MODELING AND SOLVING THE TWO-FACILITY CAPACITATED NETWORK LOADING PROBLEM

THOMAS L. MAGNANTI

Massachusetts Institute of Technology, Cambridge, Massachusetts

PRAKASH MIRCHANDANI

University of Pittsburgh, Pittsburgh, Pennsylvania

RITA VACHANI

GTE Laboratories Incorporated, Waltham, Massachusetts

(Received February 1991; revision received July 1992; accepted February 1993)

This paper studies a topical and economically significant capacitated network design problem that arises in the telecommunications industry. In this problem, given point-to-point communication demand in a network must be met by installing (loading) capacitated facilities on the arcs: Loading a facility incurs an arc specific and facility dependent cost. This paper develops modeling and solution approaches for loading facilities to satisfy the given demand at minimum cost. We consider two approaches for solving the underlying mixed integer program: a Lagrangian relaxation strategy, and a cutting plane approach that uses three classes of valid inequalities that we identify for the problem. We show that a linear programming formulation that includes these inequalities always approximates the value of the mixed integer program at least as well as the Lagrangian relaxation bound. Our computational results on a set of prototypical telecommunication data show that including these inequalities considerably improves the gap between the integer programming formulation and its linear programming relaxation: from an average of 25% to an average of 8%. These results show that strong cutting planes can be an effective modeling and algorithmic tool for solving problems of the size that arise in the telecommunications industry.

Private networks are playing an increasingly important role in carrying the communications traffic of large organizations because of the economic and strategic benefits they offer when compared to the public switched network. Private lines are transmission facilities that customers lease from a telephone company for their exclusive use. Because these lines are billed on a fixed (nonusage sensitive) rate, an organization might find it cheaper to lease a private line between any two locations that have a large amount of traffic between them rather than pay on a per-usage basis for this traffic. In addition, private networks offer customers greater flexibility to reconfigure the network to accommodate changes in traffic patterns and provide improved reliability. Due to rapid technological changes in the telecommunications industry, telephone companies are offering high bandwidth (capacity) facilities to private subscribers which allow the customers to use the private networks for a variety of applications, including voice, data, and video transfer. As a result, the demand for private lines has been increasing rapidly and is expected to continue to do so over the next five to ten years.

Private networks have dedicated access lines from the customer premises to the nearest telephone company switch (central office) and dedicated lines between the central offices that connect different locations of the customer's organization. The digital facilities that a customer leases to and between the central offices are selected from a small set of alternatives, for example, DS0 (Digital Signal Level 0), which transmits at the rate of 64 kilo bits per second (kbps), DS1 (Digital Signal Level 1), which offers the capacity equivalent to 24 DS0 facilities, and in some cases, DS3 (Digital Signal Level 3), which is equivalent (in capacity terms) to 28 DS1 facilities. The tariffs for these facilities are complex and offer strong economies of scale; typically, a DS1 circuit costs the same as only 8 to 10 DS0 circuits, depending on the length of the circuit.

One fundamental problem arises in designing a private communication network: Given an organization's forecast for traffic between its various locations, what configuration of digital transmission facilities between the central offices (referred to as the backbone network) should be leased to carry this traffic at minimum cost? The cost of any private network corresponds to the leasing cost of the facilities installed on the arcs; the user incurs no additional routing cost. This problem is difficult because of the complexity of the cost structure. The optimal solution might use complicated routes for the different commodities because by aggregating traffic on some arcs, it can take advantage of the economies of scale in the tariff structure.

Subject classifications: Networks/graphs, applications: design of telecommunications networks. Programming, integer, cutting plane/facet generation: facets for capacitated network design problems. Programming, integer, relaxation/subgradient: comparison of Lagrangian and polyhedral approaches.
Area of review: TELECOMMUNICATIONS (SPECIAL ISSUE ON TELECOMMUNICATION SYSTEMS: MODELING, ANALYSIS AND DESIGN).

0030-364X/95/4301-0142 $01.25

A similar problem arises in the context of transportation planning; in this setting, the traffic corresponds to freight and the transmission facilities to different types of trucks. These problems have substantial economic significance. For example, revenues to the long distance carriers from the lease of digital transmission circuits used for private communications networks were about \$1.7 billion per annum at the start of the study (*Business Communications Review*, May 1990), and these were expected to grow at 30–40% per annum by one estimate (*Telecommunications*, North American Edition, May 1990). In the transportation context, the total estimated expenditure on trucking was estimated to be \$313 billion in 1993 and is expected to grow at an annual rate of 6% (*US Industrial Outlook*, U.S. Department of Commerce, January 1994).

The objective of this paper is to develop modeling and solution approaches for the network loading problem, which includes the private network leasing problem as a special case. In the transportation context, a slight variation of the model would prescribe a load plan (the assignment of trucks to routes) and the loading of freight onto trucks. Since the network loading model that we study is a special version of the general capacitated network design problem, we hope that this paper might also provide some useful insights for solving the more general problem.

The rest of the paper is organized as follows. Section 1 presents a formal description and formulation of the network loading problem and briefly reviews the literature on similar models. In Section 2, we discuss alternative solution strategies for the problem and provide motivation for our proposed solution approach. Section 3 provides a partial characterization of the mixed integer polyhedron that models the problem, and Section 4 describes the solution methodology and presents our computational results. The last section presents our conclusions, and suggests some future research directions.

1. NETWORK LOADING PROBLEM: DESCRIPTION AND FORMULATION

The network loading problem models the design of capacitated networks for which the variable flow costs are zero, and facilities of fixed capacity are available to carry flow. We can install (load) these facilities on any arc of the network. The problem is to determine the number of facilities to be loaded on each arc of the network to meet given point-to-point communication demand at minimum cost. In this paper, we assume that only two types of facilities are available. In the context of the private network leasing problem, the two facilities could correspond to DS0 and DS1 circuits, which were the facilities most widely available at the start of this study, or to DS1 and DS3 circuits, the facilities most widely leased currently. In general, we may have a choice of facilities with capacities at many different levels; Mirchandani (1989) shows how our results generalize for the case of multiple facilities. This model extension would permit us to consider emerging industry practice in which telephone companies are beginning to offer DS3 facilities to private subscribers on selected segments.

We model the network loading problem with two facilities, which we refer to as the **TFLP** (for the two-facility loading problem), as follows.

