. The first approximation to the value of  $T_1$  was calculated in the linear part of the model. Then the switching times  $T_2$  and  $T_3$  and the energy consumption were calculated. Numerical methods were applied to find the optimal value of  $T_{1,opt}$  (see 4.).

The run for the run-off stage is now calculated further; the calculation determines the variation of final values for trajectory, velocity, and energy consumption with variation of  $T_1$ . The energy consumption is then compared to the original one (with original  $T_1$ ). If the energy consumption has decreased, we move the value of  $T_1$  in the same direction as long as the energy consumption goes down. At the moment when the energy consumption starts to increase, we have obtained the last interval containing  $T_{1,opt}$ . By this method the run-off stage is solved.

We solve the system:

$$\dot{x}_1(t) = x_2(t); \dot{x}_2(t) = u(x_2, k) - p_0(x_2) - s_r - p_b(x_2) .$$

Having solved the run-off stage in the linear model, i.e. having  $T_1$  and  $x_2(T_1)$ , we define the problem: The train moves with the velocity  $x_2(T_1)$  and should cover the distance  $x_1 = x_k - x_1(T_1)$  in the time  $T = T_k - T_1$ . It is to find  $T_2$  and  $T_3$  so that the required conditions be satisfied.

The solution of the system of differential equations for braking and coasting is done in a negative time sense. For braking, we obtain the optimal trajectory by solving the system of differential equations

$$\dot{x}_1(t) = x_2(t); \dot{x}_2(t) = F + k_0 + k_1 x_2(t) + k_2 x_2^2(t) ,$$

$$\text{for } x_2(0) = v_k, \quad x_1(0) = 0, \quad p_b(x_2) = F .$$

The velocity and the trajectory are

$$x_2^{(1)}(T_1) = \frac{D_1}{2k_2} \operatorname{tg} \frac{D_1}{2} (T_1 + C_1) - \frac{k_1}{2k_2} ,$$

$$x_1^{(1)}(T_1) = - \frac{1}{k_2} \left[ \ln \cos \frac{D_1}{2} (T_1 + C_1) - \frac{k_1}{2} T_1 - \ln \cos \frac{D_1 C_1}{2} \right] ,$$

where  $D_1 = \sqrt{4(F + k_0)k_2 - k_1^2}$  ;  $C_1 = \frac{2}{D_1} \operatorname{arctg} \frac{2k_2 V_k + k_1}{D_1}$  .

The optimal trajectory in coasting is obtained by solving the system:

$$\dot{x}_1(t) = x_2(t); \quad \dot{x}_2(t) = k_0 + k_1 x_2(t) + k_2 x_2^2(t) ,$$

for  $x_2(T_1) = x_2^{(1)}(T_1)$ ;  $x_1(T_1) = x_1^{(1)}(T_1)$ ,  $u = p_b(x_2) = 0$  .

Calculations are done for

(A)  $4k_0k_2 - k_1^2 > 0$  ,  $D_2 = \sqrt{4k_0k_2 - k_1^2}$  ,

(B)  $4k_0k_2 - k_1^2 < 0$  ,  $D_2 = \sqrt{k_1^2 - 4k_0k_2}$  ,

(C)  $4k_0k_2 - k_1^2 = 0$  .

$$x_2^{(2)}(t) = \frac{1}{k_2} \left\{ \frac{D_2}{2} \operatorname{tg} \frac{D_2}{2} \left[ t - T_1 + \frac{2}{D_2} \operatorname{arctg} \frac{D_1}{D_2} \operatorname{tg} \frac{D_1}{D_2} (T_1 + C_1) \right] - \frac{k_1}{2} \right\} ,$$

$$x_1^{(2)}(t) = \frac{1}{k_2} \left\{ - \ln \cos \frac{D_2}{2} \left[ t - T_1 + \frac{2}{D_2} \operatorname{arctg} \frac{D_1}{D_2} \operatorname{tg} \frac{D_1}{D_2} (T_1 + C_1) \right] - \frac{k_1}{2} t + \ln \cos \frac{D_1}{2} C_1 + \ln 1 / \sqrt{1 + \left[ \frac{D_1}{D_2} \operatorname{tg} \frac{D_1}{D_2} (T_1 + C_1) \right]^2} \cos \frac{D_1}{2} (T_1 + C_1) \right\}$$

