TR/12/83 November 1983

An Extension of Set Partitioning With Application to Scheduling Problems

by

K. Darby-Dowman and G. Mitra

Contents

O. Abstract

The well known problems of set covering, set partitioning and set packing are defined and their interrelationship is considered. A natural generalisation called the extended set partitioning model is presented and the three standard models are shown to be special cases of this generalisation. In addition, the extended model includes another type of set problem which can be of greater use in certain applications. The model forms the basis of a computer assisted bus crew scheduling system developed by the authors. The system is in regular use by Dublin City Services in the Republic of Ireland. Finally, the equivalence between a special case of the set partitioning problem and the shortest route problem is considered and it is shown that this equivalence also applies to the extended model.

1. Introduction

The well-known problems of set covering, set partitioning and set packing have attracted wide attention for many years. Applications include airline crew scheduling [1,11,15], bus crew scheduling [11,12] plant location [5,18], circuit switching [13], information retrieval [7], assembly line balancing [16], political districting [10] and truck delivery [3]. The use of these models in practice has not been an unqualified success. In many applications the models become too large to solve exactly [11]. In such cases heuristics are often used either to obtain an approximate solution [1] or to reduce the model to a more manageable size. A difficulty with the latter approach is that the reduction may destroy the appropriateness of the model to the application. In this paper an extended model is proposed in which the three standard problems are special cases and which can be of greater applicability. The model, called the extended set partitioning model, forms the basis of a computer assisted bus crew scheduling system that the authors have developed and implemented at Dublin City Services in the Republic of Ireland.

In the next section the set covering, set partitioning and set packing problems are defined and their interrelationships examined. The extended set partitioning model and its properties are presented in section 3. In section 4 an application of the models to the problem of bus crew scheduling is discussed. The extended set partitioning model is shown to be more useful than the standard set partitioning model. In section 5 the equivalence between a special case of the set partitioning problem and a formulation as a shortest route problem is stated. It is shown that the corresponding special cases of the other models, in particular the extended set partitioning problem, can also be formulated as shortest route problems. In the final section it is noted that the extended set partitioning model may be considered as a goal programming [6] formulation of the set partitioning problem.

2. Background

In this section the set covering, partitioning and packing problems are defined in 0-1 integer programming terms. The three problems refer to different criteria for selecting a subclass from a class of sets.

Let N denote the set of m integers 1,2,...m and let Q denote a class of n subsets of N . Thus,

$$N = \{1, 2, \ldots m\},$$

and $Q = \{Q_1, Q_2, \ldots Q_n\}$ where $Q_j \subseteq N, j = 1,2,\ldots n.$

Let
$$\left.\begin{aligned} a_{ij} &= 1 \text{ if } i \in Q_j \\ &= 1 \text{ if } i \notin Q_j \end{aligned}\right\} \quad \begin{aligned} i &= 1,2,\ldots m, \\ j &= 1,2,\ldots n. \end{aligned}$$

The set covering problem may be defined as

$$\text{Minimise} \sum_{j=1}^{n} c_j x_j$$

$$\text{subject to} \sum_{j=1}^{n} a_{ij} x_j \geq 1 \quad i = 1,2,\ldots m,$$

$$x_j \in \{0,1\} \quad j = 1,2,\ldots n,$$

where c_j represents the cost of selecting the subset Q_j (j = 1,2,...n). The problem represents the minimum cost selection such that each member of the set N is included at least once.

The set partitioning problem may be defined as

$$\text{Minimise} \sum_{j=1}^{n} c_j x_j$$

$$\text{subject to} \sum_{j=1}^{n} a_{ij} x_j = 1 \quad i = 1,2,\ldots m,$$

$$x_j \in \{0,1\} \quad j = 1,2,\ldots n,$$

and represents the minimum cost selection such that each member of N is included exactly once.