TFLP

Minimize

$$\sum_{\{i,j\}\in A} (a_{ij}x_{ij} + b_{ij}y_{ij})$$

subject to

$$\sum_{j\in N} f_{ji}^k - \sum_{j\in N} f_{ij}^k = \begin{cases} -d_k & \text{if } i = O(k) \\ d_k & \text{if } i = D(k) \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } i \in N, \text{ for all } k \in K \quad (1)$$

$$\sum_{k\in K} (f_{ij}^k + f_{ji}^k) \le x_{ij} + Cy_{ij} \quad \text{for all } \{i, j\} \in A \quad (2)$$

$$x_{ij}, y_{ij} \ge 0 \text{ and integer for all } \{i, j\} \in A;$$

$$f_{ij}^k, f_{ji}^k \ge 0 \text{ for all } \{i, j\} \in A, \text{ for all } k \in K. \quad (3)$$

In this formulation, N denotes the set of nodes of the network, A the set of arcs, and K the set of commodities; commodity k has origin $O(k)$, destination $D(k)$, and demand d_k. We refer to the two types of facilities as the low capacity (LC) and the high capacity (HC) facilities; the LC facility has capacity 1 and the HC facility has capacity C. (For the telecommunications private network leasing problem with DS0 and DS1 facilities, $C = 24$ for problems with DS1 and DS3 facilities, $C = 28$.) The formulation contains two sets of variables: design variables x_{ij} and y_{ij} that define the number of LC and HC facilities loaded on the undirected arc $\{i, j\}$, and flow variables f_{ij}^k that model the flow of commodity k on arc $\{i, j\}$ in the direction i to j. The coefficients a_{ij} and b_{ij} represent the cost of loading a single LC and HC facility, respectively, on arc $\{i, j\}$ and the objective function minimizes the total cost incurred in loading all the facilities. Constraints (1) correspond to the usual flow conservation constraints for each commodity at each node. The capacity constraints (2) model the requirement that the total flow (in both directions) on an arc cannot exceed the capacity loaded on that arc.

Note that this formulation assumes that the traffic for each origin-destination pair is known and deterministic and the capacity utilization on every arc can be 100%. In the context of communications networks, analysts would specify the model parameters using preliminary analysis of the expected busy hour traffic between different locations. The first step of this analysis would be to determine the characteristics for all types of traffic (voice, video or data) such as the average length of a call and the maximum blocking probability of a call for voice traffic. The next step would be to estimate the number of circuits required between the different origin-destination pairs to serve this traffic with the specified blocking probability or the specified maximum delays. The required number of circuits derived from this analysis form the demand input for the **TFLP**. Other application contexts might require some other form of preliminary analysis to calibrate the parameters of the **TFLP**.

Note that our model assumes it is sufficient to install just enough capacity to meet demand—we do not need to provide extra capacity to address reliability issues. Although this assumption is valid in some situations, such as the transportation of freight, it might be less so in the telecommunications setting where the network reliability is of greater concern. However, our discussions with planners in the telecommunications industry indicate that the model we are considering is valuable as a first-cut design tool for their planning activities and that they would typically use other ancillary models to address reliability issues. (For an approach to telecommunications network survivability problems, see Groetschel, Monma and Stoer 1992 and Balakrishnan, Magnanti and Mirchandani 1994.)

Despite the importance of this network design problem in a variety of settings, the available research on it is quite limited. Research in the network design area has primarily focused on the uncapacitated, fixed-charge design problem for which the gap between the disaggregate integer formulation of the problem and its linear programming relaxation is small (see Balakrishnan, Magnanti and Wong 1989). The algorithms and heuristics developed for this uncapacitated problem work well because of this small integrality gap, which is not true for the **TFLP**, as we point out later.

Previously, in the context of either computer networks or delivery and transportation systems, several researchers have studied capacitated network design problems with a limited number of capacity options available for each arc. Gavish and Altinkemer (1990) discuss a problem similar to the **TFLP**. This paper addresses the design of a computer communication backbone network. The problem is to simultaneously assign capacities to the arcs and select the single route to be used by each commodity. The relevant costs are the fixed costs of the cables installed on the arcs, variable costs proportional to the flow on each arc, and a delay cost due to queueing that varies with the capacity utilization of each arc. The computational results for this study for situations with zero variable costs and zero delay costs (corresponding to our model) show large gaps between the heuristic solutions and the Lagrangian lower bounds, indicating that this is a hard combinatorial problem. Gavish and Altinkemer's work improves upon an earlier method proposed by Gavish and Neuman (1989) for the same problem. Maruyama, Fratta and Tang (1977), Gerla and Kleinrock (1977), Ng and Hoang (1987), and LeBlanc and Simmons (1989) have conducted work on related problems.

In the context of transportation systems, Leung, Magnanti and Singhal (1990) also use a Lagrangian relaxation solution approach in their study of a related model. Their results confirm that the gaps between the heuristic solutions and the lower bounds are large, indicating that further work is needed to obtain better problem representation and sharper lower bounds.

The next section discusses our solution approach for the **TFLP** and provides motivation for why we chose to investigate this particular approach.

2. MODELING AND SOLUTION APPROACHES: MOTIVATION

In general, linear programming lower bounds are weak for most capacitated network design problems (including the **TFLP**) and so these models are much more difficult to solve than the corresponding uncapacitated problems. Our objective is to develop stronger formulations for the **TFLP** than its linear programming relaxation and use these results to develop more efficient solution techniques than a linear programming based branch-and-bound procedure.

One approach to obtaining stronger lower bounds is embodied in the burgeoning field of polyhedral combinatorics, which attempts to improve the linear programming approximation to an integer programming problem by adding (strong) valid inequalities, either a priori to the original formulation of the problem, or dynamically, via a cutting plane approach, to a series of linear programming models. This methodology has been very successful in solving problems in a variety of contexts: the traveling salesman problem (Groetschel and Padberg 1985), the economic planning and linear ordering problem (Groetschel, Juenger and Reinelt 1985a, b), production planning models (Barany, Van Roy and Wolsey 1984, Magnanti and Vachani 1990), the fixed charge problem (Padberg, Van Roy and Wolsey 1983), the lot sizing problem (Leung, Magnanti and Vachani 1989, and Pochet 1988), the spin glass problem (Barahona et al. 1988), and models for planning capacity expansion in local access telecommunications systems (Balakrishnan, Magnanti and Wong 1995a, b).