Switching time  $T_2$  is determined by using the condition

$$x_2^{(2)}(T_2) = v_1 \quad ;$$

$$T_2 = T_1 + \frac{2}{D_2} \left[ \operatorname{arctg} \frac{2k_2 v_1 + k_1}{D_2} - \operatorname{arctg} \frac{D_1}{D_2} \operatorname{tg} \frac{D_1}{2} (T_1 + C_1) \right] \quad ,$$

then  $x_1^{(2)}(T_2) = x_k - (T_k - T_2)v_1 \quad .$

Switching time  $T_1$  is determined by substituting  $T_k$  for  $T_2$  into

$x_1^{(2)}(T_2)$ , i.e.  $x_1^{(2)}(T_k) = x_k$ , which can be solved. Similarly, the calculations under (B) and (C) can be performed. When covering the distance in the time  $T_k$  min, coasting is abandoned. The switching time  $T_1$  is determined by the condition  $x_2^{(1)}(T_1) = v_1$  :

$$T_1 = -C_1 + \frac{2}{D_1} \operatorname{arctg} \frac{2k_2 v_1 + k_1}{D_1} \quad .$$

$T_{k,\min}$  is determined from:

$$x_1^{(1)}(T_1) = x_k - (T_{k,\min} - T_1)v_1 \quad ,$$

$$T_{k,\min} = \frac{1}{k_2 v_1} \left\{ x_k k_2 + \frac{2k_2 v_1 + k_1}{2} \left[ -C_1 + \frac{2}{D_1} \operatorname{arctg} \frac{2k_2 v_1 + k_1}{D_1} \right] - \frac{1}{2} \ln \left[ 1 + \frac{2k_2 v_1 + k_1}{2} \right]^2 - \ln \cos \frac{D_1}{2} C_1 \right\} \quad .$$

The relations derived for the non-linear model can serve to set up the optimal train movement on a given track segment, thus verifying theoretical works on optimal train control, and also to simulate the real-time control of a prime mover. The combination of analytical solutions and numerical methods ensures relatively good agreement with reality with minimal simplifications.

The computer program permits one to obtain:

- The optimal  $x_2(x_1)$ ,  $x_2(t)$ , and  $x_1(t)$  both calculated and observed, for given train characteristics.
- The energy consumption dependent on the running time,  $T_k$ , for a given track length,  $x_k$ .
- The switching times  $T_1, T_2, T_3, T_{k,\min}$ , and  $T_{\text{crit}}$ .

#### BIBLIOGRAPHY

- [1] Ichikawa, K., *Application of Optimization Theory for Bounded State Variable Problems to the Operation of a Train*, Tokyo University, Tokyo, 1968.
- [2] Horn, P., *Experimentelle Untersuchungen eines Verfahrens zur energieoptimalen Zugsteuerung*, Transpress, Berlin, 1973.
- [3] Strobel, H., et al., Über die Ermittlung und praxisnahe Erprobung von Regelungsstrategien für eine rechnergestützte Optimierung, *Schienen der Welt*, 5 (1974).

## APPENDIX



List of Participants and Authors

O.I. Aven  
Institute of Control Sciences  
Moscow  
USSR

S.N. Baibakov  
VNIPI Transprogress  
Kutuzovsky Prospect 31  
Moscow  
USSR

G.M. Barber  
Department of Geography  
Northwestern University  
Chicago  
USA

Z. Bubnicki  
Technical University of Wroclaw  
PL - Wroclaw  
27 ul. Wyspianski  
POLAND

L. Burrow  
Warwick University  
Coventry  
W. Midlands CV4 7AL  
UK

R. Camus  
University of Trieste  
Istituto di Elettrotecnica  
Via Valerio 10  
Trieste  
ITALY

M. Cremer  
Technische Universität  
Postfach 2024 20  
8 München  
FRG

Antonio D'Amore  
University of Trieste  
Istituto di Elettrotecnica  
Via Valerio 10  
Trieste  
ITALY

Ghulam Dastgir  
Brighton Polytechnic  
Department of Computing  
and Cybernetics  
Brighton BN2 4GJ  
Sussex  
UK

M. Dechicha  
Sonitex Dep. Etudes Dir.  
Planification  
5, Rue Abane Ramdane Alger  
ALGERIA

H.V. Ellingsen  
Norges Skipstforskningsinst.  
Oslo 6, Grenseveien 99  
NORWAY

M.M. Etschmaier  
Department of Industrial  
Engineering, Systems Management  
Engineering and Operations Research  
University of Pittsburgh  
Pittsburgh, Pa. 15261  
USA

A. Fauland  
Mag. d. Stadt Wien  
MA 8 Ref. 5  
Rathaus  
A-1010 Wien  
AUSTRIA

J.J. Fearnside  
US Department of Transportation  
UTG Engineering Department  
Washington, DC 20590  
USA

J.C. Ferrer  
Renfe Direccion de Innovacion Madrid  
Torre de Madrid, Princesa  
SPAIN 1

J. Fonlupt  
IRT  
2, Ave. du Gl. Malleret-Joinville  
94 Arceuil  
FRANCE



M. Genete  
SNCF Direction des Etudes Générales  
de la Recherche  
162 Rue du Fg. St. Martin  
75475 Paris, Cedex 10  
FRANCE