The set packing problem may be defined as

$$\text{Maximise} \sum_{j=1}^{n} p_j x_j$$

$$\text{subject to} \sum_{j=1}^{n} a_{ij} x_j \leq 1 \quad i = 1,2,\ldots m,$$

$$x_j \in \{0,1\} \quad j = 1,2,\ldots n,$$

where p_j represents the profit associated with selecting the subset Q_j , (j = 1,2,...n). This problem represents the maximum profit selection such that no member of N is included more than once.

The terms cover, partition and packing refer to any integer feasible solution to the corresponding problems. By inspection of the three sets of constraints it can be seen that a partition is also a cover and a packing although the converse is not necessarily true. Indeed, as pointed out by several authors including Balas and Padberg [2] in their comprehensive survey of set partitioning, a set partitioning problem may be transformed to a set covering problem as follows.

The set partitioning problem may be rewritten as

$$\text{Minimise} \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} \theta y_i$$

$$\text{subject to} \sum_{j=1}^{n} a_{ij} x_j - y_i = 1 \qquad i = 1,2,\dots m,$$

$$x_j \in \{0,1\}, \qquad j = 1,2,\dots n,$$
$$y_j \geq 0 \qquad i = 1,2,\dots m,$$

where θ is a sufficiently large positive number.

By substituting $y_i = \sum_{j=1}^{n} a_{ij} x_j - 1$ in the objective function the following formulation is obtained.

$$\text{Minimise } (-m\theta + \sum_{j=1}^{n} c'_j x_j)$$

$$\text{subject to} \sum_{j=1}^{n} a_{ij} x_j \geq 1 \qquad i = 1,2,\dots m,$$

$$x_j \in \{0,1\}$$

$$\text{where } c'_j = c_j + \theta \sum_{i=1}^{m} a_{ij},$$

which is a set covering problem (since the constant term —mθ may be dropped) with the same set of optimal solutions as the set partitioning problem. By a similar argument the set partitioning problem may also be formulated as a set packing problem. These equivalences, however, only apply when the set partitioning problem has a feasible solution. Unfortunately, when set partitioning is applied in practice it is often the case that no feasible solution exists for the formulated model. There are two principal reasons for this. One is that the formulation may represent an 'ideal' which is in fact unattainable. The other reason is that, in order to arrive at a model which is manageable in terms of both size and computational workload, the number of variables (subsets Q_j available for selection) may have to be reduced by some heuristic which excludes from the model those subsets which seem highly unlikely to be included in the optimal solution. The full (impractical to solve) model may have a feasible solution whereas the reduced (practical to solve) model may not.

Our experience with set partitioning problems with no feasible solutions led us to develop a more general model which yields a solution as near as possible to the desired set partition. This model called the extended set partitioning problem is explained in the next section.

3. The Extended Set Partitioning Problem

A set partitioning problem with no feasible solution is one in which an ‘exact cover’ does not exist. The set packing problem is a relaxation in which ‘undercover’ is permitted whereas the set covering problem is a relaxation in which 'overcover' is permitted. Whenever an exact cover does not exist one might be tempted to solve either the set packing problem or the set covering problem whichever is more appropriate to the given application. For example, in crew scheduling applications, it is desired to cover each trip of a given timetable with exactly one crew. This approach can lead to a formulation which is a set partitioning problem [11,12] . However, such a formulation often results in a problem with no feasible solution caused by the many regulations and desirable features governing the validity of crew duties. Consequently a formulation based on set covering has been suggested [12,15] in which the interpretation is that each trip must be covered at least once. The resultant solution can then be amended manually to remove the over cover and produce a crew schedule that can be implemented. The motivation behind this approach is to find the most useful solution that exists given that an exact cover may not exist. This purpose is better served by considering a more general model. The extended set partitioning formulation, defined below, has set packing, set partitioning and set covering as special cases and also encompasses a fourth model to be considered in which both undercover and overcover are permitted. The model is defined as follows.