Lagrangian relaxation is another successful approach for obtaining tight lower bounds for integer programming problems (see, for example, Gavish and Neuman 1989, Gavish and Altinkemer 1990, for applications in the network design area; Geoffrion and McBride 1978, for the capacitated facility location problem; and Graves 1982, for a hierarchical production planning problem). For an application-oriented survey of Lagrangian relaxation, see Fisher (1981).

In this section, we illustrate and compare both these solution approaches as applied to the **TFLP** and then discuss how we could combine the two approaches to obtain strong lower bounds. First, consider a *Lagrangian relaxation approach* to solving the **TFLP**. When using this approach, we can dualize either constraints (1) or (2). If we relax (2), because the resulting Lagrangian problem is a network flow problem that satisfies the integrality property (i.e., its linear programming relaxation has an integer optimal solution), the Lagrangian dual problem gives the same lower bound as the linear programming relaxation of **TFLP** (Geoffrion 1974). On the other hand, if we relax (1) using multipliers v_i^k, then the resulting Lagrangian problem, with $v_{O(k)}^k = 0$, is:

minimize

$$\sum_{\{i,j\}\in A} \left\{ a_{ij}x_{ij} + b_{ij}y_{ij} + \sum_{k\in K} (f_{ji}^k - f_{ij}^k)(v_i^k - v_j^k) \right\} + \sum_{k\in K} v_{D(k)}^k d_k$$

subject to (2) and (3).

Note that we can set $v_{O(k)}^k = 0$ without any loss of generality because one of the flow conservation constraints in **TFLP** is redundant and, hence, we can set the value of one of the dual variables arbitrarily. To this relaxed problem, we add the following *upper bound* constraints.

$$f_{ij}^k + f_{ji}^k \leq d_k \quad \text{for all } \{i, j\} \in A, \text{ for all } k \in K. \tag{4}$$

These constraints do not affect the optimal objective value of the original formulation (because we can always delete the flow around cycles for any commodity), but improve the Lagrangian lower bound. We refer to the resulting Lagrangian problem as **P(LAG)**.

P(LAG) does not satisfy the integrality property and, therefore, we can expect the lower bound obtained from the Lagrangian dual to be stronger than that obtained from the linear programming relaxation of **TFLP**. Note that **P(LAG)** decouples into separate subproblems, one for each arc of the network. The subproblem for each arc is a knapsack-type problem that can be solved efficiently by an incremental strategy of "loading" the "profitable" commodities (relative to the facility costs) on each arc. Vachani (1988) uses this Lagrangian relaxation strategy, with subgradient optimization to update the Lagrange multipliers and improve the Lagrangian bound, to solve the **TFLP**. Her results show that the lower bounds from using this approach significantly improve upon the linear programming relaxation bounds.

The second, *polyhedral*, approach for obtaining better lower bounds uses results about the polyhedral structure of the problem to strengthen the formulation. To illustrate this approach, consider a 3-node network with an arc between every pair of nodes and a demand of d between every pair of nodes. Assume that only HC-type facilities are available and that the facility cost is the same on all the arcs, i.e., $b_{ij} = 20$ for all $\{i, j\}$. Since the demand between node 1 and the other two nodes of the network (nodes 2 and 3) is $2d$ units, the network must contain at least $\lceil 2d/C \rceil$ HC facilities between node 1 and the other two nodes to carry this traffic. Thus, the constraint

$$y_{12} + y_{13} \geq \lceil 2d/C \rceil \tag{5}$$

is valid for the problem. Similar constraints for the other two nodes are also valid. Note that the linear programming relaxation of the **TFLP** requires only that

$$y_{12} + y_{13} \geq 2d/C$$

and thus, the (cutset) constraint (5) will strengthen the linear programming relaxation if $\lceil 2d/C \rceil$ is significantly larger than $2d/C$. For example, assume that $C = 24$. If $d = 13$, then the addition of these three constraints (one for each node) in the linear programming relaxation of **TFLP** is sufficient to obtain an optimal integer solution, whereas if $d = 12$ these constraints are not effective at all.

On the other hand, the Lagrangian relaxation approach applied to the same 3-node example with $d = 12$ results in a Lagrangian dual value of 40 (equal to the optimal solution value); however, when $d = 13$, the Lagrangian dual value (obtained by using subgradient optimization to update the Lagrange multipliers) is about 42, substantially lower than the optimal solution value of 60.

This example shows that both solution approaches result in stronger lower bounds than the linear programming relaxation, though their performance depends upon the problem parameters, in this case the values of d and C. One of them is more effective when $d = 12$, and the other is more effective for $d = 13$. However, changing the demand to 8 units shows that neither approach generates a good lower bound. With $d = 8$, the optimal solution value for the linear programming relaxation is 20, the optimal integer solution has a value of 40; the Lagrangian dual value is 30, and the linear program with constraints (5) included has an optimal value of 30.

Thus, our discussion of the two different solution approaches raises the following questions: Can we identify situations in which one approach is likely to provide better lower bounds than the other and, more importantly, can we combine the two approaches to obtain lower bounds stronger than those that would be obtained from using either approach by itself? This paper develops one way of combining the two approaches.

The strategy that we adopt to combine the two approaches is to identify facets of the arc subproblems of the Lagrangian relaxation that are valid inequalities (in fact, facets) for the **TFLP**. We use these inequalities to strengthen the formulation of the **TFLP**. These inequalities apply to *individual* arcs of the network and relate the flow of the commodities on the arc with the capacity on the arc; in contrast (the cutset) inequalities (5) apply to a *set* of arcs across a cutset. In fact, we identify inequalities that together with constraints (2) and (4) completely characterize the convex hull of the feasible solutions to the single-arc subproblems of the Lagrangian relaxation. We also identify another class of valid inequalities for the problem that are obtained by partitioning the network into three sets of nodes. The next section provides our main technical results and the following section then discusses the solution method in more detail.

