Abstract

Network design applications are prevalent in transportation and logistics. We consider the multicommodity capacitated fixed-charge network design problem (MCND), a generic model that captures three important features of network design applications: the interplay between investment and operational costs, the multicommodity aspect, and the presence of capacity constraints. We focus on mathematical programming approaches for the MCND and present three classes of methods that have been used to solve large-scale instances of the MCND: a cutting-plane method, a Benders decomposition algorithm, and Lagrangian relaxation approaches.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the Organizing Committee.

Keywords: multicommodity capacitated fixed-charge network design; mathematical programming; cutting-plane algorithm; Benders decomposition; Lagrangian relaxation.

1. Introduction

Network design applications are prevalent in transportation and logistics (Magnanti and Wong, 1984; Crainic, 2000). In such applications, decision-makers have to select the “best” investments among a large set of discrete alternatives (typically, roads, transportation services, or consolidation points) by taking into account tradeoffs between investment and operational costs. Typically, given a network with multiple commodities (types of goods or vehicles) flowing from supply to demand points, one wishes to minimize a “complex” (non-convex) objective function, while satisfying supplies and demands for all commodities, as well as a large number of additional constraints that deal with capacity or congestion issues, budget limitations, topological restrictions, robustness (or resilience), etc.

Network design models can be classified according to a number of characteristics. Most often, the models assume a central authority (a firm or a public administration), but the development of models and methods to account for collaboration/competition issues is a challenging area of research, in particular in the field of logistics and supply chain management (Lehoux, D’Amours and Langevin, 2009). The network is often represented as a static entity, but an increasing number of models explicitly take into account the time dimension by building a dynamic, space-time, expansion of the network; this is the case in particular for service network design applications for which...
transportation services (railway, less-than-truckload, or intermodal itineraries) have to be scheduled. Network design models often assume data (in particular, supplies and demands) to be known; such a simplistic hypothesis might be realistic enough to capture the essence of applications involving “heavy” investment costs over a strategic/tactical planning horizon, but might be insufficient to design robust networks that will easily adapt themselves to significant variations in the data. This is why an increasing number of models explicitly integrate uncertainty; as an example, recent advances in the area of stochastic programming have been adapted to network design models (Crainic et al., 2009).

Network design models span the entire spectrum of planning applications: strategic, tactical, and operational. At the strategic level, investment decisions have an impact over several years and typically consist in building infrastructures, such as roads and plants. At this level, models are typically static and deterministic, with capacity or congestion issues often overlooked. At the tactical level, the design of transportation services over a dynamic network is considered over a mid-term planning horizon (typically, a few months); for such applications, capacity and congestion issues are central, and uncertainty is increasingly taken into account. At the operational level, the design of the network is adaptive, in the sense that the locations of the facilities (consolidation points corresponding to cross-docking terminals and parking spaces) change from one day to the next to account for variations in the demands (Gendron and Semet, 2009). Such models explicitly take into account capacity and congestion issues.

In this paper, we consider the multicommodity capacitated fixed-charge network design problem (MCND), a generic model that captures three of the most important features of network design problems: the interplay between investment and operational costs, the multicommodity aspect, and the capacity issue. A large number of methods have been developed to solve large-scale instances of the MCND, in particular heuristic approaches (see Hewitt, Nemhauser and Savelsbergh, 2010, and the references therein, for an account of heuristic algorithms inspired by metaheuristics principles). Here, we review algorithms based on mathematical programming. After a description of the problem and its formulation as a mixed-integer program (MIP), we will see how this model can be improved. We will follow with the presentation of three classes of methods that have been used to solve the problem: a cutting-plane approach, a Benders decomposition algorithm, and Lagrangian relaxation methods. We will conclude this work by suggesting promising research avenues for the MCND and more generally for network design problems.

2. Multicommodity capacitated fixed-charge network design

Given a directed graph $G = (N, A)$, where $N$ is the set of nodes and $A$ is the set of arcs, and a set of commodities $K$ to be routed according to a known positive demand $d^k$ flowing from an origin $O(k)$ to a destination $D(k)$ for each commodity $k$, the problem is to satisfy the demand at minimum cost. The objective function consists of the sum of transportation costs and fixed design costs, the latter being charged whenever an arc is used. The nonnegative transportation cost on arc $(i,j)$ is denoted $c_{ij}$, while the nonnegative fixed design cost for arc $(i,j)$ is denoted $f_{ij}$. In addition, there is a positive capacity $u_{ij}$ on the flow of all commodities circulating on arc $(i,j)$.