$$\text{Minimise} \quad \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} w_i^u u_i + \sum_{i=1}^{m} w_i^o o_i$$

$$\text{subject to} \quad \sum_{j=1}^{n} a_{ij} x_j + u_i - o_i = 1 \qquad i = 1,2,\dots m,$$

$$x_j \in \{0,1\} \qquad j=1,2,\dots\dots n,$$

$$\left.\begin{array}{l} u_i \in \{0,1\} \\ o_i \geq 0 \text{ and integer} \end{array}\right\} \qquad i=1,2,\dots\dots m,$$

where w_i^u $(\geq o)$ is the penalty associated with not covering the ith member,(undercover), and w_i^o $(\geq o)$ is the penalty associated with each overcover of the ith member.

The trivial case of $w_i^u = w_i^o = 0$ for any $1 \leq i \leq m$ is not considered since it is equivalent to the freeing of constraints. Thus it is assumed that at least one of the pair (w_i^u, w_i^o) is strictly positive for each i = 1,2,...m.

The interpretations associated with u_i and o_i are as follows.

u_i = 1 if the ith member is not covered,
= 0 otherwise;
o_i = the number of times that the ith member is overcovered.

3.1 Interpretations of the Model

Set Covering If w_i^u ,(i = 1,2,...m), is set to a sufficiently large positive number (greater than $\sum_{j=1}^{n} c_j$, say) and w_i^o, (i = 1,2, ...m) , is set to zero, then the extended model and the set covering model have the same optimal solution(s).

Set Packing If w_i^o, i = 1,2,...m, is set to a sufficiently large positive number and w_i^u, i = 1,2,...m, is set to zero then the extended model (with $c_j = -p_j$, j = 1,2,...n) and the set packing model have the same optimal solution(s).

Set Partitioning If both w_i^o and w_i^u, i = 1,2,...m, are set to a sufficiently large positive number then the extended model and the set partitioning model have the same optimal solution(s), assuming that the latter is feasible. The extended model is guaranteed to be feasible and, in cases when the set partitioning model has no feasible solution, will yeild an optimal solution involving undercover and/or overcover. The solution is one in which the sum of undercover and overcover is minimised and, in this sense, is the nearest possible to a partition.

General Interpretation The nature of the solution to the extended model depends on the size of the penalties applied to undercover and overcover. The model was originally developed to be used in situations in which a partition is desired. However, deviations from an exact cover (partition) are permitted, but at a cost defined by the individual penalties applied to each member. Even in cases where a partition exists the solution to the extended model may not be a partition. The least cost partition may be more expensive than a solution involving undercover and/or overcover. Thus the extended model provides a flexibility to enable a wider class of solutions to be considered which, in certain cases, leads to a better solution.

4. Applications To Crew Scheduling

Crew scheduling by computer has received considerable attention during the past 20 years or so. Annual symposia organised by the Airline Group of the International Federation of Operations Research Societies (AGIFORS) have been held since 1961. Airline crew scheduling and allied problems play a prominent role in these symposia. Attention to crew scheduling in urban mass transit systems has gained momentum with international workshops on vehicle and crew scheduling held in Chicago (1975), Leeds (1980) [19] and Montreal (1983) [14] .