R. Genser  
Österr. Bundesbahn - GD/IK  
Trattnerhof 2  
A-1010 Wien  
AUSTRIA

J. Goy  
IRT  
2, Ave. du Gl. Malleret-Joinville  
94 Arceuil  
FRANCE

G. Gilnreiner  
Mag. d. Stadt Wien  
MA 8 Ref. 5  
Rathaus  
A-1010 Wien  
AUSTRIA

T. Hasegawa  
Dept. of Applied Mathematics  
and Physics  
Faculty of Engineering  
Kyoto University  
Kyoto  
JAPAN

K. Heinrich  
IABG  
8012 Ottobrunn bei München  
Einsteinstrasse  
FRG

M. Helfert  
Inst. f. Regelungstechn.  
u. Prozessautomat.  
Universität Stuttgart  
Seidenstr. 36  
D-7 Stuttgart  
FRG

Edith Heurgon  
RATP  
53 ter quai des Edo Augustins  
75016 Paris  
FRANCE

A. Hitchcock  
Dept. of the Environment  
Room S10/80  
2 Marsham Street  
London SW1  
UK

M.J. Hoek  
Venerigde Befrijven Bredero N.V.  
Nieuwe Gracht 6  
Utrecht  
THE NETHERLANDS

Y. Istefanopolos  
Bogazici University  
Dept. of Electrical Engineering  
P.K. 2 Bebek  
Istanbul  
TURKEY

E. Jacquet-Lagreze  
Marketing et Modèles de Decision  
16-18 Rue Barbes  
92128 Montrouge  
FRANCE

O.H. Jensen  
IMSOR  
Technical University of Denmark  
2800 Lyngby  
DENMARK

V.J. Klima  
ÖMV  
Otto Wagner Pl. 5  
A-1090 Wien  
AUSTRIA

S.E. Lovetsky  
Institute of Control Sciences  
Moscow  
USSR

R.D. MacKinnon  
University of Toronto  
Dept. of Geography  
Sidney Smith Hall  
100 St. George Street  
Toronto, Ontario  
CANADA

O.B.G. Madsen  
IMSOR  
Technical University of Denmark  
2800 Lyngby  
DENMARK

C. Mandl  
Institute for Advanced Studies  
Stumperg. 56  
A-1060 Wien  
AUSTRIA

G.E. Moiseenko  
Institute of Control Sciences  
Moscow  
USSR

P. Mitter  
Institute for Advanced Studies  
Stumperg. 56  
A-1060 Wien  
AUSTRIA

Y. Ogawa  
Kyoto University  
Kyoto  
JAPAN

C. van Ouwerkerk  
Shell Laboratories  
Dept. MSE  
Badhuisweg 3  
Amsterdam  
THE NETHERLANDS

K. Pierick  
Institut für Verkehr T.U. Braunschweig  
Pockelstr. 4  
3300 Braunschweig  
FRG

M. Pollack  
Lockheed  
California Company  
Airline Research & Planning Dept. 95-8  
Burbank, California 91520  
USA

G. Rogenstad  
Norges Skipstforskningsinst.  
Oslo 6, Grensevein 99  
NORWAY

B. Roy  
University of Paris and SEMA  
16-20, rue Barbes  
92128 Montrouge  
FRANCE

J.L. Schlaefli  
Applied Transportation Systems, Inc.  
505 Hamilton Avenue  
Palo Alto, California 94301  
USA

H.W. Schleife  
Ministerium fur Verkehrswesen  
108 Berlin  
Voss Strasse 33  
GDR

P. Scioldo-Zurcher  
Institut Universitaire Technique  
63170 Aubiere  
FRANCE

P. Sipala  
University of Trieste  
Istituto di Elettrotecnica  
Via Valerio 10  
Trieste  
ITALY

A.M. Skarke  
International Institute for  
Applied Systems Analysis  
A-2361 Laxenburg  
AUSTRIA

L. Skyva  
VSD  
Moyzesova 25  
Zilina  
CSSR

J. Sokolowski  
Institute for Organization  
Management & Control Sciences  
KRN 55 Warszawa  
POLAND

G. Stemmer  
IPA  
Holzgartenstr. 17  
7000 Stuttgart  
FRG

H. Strobel  
International Institute for  
Applied Systems Analysis  
A-2361 Laxenburg, AUSTRIA  
and  
Hochschule für Verkehrswesen  
"Friedrich List"  
Friedrich-List-Platz 1  
801 Dresden  
GDR

M. Watanabe  
Kyoto University  
Kyoto  
JAPAN

K. Wiegand  
Institut für Verkehr T.U. Braunschweig  
Pockelstr. 4  
3300 Braunschweig  
FRG

M. Zemouli  
Sanitex Dept. Etudes Dir.  
Planification  
5, Rue Abane Ramdane Alger  
ALGERIA