3. POLYHEDRAL RESULTS

Mirchandani has shown that the **TFLP** is strongly NP-hard by transforming the 3-partition problem to it. Hence, we do not expect to be able to provide a complete characterization of the convex hull of its feasible solutions (Groetschel, Lovasz and Schrijver 1981, Karp and Papadimitriou 1982). In this section, we derive three classes of valid inequalities for the **TFLP**—the cutset, arc residual capacity, and 3-partition inequalities—and prove that these inequalities define facets of the underlying polyhedron. The partial characterization that we obtain not only guarantees a linear programming lower bound that is at least as strong as that obtained using Lagrangian relaxation to dualize constraints (1), but is also, as our computational results show, sufficient to reduce the integrality gap significantly. For details of polyhedral terminology used in this paper, we refer the reader to Schrijver (1986), Nemhauser and Wolsey (1988) and Pulleyblank (1989).

Let Conv(**TFLP**) denote the convex hull of feasible solutions to **TFLP** and let dim(Conv(**TFLP**)) denote its dimension. The formulation **TFLP** contains $2^*|A| + 2^*|A|^*|K|$ variables and $(|N| - 1)^*|K|$ nonredundant equality constraints. Therefore, dim(Conv(**TFLP**)) $\leq 2^*|A| + 2^*|A|^*|K| - (|N| - 1)^*|K|$. Proposition 1 shows that dim(Conv(**TFLP**)) is exactly equal to this bound. This proof uses arguments similar to those used in Theorem 1 and we, therefore, omit it.

Proposition 1. Dim(Conv(**TFLP**)) $= 2^*|A| + 2^*|A|^*|K| - (|N| - 1)^*|K|$.

3.1. The Cutset Inequality

The previous section introduced the cutset inequality (5) for problems with only HC-type facilities. The *cutset inequality* for the **TFLP**, i.e., for situations with both LC- and HC-type facilities, is described by

$$X_{S,T} + rY_{S,T} \geq r\lceil D_{S,T}/C\rceil \quad \text{for all } S,\ T;\ S \in N,\ T = N\backslash S. \qquad (6)$$

In this expression, the node set N is partitioned into nonempty subsets S and T. $X_{S,T}$ is the number of LC facilities on arcs across the cutset $\{S, T\}$, i.e., $X_{S,T} = \sum_{\{i,j\}\in\{S,T\}} x_{ij}$, $Y_{S,T}$ denotes the number of HC facilities on all arcs across the cutset $\{S, T\}$, i.e., $Y_{S,T} = \sum_{\{i,j\}\in\{S,T\}} y_{ij}$, $D_{S,T}$ denotes the demand of all commodities with origin and destination in different subsets, i.e., with $O(k) \in S$ ($O(k) \in T$) and $D(k) \in T$ ($D(k) \in S$), and $r = D_{S,T} \bmod(C)$. By convention, we set $r = C$ if $D_{S,T}$ is an integer multiple of C; thus, in all cases, $D_{S,T} = (\lceil D_{S,T}/C\rceil - 1)C + r$.

Proposition 2. *The cutset inequality* (6) *is valid for the* **TFLP**.

Proof. For any feasible solution to the problem, the aggregate capacity across the cutset must be no less than the demand across the cutset. Thus, the "aggregate capacity demand inequality" is

$$X_{S,T} + CY_{S,T} \geq D_{S,T}. \qquad (7)$$

If $r = C$, then the cutset inequality is valid because it is equivalent to the aggregate capacity demand inequality (7). Now assume that $r < C$ and let (x^*, y^*) be any feasible solution for **TFLP**. If $Y^*_{ST} \geq \lceil D_{S,T}/C\rceil$, then the solution trivially satisfies inequality (6). If $Y^*_{ST} = \lceil D_{S,T}/C\rceil - 1$, then the capacity of the HC facilities across the cutset is $D_{S,T} - r$, and so the capacity of the LC facilities across the cutset must be at least r and inequality (6) is again valid. This reasoning accounts for the coefficient r in the cutset inequality. On the other hand, if $Y^*_{ST} = \lceil D_{S,T}/C\rceil - p$ for some integer $p \geq 2$, the capacity of the HC facilities across the cutset is $D_{S,T} - r - (p - 1)C$ and the capacity of the LC facilities must be at least $r + (p - 1)C$. In this case, the cutset inequality, which becomes $X^*_{ST} \geq rp$, is weak and so it is valid.

This proof is similar to the one used by Leung, Magananti and Vachani in their study of capacitated production planning problems.

Mirchandani describes an alternate proof (with a noteworthy graphical interpretation) for establishing the validity of inequality (6) that is based on the Chvátal–Gomory procedure (see, for example, Nemhauser and Wolsey) for tightening a set of inequalities.

The proof of Proposition 2 shows that the cutset inequality (6) is stronger than the aggregate capacity demand inequality (7) when $0 \leq X_{S,T} < r$, and the next theorem shows that it defines a facet of Conv(**TFLP**) under fairly mild conditions.

Theorem 1. *The cutset inequality* (6) *is a facet of Conv*(**TFLP**) *if and only if*

i. $D_{S,T} > 0$, *and*
ii. *the subgraphs defined by S and T are connected.*

Proof. (**Necessity**) If $D_{S,T} = 0$, then (6) is a linear combination of the nonnegativity constraints and hence, cannot be a facet. To establish necessity of the second condition, assume without loss of generality, that S is not connected. Let $S = U \cup V$, where U and V are two "separated" components of S satisfying the conditions $U \cap V = \phi$ and the arc set $\{U, V\} = \phi$. To simplify the notation in this proof, let $r_U \equiv r_{U,\{V\cup T\}}$, $r_V \equiv r_{V,\{U\cup T\}}$, and $r_S \equiv r \equiv r_{\{U\cup V\},T}$. Define single indexed aggregate design and demand variables similarly. Note that the definition of separated components implies that $X_S = X_U + X_V$ and $Y_S = Y_U + Y_V$. Moreover, if $D_U = 0$ (or $D_V = 0$), then the inequality corresponding to the cutset

$\{V, U \cup T\}$ (or $\{U, V \cup T\}$) is tighter than the inequality corresponding to $\{S, T\}$. So, assume that $D_U > 0$ and $D_V > 0$. In addition, we can assume that $r < C$.

Now, either $r = r_U + r_V$ or $r = (r_U + r_V) \bmod(C)$. If $r = r_U + r_V$, then an argument similar to that used to prove Proposition 1 shows that the inequalities

$$X_U + rY_U \geq r\left\lfloor\frac{D_U}{C}\right\rfloor + r_U \text{ and } X_V + rY_V \geq r\left\lfloor\frac{D_V}{C}\right\rfloor + r_V$$

are valid. Adding these inequalities, we obtain inequality (6); thus, it cannot be a facet.