The MCND can be modeled as a MIP by using nonnegative continuous flow variables $x_{ij}^k$ which reflect the amount of flow on each arc $(i,j)$ for each commodity $k$, and 0-1 design variables $y_{ij}$, which indicate if arc $(i,j)$ is used or not ($N_i^+$ and $N_i^−$ are, respectively, the sets of outward and inward neighbors of node $i$):

$$\min \sum_{k \in K} \sum_{(i,j) \in A} c_{ij} x_{ij}^k + \sum_{(i,j) \in A} f_{ij} y_{ij}$$

(1)

$$\sum_{j \in N_i^+} x_{ij}^k - \sum_{j \in N_i^−} x_{ji}^k = \begin{cases} d^k, & \text{if } i = O(k), \\ -d^k, & \text{if } i = D(k), \\ 0, & \text{otherwise,} \end{cases} i \in N, k \in K,$$

(2)

$$\sum_{k \in K} x_{ij}^k \leq u_{ij} y_{ij}, \quad (i,j) \in A,$$

(3)

$$x_{ij}^k \geq 0, \quad (i,j) \in A, \quad k \in K,$$

(4)

$$0 \leq y_{ij} \leq 1, \quad (i,j) \in A,$$

(5)

$$y_{ij} \text{ integer} \quad (i,j) \in A.$$  

(6)
The objective function, (1), minimizes the sum of transportation costs and fixed costs. The flow conservation equations, (2), ensure that supplies and demands for all commodities are satisfied. Capacity constraints, (3), ensure that there is no flow circulating through a closed arc. Constraints (4)-(6) specify the nature of each variable.

This model can be solved with state-of-the-art MIP solvers. However, several difficulties arise. In particular, the linear programming (LP) relaxations are rather weak (optimality gaps of more than 20% are the norm, rather than the exception). Also, when the number of commodities is large, solving the LP relaxation becomes a hard task, because the multicommodity flow model that arises exhibits a lot of degeneracy. The combinatorial explosion observed when increasing the number of arcs also limits the size of the problems that can be solved. In general, the dominant factors in determining the complexity of any given instance are: relatively high fixed charges compared to transportation costs; tight capacities; a large number of commodities (say, more than 100).

Because of the weakness of the LP relaxation lower bounds, it is necessary to improve the model by adding valid inequalities that can be violated by the optimal solution to the LP relaxation. The general idea is then to use (some of) these valid inequalities in algorithms based on decomposition in mathematical programming. Before studying these algorithms, we first examine how to improve the model by adding valid inequalities.

3. Improving the model

A first class of valid inequalities, called the strong inequalities, is derived from the simple observation that no flow of any commodity can circulate on a closed arc:

$$x^k_{ij} \leq d^k_{ij}, \quad (i, j) \in A, k \in K.$$  \hspace{1cm} (7)

These inequalities are valid for the MIP, but not for its LP relaxation. In general, the addition of these inequalities significantly improves the LP relaxation lower bounds (optimality gaps below 10% are usually obtained). The model obtained by adding these inequalities will be called the strong model (with corresponding strong LP relaxation).

A second class of valid inequalities, called the cutset inequalities, is obtained by summing the flow conservation equations over any non-empty subset $S$ of $N$ (a cut) and by combining the resulting equation with the capacity constraints. By noting $S^c$, the complement of $S$, $(S, S)$ the set of arcs that connect a node in $S$ to a node in its complement (a cutset), $K(S, S)$ the set of commodities with the origin in $S$ and the destination in $S^c$, one has the following inequalities:

$$\sum_{(i, j) \in (S, S)} u_{ij}y_{ij} \geq \sum_{k \in K(S, S)} d^k, \quad S \subset N, S \neq \emptyset.$$ \hspace{1cm} (8)

These inequalities simply state that there should be enough capacity installed on the arcs of the cutset to satisfy the demands across the cut. The cutset inequalities are valid not only for the MIP, but also for its LP relaxation, since they are obtained by linear combinations of the flow conservation and capacity constraints. There are two approaches to obtain valid inequalities that are not redundant for the LP relaxation: one is to remark that each cutset inequality defines a 0-1 knapsack set and to use known knapsack inequalities, such as cover (Balas, 1975; Wolsey, 1975) and minimum cardinality inequalities (Martello and Toth, 1997); the other is to take one step back when summing the flow conservation equations over $S$ and to consider inequalities that also involve the flow variables (aggregated over any commodity subset $L$), i.e., the well-known flow cover and flow pack inequalities (Padberg, Van Roy and Wolsey, 1985; Atamtürk, 2001).