In its simplest form the problem may be stated as follows. Given a timetable with m trips, find the least cost set of crew duties such that each trip is covered by one crew. One approach is to generate the set of all possible duties that comply with the rules governing duty validity and then make a selection from this set of n generated duties. Such a formulation leads to the set partitioning problem defined in section 2 in which c_j denotes the cost of selecting the jth generated duty and $a_{ij}=1$ if the jth generated duty covers the ith trip. In bus crew scheduling a trip is defined as a portion of a bus journey between two consecutive relief points at which crews can join or leave the bus. The rules governing duty validity are many and vary from operator to operator but generally involve issues such as the maximum number of buses that a crew can work in a single duty (usually between 1 and 3), the maximum work time, the lengths of breaks and time bands within which meal breaks must be taken. In addition many unwritten rules apply. An experienced scheduler can rapidly judge whether a given crew schedule is operationally acceptable. He may decide, for example, that a balanced schedule is needed in which the minimum as well as the maximum work time for a duty is specified. The large number of rules is a mixed blessing for the set partitioning formulation. On one hand the model is restricted to a more manageable size since the number of generated duties is reduced. On the other hand no feasible solution may exist. A relaxation to the set covering model overcomes the difficulty and provides a solution in which the overcover can be removed in a way which the scheduler considers most appropriate. Unfortunately such an approach often results in a schedule that is more expensive and requires more crews than is desired. For example, a minor alteration in the timetable or a small relaxation in one of the rules may lead to a schedule costing significantly less. If a schedule is considered to be unsatisfactory in this respect it is far from a simple task to amend the set covering solution to produce an acceptable schedule.

Application of the extended model to crew scheduling can result in a solution in which some trips are covered more than once (as in the set covering approach) and some trips left uncovered. The undesirability of both these occurrences is controlled by the undercover and overcover penalties applied to each trip. Consequently the model does not insist that each trip is covered if the cost of doing so is excessively large. In such cases, however, the schedule is incomplete and hence operationally unacceptable. Fortunately, with practical crew scheduling problems, the incompleteness is confined to only a very few trips. The solution from the extended model can, with usually little trouble, be amended to form a crew schedule that can be imp lamented.

The extended model forms the basis of a computer assisted crew scheduling system developed by the authors. The system is in regular operational use by Dublin City Services in the Republic of Ireland. Full details of the development and implementation of both the model and the system are contained in [12]. The experience of the schedulers at Dublin City Services is that the system usually produces complete schedules. In those cases in which an incomplete schedule is produced it is a relatively simple task to make the necessary amendments to render it acceptable. Their opinion is that the system 'breaks the back' of the problem.

5. Special Cases Reformulated As Shortest Route Problems

In this section the special case of applications in which each column of the constraint matrix has only one segment of ones is considered. A column with a segment of ones is defined as a set of column elements a_{ij},

$a_{ij} = 0 \qquad i=1,2,\ldots k-1$ for $k > 1$,

$a_{ij} = 1 \qquad i=k,\ldots k + p$,

$a_{ij} = 0 \qquad i = k + p + 1 ,\ldots m$ for $k + p < m$.

In crew scheduling terms this represents a problem in which each crew must do a single stretch of working on a single bus. Such duties comprise a consecutive set of trips and the problem is known as the one-part duty crew scheduling problem.

It has been pointed out by Shepardson and Marsten [17] and others, that, when formulated as a set partiioning problem, the one-part duty problem is equivalent to a shortest route problem. The network contains (m+1) nodes and n edges, one for each column (generated duty). If the jth generated duty covers trips k,...k + p then the associated edge runs from node k to node k+ p +1 with length c_j . The problem of covering each trip exactly once at minimum total cost is thus the problem of finding the shortest route between node 1 and node m + 1 . It may be observed that the network contains no cycles since j > i for every edge (i,j).

It is less widely known that the set covering formulation of the one-part duty problem is also equivalent to a shortest route problem. In this case the network is augmented by m edges of the form (i + 1, i), i = 1,2,...m, of zero length.

The extended model, when applied to the one-part duty problem can also be formulated as a shortest route problem. The network of the set partitioning formulation is augmented by the addition of 2m edges; m edges of the form (i + 1, i) of length w_i^o, i=1,2,...m and m edges of the form (i, i + 1) of length w_i^u,i=1,2,...m are added.

The set packing problem can also be equivalenced to the shortest route problem by letting $c_j = -p_j$, j =,...n, and adding edges to the set partitioning network of the form (i, i + 1), i=1,2,...m, of zero length.