On the other hand, if $r = (r_U + r_V) \bmod(C)$, then $r \leq \min(r_U, r_V)$. Therefore,

$$X_U + rY_U \geq r\left\lceil\frac{D_U}{C}\right\rceil \text{ and } X_V + rY_V \geq r\left\lceil\frac{D_V}{C}\right\rceil.$$

Adding these inequalities and noting that $\lceil D_U/C\rceil + \lceil D_V/C\rceil \geq \lceil(D_U + D_V/C)\rceil \geq \lceil D_S/C\rceil$, we obtain inequality (6); thus, it cannot be facet defining in this case either.

(Sufficiency) Define the face

$L = \{(\mathbf{x}, \mathbf{y}, \mathbf{f}) \in \text{Conv}(\textbf{TFLP}): (\mathbf{x}, \mathbf{y}, \mathbf{f})$ satisfies (6) as an equality$\}$.

Let

$$\sum_{\{i,j\}\in A} \alpha_{ij}x_{ij} + \sum_{\{i,j\}\in A} \beta_{ij}y_{ij} + \sum_{k\in K}\sum_{\{i,j\}\in A} (\gamma_{ij}^k f_{ij}^k + \gamma_{ji}^k f_{ji}^k) \geq \delta \qquad (8)$$

with α_{ij}, β_{ij}, γ_{ij}^k and $\delta \in \Re^+$, represent an arbitrary inequality that is satisfied as an equality by all $(\mathbf{x}, \mathbf{y}, \mathbf{f}) \in L$. We show that (8) must be a linear combination of (6) and the other equality constraints of **TFLP**, which establishes that $\dim L = \dim(\text{Conv}(\textbf{TFLP})) - 1$, or equivalently, that (6) is a facet of Conv(**TFLP**). We use the following interchange argument to prove this result. Suppose that two vectors $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{f}^1)$ and $(\mathbf{x}^2, \mathbf{y}^2, \mathbf{f}^2)$ belong to L, and every component of $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{f}^1)$ equals the corresponding component of $(\mathbf{x}^2, \mathbf{y}^2, \mathbf{f}^2)$ except for components x_{pq}^1 and x_{pq}^2. Substituting these two solutions in (8) and subtracting the resulting equations, we obtain $\alpha_{pq} = 0$. On the other hand, if all components of $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{f}^1)$ and $(\mathbf{x}^2, \mathbf{y}^2, \mathbf{f}^2)$ are equal except that $x_{rs}^1 = x_{pq}^2 = 0$ and x_{pq}^1 and $x_{rs}^2 > 0$ (i.e., we have interchanged x_{pq}^1 LC facilities in $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{f}^1)$ with x_{rs}^2 LC facilities in $(\mathbf{x}^2, \mathbf{y}^2, \mathbf{f}^2)$), then a similar substitution of both solutions in (8) shows that $\alpha_{pq}/\alpha_{rs} = x_{rs}^2/x_{pq}^1$.

Construct a feasible solution $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{f}^0)$ satisfying (6) as an equality as follows. For all commodities $k \in \{S, S\}$ (or $k \in \{T, T\}$) connect $O(k)$ and $D(k)$ by $\lceil d_k/C\rceil$ HC facilities along a path fully contained in S (in T). This choice is possible because of condition ii of the theorem. Send a flow of d_k along this path from $O(k)$ to $D(k)$. Choose a node $u \in S$ and a node $v \in T$ for which $\{u, v\} \in A$. Install $\lceil D_{S,T}/C\rceil$ HC facilities on arc $\{u, v\}$. All commodities $k \in \{S, T\}$ flow from S to T across this arc $\{u, v\}$. Since the network is undirected, we can assume without loss of generality that $O(k) \in S$ for all $k \in \{S, T\}$. For each $k \in \{S, T\}$, connect $O(k)$ to u by $\lceil d_k/C$ HC facilities installed on a path $\{O(k), \ldots, u\}$ fully contained in S and v to $D(k)$ by $\lceil d_k/C\rceil$ HC facilities installed on a path $\{v, \ldots, D(k)\}$ fully contained in T. Send a flow of d_k along these paths. Let $(f_{uv}^k)^0 = d_k$ for all $k \in \{S, T\}$. This solution is feasible, has $X_{S,T} = 0$, $Y_{S,T} = \lceil D_{S,T}/C\rceil$, and satisfies (6) as an equality.

Using the interchange argument with one of the solutions as $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{f}^0)$, we can show that

1. $\alpha_{ij} = \beta_{ij} = 0$ for all $\{i, j\} \in \{S, S\}$ or $\{T, T\}$,
2. $r\alpha_{uv} = \beta_{uv}$, and since the choice of arc $\{u, v\}$ is arbitrary, $r\alpha_{ij} = \beta_{ij}$ for all $\{i, j\} \in \{S, T\}$, and
3. $\gamma_{ij}^k = -\gamma_{ji}^k$ for all $\{i, j\} \in \{S, S\}$ or $\{T, T\}$ for all $k \in K$.

Now, consider arc $\{u, v\}$. Since $r < C$ and $D_{S,T} > 0$, the $\lceil D_{S,T}/C\rceil$ HC facilities installed on arc $\{u, v\}$ have a residual capacity of at least one unit. So define, for some $k_1 \in K$ and $0 < \epsilon \leq 1/2$, a solution $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{f}^1) \in L$ with $\mathbf{x}^1 = \mathbf{x}^0$, $\mathbf{y}^1 = \mathbf{y}^0$, and $(f_{uv}^{k_1})^1 = (f_{uv}^{k_1})^0 + \epsilon$, $(f_{vu}^{k_1})^1 = (f_{vu}^{k_1})^0 + \epsilon$, $(f_{ij}^k)^1 = (f_{ij}^k)^0$ otherwise. Comparing $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{f}^1)$ with $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{f}^0)$, we see that $\gamma_{uv}^k = -\gamma_{vu}^k$ for $k = k_1$. But since k_1 and $\{u, v\}$ are arbitrary, $\gamma_{ij}^k = -\gamma_{ji}^k$ for all $\{i, j\} \in \{S, T\}$ and for all $k \in K$.