4. Cutting-plane method

Five sets of valid inequalities have been derived for the MCND: strong, cover, minimum cardinality, flow cover, and flow pack. The strong inequalities involve a polynomial, albeit large, number of inequalities, while the four other sets of inequalities, called cutset-based inequalities, have exponential size. A proven approach to include these inequalities when solving the model is the cutting-plane method, which consists in an iterative scheme that solves an LP relaxation at each iteration, verifies if there are valid inequalities violated by the optimal solution to the LP relaxation (the so-called separation step), and if so, adds the violated valid inequalities to the LP relaxation; otherwise, if no violated valid inequalities are found, the algorithm stops. The method is complemented with a
branch-and-bound algorithm to find an optimal (integer) solution; at each node of the branch-and-bound tree, the cutting-plane method can be repeated, giving rise to a branch-and-cut algorithm.

Chouman, Crainic and Gendron (2009) have implemented and tested this approach for the MCND. Their implementation involves tailored separation algorithms for generating cutset-based inequalities. In particular, single-node cuts are generated every iteration of the cutting-plane method, while a neighborhood search is used to generate cuts of cardinality 2 or more. For each set of cuts obtained by this neighborhood search, the four cutset-based inequalities are tested for violation using separation algorithms that exploit the structure of the MCND. The strong inequalities, since they can be generated efficiently, are always tested first, then the knapsack inequalities, and finally, the flow cover/pack inequalities. The computational results show that the strong, flow cover and flow pack inequalities are extremely effective in reducing the gaps, but that the strong inequalities can be generated more quickly (only some instances with few commodities, say less than 100, can benefit from the generation of flow cover/pack inequalities). The knapsack inequalities are easy to generate and can be used to restart the cutting-plane method when the search for strong violated inequalities has stopped. The model obtained at the end of the cutting-plane method was solved by branch-and-bound, obtaining results that are competitive with the state-of-the-art MIP solver CPLEX. An implementation of the cutting-plane method within a branch-and-cut algorithm is currently underway.

5. Benders decomposition

When fixing the design variables to some values, we obtain the following multicommodity flow subproblem:

$$z(\tilde{y}) = \min \sum_{k \in K} \sum_{(i,j) \in A} c_{ij}^k x_{ij}^k$$  \hspace{2cm} (9)

$$\sum_{k \in K} x_{ij}^k \leq u_{ij} \tilde{y}_{ij}, \quad (i, j) \in A,$$  \hspace{2cm} (10)

plus the flow conservation equations, (2), and the nonnegativity requirements on the flow variables, (4). This is a linear program, which is thus equivalent to its dual. Since one of the flow conservation equations, say the one associated to $O(k)$, is redundant for each commodity $k$, we can fix its corresponding dual variable to 0 and obtain the following dual problem:

$$z(\bar{y}) = \max \sum_{k \in K} d_k \bar{\pi}_D^k \left( k \right) - \sum_{(i,j) \in A} u_{ij} \bar{\pi}_{ij} \alpha_{ij}$$  \hspace{2cm} (11)

$$\bar{\pi}_j^k - \bar{\pi}_i^k - \alpha_{ij} \leq c_{ij}, \quad (i, j) \in A, k \in K,$$  \hspace{2cm} (12)

$$\alpha_{ij} \geq 0, \quad (i, j) \in A,$$  \hspace{2cm} (13)

$$\bar{\pi}_O^k = 0, \quad k \in K.$$  \hspace{2cm} (14)

There are two possibilities: either this problem is bounded, which means the multicommodity flow subproblem is feasible and the two problems have the same optimal value, or it is unbounded, which implies that the multicommodity flow subproblem is infeasible. In the first case, if we capture the optimal value of the multicommodity flow subproblem in a single variable $z$, we aim that $z$ never exceeds the optimal value associated to the fixed design variables, i.e,

$$\sum_{k \in K} d_k \bar{\pi}_D^k \left( k \right) - \sum_{(i,j) \in A} u_{ij} \bar{\pi}_{ij} \alpha_{ij} > z.$$  \hspace{2cm} (15)