This shortest route formulation of set packing applications is likely to result in negative edge lengths but, as in the formulation of the set partitioning problem, the network contains no negative cycles and therefore an algorithm such as that due to Ford [9] may be applied. In the shortest route formulation of the other problems it is likely, in practice, that all edge lengths will be non-negative and the more efficient algorithm due to Dijsktra [8] can be applied.

6. Conclusions

A new model has been proposed which is different from the standard problems of set covering, set partitioning and set packing but which includes them as special cases. The model includes a fourth type of set problem which has useful applications. It is especially useful in situations in which a partition, although desirable, either does not exist or is not necessarily the best solution. The model, when used in this mode, may be considered as a (discrete) goal programming [6] formulation of the set partitioning problem.

REFERENCES

1. BAKER, E.K., BODIN, L.D., FINNEGAN, W.F. and PONDER, R.J., Efficient Heuristic Solutions to an Airline Crew Scheduling Problem, Trans. AIIE., 11 (1979), 79-85.
2. BALAS, E. and PADBERG, M.W., Set Partitioning - A Survey, SIAM Review, 18, 4, (1976), 710-760.
3. BALINSKI, M.L. and QUANDT, M.H., On an Integer Program for a Delivery Problem, Opns. Res., 12, (1964), 300-304
4. BELLMAN, R.E., On a Routing Problem, Quart .Appl .Math., 16, (1958), 87-90.
5. BILDE, O. and KRARUP, J., Sharp Lower Bounds and Efficient Algorithms for the Simple Plant Location Problem, Annals of Discrete Mathematics, 1. (1977), 79-97.
6. CHARNES, A., and COOPER, W.W., Management Models and Industrial Applications of Linear Programming, Wiley (New York), (1961).
7. DAY, R.H., On Optimal Extracting from a Multiple File Data Storage System: an Application of Integer Programming, Opns.Res., 13, (1965), 482-494.
8. DIJKSTRA, E.W., A Note on Two Problems in Connexion with Graphs, Num. Math., 1, (1959), 269-271.
9. FORD, L.R. Jr., Network Flow Theory, The Rand Corporation, P923, August 1956.
10. GARFINKEL, R.S., and NEMHAUSER, G.L., Optimal Political Districting by Implicit Enumeration Techniques, Man. Sci., 16, (1970), B495-B508.
11. MARSTEN, R.E. and SHEPARDSON, F., Exact Solution of Crew Scheduling Problems Using the Set Partitioning Model: Recent Successful Applications, Networks, 11, (1981), 165-177.
12. MITRA, G., and DARBY-DOWMAN, K., CRU-SCHED - A Computer Based Bus Crew Scheduling System using Integer Programming, presented at the Third International Workshop on Transit Vehicle and Crew Scheduling, in Montreal (June 1983), and published in [14] .
13. PYNE, I.B., and McCLUSKEY, E.J. Jr., An Essay on Prime Implicant Tables, SIAM, J., 9, (1961), 604-631.
14. ROUSSEAU, J.M., (ed), Proceedings of the Third International Workshop on Transit Vehicle and Crew Scheduling, held in Montreal, June 27-30, 1983, to be published by North-Holland (1984).
15. RUBIN, J., A Technique for the Solution of Massive Set Covering Problems with Applications to Airline Crew Scheduling, Trans.Sci., 7, (1973), 34-48.

16. SALVESON. M.E., The Assembly Line Balancing Problem, Trans. ASME, 77, (1955), 939-947.

17. SHEPARDSON, F. and MARSTEN, R.E., A Lagrangian Relaxation Algorithm for the Two Duty Period Scheduling Problem, Man. Sci., 26, 3, (1980), 274-281.

18. TORREGAS, C., SWAIN, R., REVELLE, C., and BERGMAN, L., The Location of Emergency Service Facilities, Opns. Res., 19, (1971), 1363-1373.

19. WREN, A. (ed), Computer Scheduling of Public Transport, (papers based on presentations at the International Workshop held at the University of Leeds, 16-18 July 1980), North-Holland, (1981).