We now show that $\sum_{k\in K}\sum_{\{i,j\}\in A} (\gamma_{ij}^k f_{ij}^k + \gamma_{ji}^k f_{ji}^k)$ is a constant by showing that the sum of the γ coefficients corresponding to any cycle in the network equals zero. Let Δ denote the set of (directed) cycles in the network using two directed arcs instead of each undirected arc in A. We denote the arc directed from i to j by (i, j). We can assume that each directed arc is traversed at most once for all the cycles belonging to Δ.

Consider a particular cycle $\zeta \in \Delta$. Let $\gamma_\zeta^k = \sum_{(i,j)\in\zeta} \gamma_{ij}^k$. Call ζ an *s-intersection cycle* if it contains exactly s directed $\{S, T\}$ cutset arcs. Note that s must be even because ζ is a cycle. If ζ is a 0 intersection cycle, i.e., it does not contain any arcs in the cutset, change $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{f}^0)$ as follows: Install one additional LC facility on all arcs in the cycle and for some $k_1 \in K$, send an additional 1/2 unit of flow around the cycle. This solution maintains feasibility and satisfies (6) as an equality (i.e., belongs to L). Comparing the coefficients of $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{f}^0)$ with the modified solution, we find that $\gamma_\zeta^{k_1} = 0$ because $\alpha_{ij} = 0$ for all $\{i, j\} \in \{S, S\}$ or $\{T, T\}$. Thus, $\gamma_\zeta^k = 0$ for all 0-intersection cycles for all $k \in K$.

Now, suppose that ζ is a 2-intersection cycle. Assume that (u, v) and (p, q) are the cutset arcs belonging to cycle ζ. If $\{u, v\} = \{p, q\}$, then ζ can be decomposed into three cycles, one in S, the other in T, and the cycle (u, v), (v, u). The cycles contained in S and T are 0-intersection cycles and we have shown that $\gamma_{ij}^k = -\gamma_{ji}^k$ for all $\{i, j\} \in \{S, T\}$ and for all $k \in K$. Thus, $\gamma_\zeta^k = 0$ for all $k \in K$. If $\{u, v\} \neq \{p, q\}$, then construct a solution $(\mathbf{x}^2, \mathbf{y}^2, \mathbf{f}^2)$ as follows: Send the flow of commodities belonging to $\{S, S\}$ (or $\{T, T\}$) on paths fully contained in S (or T), as in solution $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{f}^0)$. Let $y_{uv} = \lceil D_{S,T}/C\rceil - 1/2$ and $y_{pq} = 1/2$. By installing additional facilities on arcs in $\{S, S\}$ and $\{T, T\}$, route the commodities belonging to $\{S, T\}$ so that each of the arcs

$\{u, v\}$ and $\{p, q\}$ contains at least 1/2 units of residual capacity. We can determine a feasible flow that meets this condition because $r < C$. Note that this solution belongs to L, and we can modify this solution to obtain another solution $\in L$ that differs from $(\mathbf{x}^2, \mathbf{y}^2, \mathbf{f}^2)$ only in two ways: There is an additional flow of 1/2 units along ζ for some $k_1 \in K$, and an additional LC facility is installed on all arcs in the cycle that are not cutset arcs. The cycle flow is possible without increasing the capacity on the cutset arcs because of the residual capacity on those arcs. Comparing the coefficients of $(\mathbf{x}^2, \mathbf{y}^2, \mathbf{f}^2)$ with the modified solution, we conclude that $\gamma_\zeta^k = 0$ for all 2-intersection cycles for all $k \in K$.

Now consider an arbitrary s-intersection cycle ζ. We will show that some 0-intersection cycle, say ψ, satisfies $\gamma_\zeta^k = \gamma_\psi^k$. Since we have already shown that $\gamma_\psi^k = 0$, this result would complete our argument. Let ζ be defined by $\{(r_1, r_2), (r_2, r_3), \ldots, (r_t, r_1)\}$ with $r_1 \in S$. Let (r_i, r_j) be the first arc of the cycle that crosses the $\{S, T\}$ cutset and let (r_k, r_l) be the first subsequent arc that re-enters the set S. Note that node r_i can equal r_l and/or r_k can equal r_j. In all cases $\{(r_i, r_j), \ldots, (r_k, r_l), (r_l, r_i)\}$ is a 2-intersection cycle. (Arc $\{r_l, r_i\}$ need not exist in the underlying network, but this condition does not change the essence of the following argument.) The sum of the γ^k's on this subcycle must equal 0, thus $\gamma_{r_i r_l}$ equals the sum of the γ^k's on the path $(r_i, r_j), \ldots, (r_k, r_l)$. Thus, we can replace the path $(r_i, r_j), \ldots, (r_k, r_l)$ by the arc (r_i, r_l). Repeating this argument, if necessary, we can construct a 0-intersection cycle ψ that satisfies $\gamma_\zeta^k = \gamma_\psi^k$.

We have shown that $\gamma_\zeta^k = 0$ for all $\zeta \in \Delta$, for all $k \in K$. The above argument also implies $\sum_{(i,j)\in \text{path } \pi} \gamma_{ij}^k =$ constant, say, γ_{pq}^k for any (directed) path π connecting nodes p and q. In particular, suppose that we chose $p = O(k)$ and q arbitrarily in this argument, then the sum of the γ's for all arcs $\{i, j\}$ (with proper signs) belonging to any path connecting $O(k)$ and q is the same. Let $\gamma_{O(k)q}^k$ denote this quantity. Thus, by setting $v_{O(k)}^k = 0$, we can find *unique* multipliers v_i^k satisfying the condition $v_j^k - v_i^k = \gamma_{ij}^k = -\gamma_{ji}^k$. Now, using these multipliers for the flow conservation constraints, we obtain

$$\begin{aligned}
&\sum_{\{i,j\}\in A} \{(v_j^k - v_i^k) f_{ij}^k + (v_i^k - v_j^k) f_{ji}^k\} \\
&= \sum_{\{i,j\}\in A} (\gamma_{ij}^k f_{ij}^k + \gamma_{ji}^k f_{ji}^k) \\
&= (\gamma_{O(k)D(k)}^k)\, d_k \\
&= v_{D(k)}^k\, d_k.
\end{aligned}$$

We can now show that $\alpha_{ij} = \alpha$ and $\beta_{ij} = \beta$ for all $\{i, j\} \in \{S, T\}$. Choose $\{p, q\} \in \{S, T\}$, so that $\{p, q\} \neq \{u, v\}$, and $k_1 \in K$. Let $P(u, p)$ be a path from nodes u to p fully contained in S, and $P(q, v)$ be a path from nodes q to v fully contained in T. Define $(\mathbf{x}^3, \mathbf{y}^3, \mathbf{f}^3)$ with $\lceil D_{S,T}/C \rceil - 1$ HC facilities and $r - 1$ LC facilities on arc $\{u, v\}$, one LC facility on arcs $\{p, q\}$ and $\{i, j\} \in P(u, p)$ or $P(q, v)$, and $y_{ij}^3 = y_{ij}^0$ for all $\{i, j\} \neq \{u, v\}$. Define the flows as follows.