In the second case, the optimal value of the dual associated to the fixed design variables is arbitrarily large, which means that there is a dual ray, i.e., a solution to the dual cone obtained by replacing $c_{ij}$ by 0 in (12), that satisfies

$$\sum_{k \in K} d_k \bar{\pi}_D^k \left( k \right) - \sum_{(i,j) \in A} u_{ij} \bar{\pi}_{ij} \alpha_{ij} > 0.$$  \hspace{2cm} (16)
The classical Benders decomposition algorithm exploits these results. At each iteration, the method solves a MIP relaxation of the problem, called the Benders master problem, which is defined using only the design variables, along with the variable $z$ capturing the optimal value of the multicommodity flow subproblem. After solving the Benders master problem, the design variables are fixed to the optimal solution thus obtained, and the corresponding multicommodity flow subproblem is solved. If the dual problem is bounded, a so-called optimality cut is generated by complementing inequality (15) and by replacing the fixed design variables by arbitrary values. If the dual problem is unbounded, a so-called feasibility cut is generated by complementing inequality (16) and by replacing the fixed design variables by arbitrary values. The generated cut is added to the Benders master problem, which is solved again to perform another iteration. Thus, the Benders master problem has the following form, where $E(P_D)$ and $R(P_D)$ are, respectively, subsets of (extreme point) solutions and (extreme) rays of the dual problem:

$$\min \sum_{(i,j) \in A} f_{ij}y_{ij} + z \quad (17)$$

$$\sum_{k \in K} d^k s^k_{D(k)} - \sum_{(i,j) \in A} u_{i,j}y_{ij}\alpha_{ij} \leq z, \quad (\pi, \alpha) \in E(P_D), \quad (18)$$

$$\sum_{k \in K} d^k \pi^k_{D(k)} - \sum_{(i,j) \in A} u_{i,j}y_{ij}\alpha_{ij} \leq 0, \quad (\pi, \alpha) \in R(P_D). \quad (19)$$

When the dual problem is bounded, since we are minimizing $z$, either the current fixed design values will be cut when generating the corresponding optimality cut or a solution with the same objective value will be generated, which implies that an optimal solution to the MCND has been found. When the dual problem is unbounded, the current fixed design values will be cut by the corresponding feasibility cut.

This classical Benders decomposition algorithm suffers from several drawbacks. In particular, the Benders master problem is a MIP, which can be computationally elusive. Also, the process might converge slowly, especially if only one Benders cut is generated per iteration. Costa, Cordeau and Gendron (2009) and Costa et al. (2011) have implemented several refinements to the classical Benders scheme. The Benders cuts are initialized with cutset inequalities of the form (8). Then, a number of Benders cuts are obtained by solving the strong LP relaxation of the MCND by Benders decomposition, rather than the MCND itself. Also, instead of solving the MIP Benders master problem at every iteration, a heuristic approach is also used to generate a large number of tentative fixed design values for which several Benders cuts are generated by solving the corresponding multicommodity flow subproblems. Finally, when solving the MIP Benders master problem by branch-and-bound, all feasible solutions, not only an optimal one, are used to generate several Benders cuts. In spite of all these refinements, the Benders decomposition algorithm appears computationally inferior to the cutting-plane method (complemented with branch-and-bound) described in the previous section. One of the reasons might be the lack of separability in the multicommodity flow subproblem; indeed, successful applications of Benders decomposition often display an inherent decomposition of the Benders subproblem into independent, smaller, subproblems from which several Benders cuts (one for each subproblem) can be generated. Nevertheless, Benders decomposition is an interesting approach for the MCND, since Benders cuts can be used in any cutting-plane approach (in particular, the one described in Section 4) performed in combination with a heuristic approach that identifies fixed design values, since any of these has a corresponding Benders cut that can be added to the model handled by the cutting-plane approach.

### 6. Lagrangian relaxation

Another approach to compute improved lower bounds on the optimal value of the MCND relies on the concept of Lagrangian relaxation: some constraints are relaxed by appending them to the objective function with associated Lagrange multipliers, thus obtaining so-called Lagrangian subproblems. The values of the Lagrange multipliers are gradually adjusted in such a way that constraint violations are penalized. We sought the best values for the Lagrange multipliers, i.e., the ones that maximize the lower bound obtained by Lagrangian relaxation. To obtain an optimal solution, this process must be completed by branch-and-bound, i.e., at each node of the branch-and-bound tree, the same Lagrangian relaxation approach is applied. In addition, the solutions to the Lagrangian subproblems might be used to guide the search for solutions in a heuristic method.
For the MCND, two Lagrangian relaxation methods have been extensively tested. Both methods use as a starting point the strong model for the problem, (1)-(6) + (7). The first approach, called the shortest path relaxation, consists in relaxing all constraints linking the flow and the design variables, i.e., the capacity constraints (3) and the strong inequalities (7). The Lagrangian subproblem has the following structure:

\[
\begin{align*}
  z(\alpha, \beta) &= \min \sum_{(i,j) \in A} \sum_{k \in K} (c_{ij} + \alpha_{ij} + \beta_{ij}^k)x_{ij}^k + \sum_{(i,j) \in A} (f_{ij} - u_{ij}\alpha_{ij} - \sum_{k \in K} d_k^k\beta_{ij}^k)y_{ij}
\end{align*}
\]

with constraints (2) and (4), only on the flow variables, and (5)-(6), only on the design variables. The first part, which depends only on the flow variables, decomposes into \(|K|\) shortest path problems; the second part can be solved by inspection of the sign of the Lagrangian cost associated to each design variable.

The second approach, called the knapsack relaxation, consists in relaxing the flow conservation equations (2). The Lagrangian subproblem has the following structure:

\[
\begin{align*}
  z(\pi) &= \min \sum_{(i,j) \in A} \sum_{k \in K} (c_{ij} + \pi_i^k - \pi_j^k)x_{ij}^k + \sum_{(i,j) \in A} f_{ij}y_{ij} + \sum_{k \in K} d_k^k(\pi_{D(k)}^k - \pi_{O(k)}^k)
\end{align*}
\]

with constraints (3) to (7). This problem decomposes into |A| subproblems, one for each arc; each such subproblem has a single 0-1 variable and \(|K|\) continuous bounded variables. If we fix the 0-1 variable to 0, all the continuous variables also assume value 0, and the corresponding objective value is 0; if we fix the 0-1 variable to 1, the optimal solution is obtained by solving a continuous knapsack problem. If the optimal value of this continuous knapsack problem is negative, then the optimal solution for the arc subproblem has the 0-1 variable equal to 1 and the continuous variables equal to the continuous knapsack solution; otherwise, the optimal solution for the arc subproblem has all the variables equal to 0.

Gendron and Crainic (1994) show that the two Lagrangian relaxation approaches provide the same theoretical lower bound, which is also the same as the one provided by solving the strong LP relaxation. Several algorithms have been used to solve the so-called Lagrangian dual, i.e., find the optimal values of the Lagrange multipliers. In particular, subgradient and bundle algorithms have been tested for the shortest path and knapsack relaxations for the MCND. Crainic, Frangioni and Gendron (2001) show the efficiency of the bundle algorithm when applied to the two relaxations. Their results also suggest that the knapsack relaxation is in general preferable, since both the Lagrangian subproblem and the bundle master problem are solved more quickly in that case. Holmberg and Yuan (2000) and Sellman, Kliwer and Koberstein (2002) have used the knapsack relaxation in a subgradient-based algorithm embedded in a branch-and-bound algorithm. Kliwer and Timajev (2005) have improved this Lagrangian-based branch-and-bound algorithm by replacing the subgradient algorithm with a bundle method. Their implementation is probably one of the best existing specialized algorithms to solve large-scale instances of the MCND to optimality.

7. Conclusion

The MCND is representative of a large class of network design models used in transportation and logistics, as it illustrates the tradeoff between investment and operational costs coupled with the presence of multiple commodities and capacities. We have seen that decomposition approaches are essential if one wishes to solve such problems to optimality or at least find provably good solutions. The literature contains a number of contributions regarding the development of such decomposition methods. We have reviewed three of them: a cutting-plane method, a Benders decomposition algorithm, and Lagrangian relaxation approaches. Progresses are still being made and will soon be reported.

More generally, an increasing number of network design applications in transportation and logistics are being studied. Of prime importance is the development of models and methods that integrate the dynamic aspect, as well as the inherent uncertainty, of the applications. In particular, network design problems in an operational context have recently appeared; in particular, fascinating “city logistics” applications emerge as part of a global effort to improve the distribution of freight in urban areas.
References

Publication CIRRELT-2009-20, Centre interuniversitaire de recherche sur les réseaux d’entreprise, la logistique et le transport.
Publication CIRRELT-2009-03, Centre interuniversitaire de recherche sur les réseaux d’entreprise, la logistique et le transport.