$(f_{ij}^k)^3 = (f_{ij}^k)^0 + 1$ if $(i, j) \in P(u, p)$ or $P(q, v)$ and $k = k_1$

$(f_{uv}^k)^3 = (f_{uv}^k)^0 - 1$, $(f_{pq}^k)^3 = 1$ for $k = k_1$

$(f_{ij}^k)^3 = (f_{ij}^k)^0$ otherwise.

Using the interchange argument on the solutions $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{f}^0)$ and $(\mathbf{x}^3, \mathbf{y}^3, \mathbf{f}^3)$, we see that

$$\alpha_{uv} + \gamma_{uv}^k = \alpha_{pq} + \sum_{(i,j)\in P(u,p)} \gamma_{ij}^k + \gamma_{pq}^k + \sum_{(i,j)\in P(q,v)} \gamma_{ij}^k \tag{9}$$

for $k = k_1$. Since the sum of the γ's corresponding to any cycle equals zero, (9) implies that $\alpha_{uv} = \alpha_{pq}$. Furthermore, since arc $\{p, q\}$ was chosen arbitrarily, we obtain $\alpha_{ij} = \alpha$ and, thus, $\beta_{ij} = \beta = r\alpha$, for all $\{i, j\} \in \{S, T\}$.

Thus, (8) is equivalent to $\alpha X_{S,T} + r\alpha Y_{S,T}$ + constant $= \alpha^*$, which implies that $\alpha X_{S,T} + r\alpha Y_{S,T} = \alpha_0$. Since (8) is nonvacuous, $\alpha \neq 0$. Consequently, $X_{S,T} + rY_{S,T} = \alpha_0/\alpha = r\lceil D_{S,T}/C \rceil$ because (8) holds as an equality for all points in L.

3.2. The Arc Residual Capacity Inequality

Magnanti, Mirchandani and Vachani (1993) developed the arc residual inequality in their study of a core problem that arises when using a Lagrangian relaxation approach for solving many capacitated network design models. We show that a generalized version of this inequality defines a facet of Conv(**TFLP**). More significantly, we show that if we add all the generalized arc residual capacity inequalities and the upper bound constraints (4) to **TFLP**, then the lower bound that we obtain from its linear programming relaxation is the same as the Lagrangian dual value if we relax constraints (1) of **TFLP**.

Before introducing the generalized arc residual capacity inequality, we extend our model by adding the upper bound constraints (4) to the original **TFLP** formulation, that is,

$$f_{ij}^k + f_{ji}^k \leq d_k \quad \text{for all } \{i, j\} \text{ for all } k.$$

In addition, if $\{i, j\}$ is a bridge arc (i.e., an arc whose removal causes the network to separate into two disjoint components), we add the inequalities

$$f_{ij}^k = 0 \text{ and } f_{ji}^k = 0 \quad \text{for all } k \in K \backslash K(i, j). \tag{10}$$

In this expression, $K(i, j)$ denotes the set of commodities whose origin and destination nodes lie on the "opposite" sides ("shores") of the arc $\{i, j\}$. We can always add inequality (10) to the formulation for the same reason that we can add inequality (4)—that is, because we can assume that the solution for any commodity is cycle free. We refer to the formulation with constraints (1), (2), (3), (4), and (10) as the extended **TFLP** model.

The generalized arc residual capacity inequality, which we will henceforth refer to as simply the *arc residual capacity inequality*, is

$$\sum_{k\in L}(f_{ij}^k+f_{ji}^k)-x_{ij}-r_L y_{ij}\leq(\mu_L-1)(C-r_L)\equiv D_L-\mu_L r_L. \tag{11}$$

In this expression, L is any subset of K, $D_L = \sum_{k\in L} d_k$, $\mu_L = \lceil D_L/C \rceil$ and $r_L = D_L \mod(C)$. Note that if $\sum_{k\in L}(f_{ij}^k + f_{ji}^k) = D_L$ for any subset L and $x_{ij} = 0$, then this inequality forces y_{ij} to be at least μ_L; as we have seen earlier, the linear programming relaxation without this constraint would permit the fractional solution $y_{ij} = D_L/C$. Note further that because of inequality (4), which applies to any problem with nonnegative flow costs, the arc residual capacity inequality (11) reduces to the cutset inequality (6) if $\{i, j\}$ is a bridge arc and $L = K(i, j)$.

To verify the validity of the arc residual capacity inequality for the **TFLP**, we rewrite the inequality as $\sum_{k\in L}(f_{ij}^k + f_{ji}^k) \leq D_L - r_L(\mu_L - y_{ij}) + x_{ij}$. If $y_{ij} \geq \mu_L$, then the inequality is valid because $\sum_{k\in L}(f_{ij}^k + f_{ji}^k) \leq D_L$. If $y_{ij} = \mu_L - s$ for some $s \geq 1$, then the arc residual capacity inequality reduces to $\sum_{k\in L}(f_{ij}^k + f_{ji}^k) \leq D_L - r_L s + x_{ij}$ which is equivalent to or dominated by the capacity constraint $\sum_{k\in K}(f_{ij}^k + f_{ji}^k) \leq C(\mu_L - s) + x_{ij}$.

The next three theorems show, in a theoretical sense, the effectiveness of the arc residual capacity inequality in tightening the linear programming relaxation of **TFLP**.

Theorem 2. *The arc residual capacity inequality* (11) *defines a facet of the extended* **TFLP** *model if and only if*

i. *if* $r_L = C$, *then* $L = K$, *and*
ii. *if* $\{i, j\}$ *is a bridge arc, then* $L = K(i, j)$.

Proof. (**Necessity**) If $r_L = C$ and $L \subset K$, then the arc residual capacity inequality is dominated by the capacity constraint for arc $\{i, j\}$. Now, suppose, that $\{i, j\}$ is a bridge arc and let $G = L \cap K(i, j)$ and $H = L\backslash G$. Also, for simplicity of notation, let $r_{ij} = r_{K(i,j)}$, $D_{ij} = D_{K(i,j)}$ and $\mu_{ij} = \mu_{K(i,j)}$. Since $\sum_{k\in G}(f_{ij}^k + f_{ji}^k) = D_G$ and $\sum_{k\in H}(f_{ij}^k + f_{ji}^k) = 0$, the arc residual capacity inequality is equivalent to $x_{ij} + r_L y_{ij} \geq \mu_L r_L - D_H$. If $L = K(i, j)$, then this inequality becomes $x_{ij} + r_{ij} y_{ij} \geq \mu_{ij} r_{ij}$.

We first show that $\mu_G r_L \geq \mu_L r_L - D_H$. Since $\mu_G = (D_G + C - r_G)/C$ and $\mu_L = (D_G + D_H + C - r_L)/C$, we can write $\mu_G r_L - \mu_L r_L + D_H$ as $r_L(r_L - r_G)/C + D_H(1 - r_L/C)$, which is nonnegative if $r_L \geq r_G$. If $r_L < r_G$, then $r_L < \min(r_G, r_H) \leq D_H$ and, therefore, $r_L(r_L - r_G)/C + D_H(1 - r_L/C) \geq D_H(r_L - r_G)/C + D_H(1 - r_L/C) \geq 0$. Thus, $\mu_G r_L \geq \mu_L r_L - D_H$.

Case i. $r_L \leq r_{ij}$

In this case, we show that the arc residual capacity inequality $x_{ij} + r_{ij} y_{ij} \geq \mu_{ij} r_{ij}$ for $L = K(i, j)$ dominates the arc residual capacity inequality for the given choice of L. Since $\mu_{ij} \geq \mu_G$, the arc residual capacity inequality for $L = K(i, j)$ is stronger than $x_{ij} + r_{ij} y_{ij} \geq \mu_G r_{ij}$. The last inequality dominates $x_{ij} + r_L y_{ij} \geq \mu_G r_L$ if $r_L \leq r_{ij}$ and $y_{ij} \leq \mu_G$ (if $y_{ij} > \mu_G$, the inequality $x_{ij} \geq 0$ implies that $x_{ij} + r_L y_{ij} \geq \mu_G r_L$). Since $\mu_G r_L \geq \mu_L r_L - D_H$, the necessity of condition 2 follows if $r_L \leq r_{ij}$.

Case ii. $r_L > r_{ij}$

If r_L is greater than r_{ij}, then consider the following linear combination of $x_{ij} + r_{ij} y_{ij} \geq \mu_{ij} r_{ij}$ (the arc residual capacity inequality for $L = K(i, j)$) and $x_{ij} + C y_{ij} \geq D_{ij}$ (the aggregate capacity demand inequality across arc $\{i, j\}$):

$$\left(\frac{C-r_L}{C-r_{ij}}\right)(x_{ij}+r_{ij}y_{ij})+\left(\frac{r_L-r_{ij}}{C-r_{ij}}\right)(x_{ij}+Cy_{ij})$$
$$\geq\left(\frac{C-r_L}{C-r_{ij}}\right)\mu_{ij}r_{ij}+\left(\frac{r_L-r_{ij}}{C-r_{ij}}\right)D_{ij}.$$

Simplifying this inequality, we obtain $x_{ij} + r_L y_{ij} \geq r_L \mu_{ij} + r_{ij} - r_L$. The right-hand side of this inequality is greater than $r_L \mu_G$ if $\mu_G < \mu_{ij}$ and so the residual capacity inequality is no stronger than a weighted combination of the other two constraints. So assume that $\mu_G = \mu_{ij}$. This assumption implies that $\mu_L \geq \mu_{ij}$, thus

$$\begin{aligned}x_{ij}+r_Ly_{ij} &\geq r_L\mu_{ij}+r_L(\mu_L-\mu_{ij})-C(\mu_L-\mu_{ij})\\ &\qquad +r_{ij}-r_L\\ &=r_L\mu_L-C(\mu_L-\mu_{ij})+r_{ij}-r_L\\ &=r_L\mu_L-D_L+D_{ij}\\ &\geq r_L\mu_L-D_L+D_G\\ &=r_L\mu_L-D_H.\end{aligned}$$

Therefore, since the residual capacity constraint for a bridge arc is implied by a weighted combination of two valid inequalities, it cannot be a facet.

(**Sufficiency**). We will use an interchange argument, similar to the one used for Theorem 1, to prove the sufficiency part of the theorem. As earlier, define L to be the set of points that belong to Conv(**TFLP**) and satisfy (11) as an equality. Let (8) be an arbitrary inequality that is satisfied as an equality by all points belonging to L. First, construct a feasible solution $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{f}^0)$ that belongs to L.

Consider a (nonbridge) arc $\{u, v\}$. For each $k \in K\backslash L$, install $\lceil d_k/C \rceil$ HC facilities on a path connecting $O(k)$ and $D(k)$ that does not contain arc $\{u, v\}$ and set $f_{ij}^k = d_k$ for all arcs lying on this path. For each $k \in L$, consider a path connecting $O(k)$ and $D(k)$ that contains arc $\{u, v\}$ and install $\lceil d_k/C \rceil$ HC facilities on all arcs of this path except arc $\{u, v\}$. Now send d_k units of flow on arc $\{u, v\}$ for all $k \in L$ by installing $\lceil \sum_{k\in L} d_k/C \rceil$ HC facilities on this arc.

Arguments similar to those used to prove Theorem 1 permit us to show that

1. $\alpha_{ij} = \beta_{ij} = 0$ for all $\{i, j\} \neq \{u, v\}$,
2. $r_L \alpha_{uv} = \beta_{uv}$,

3. $\sum_{(i,j)\in\zeta}\gamma_{ij}^k = 0$

$$\begin{cases}\text{for all cycles } \zeta, \text{ for all } k\in K\backslash L\\ \text{for all cycles } \zeta \text{ for which neither } (u,v)\\ \quad \text{nor } (v,u)\in\zeta, \text{for all } k\in L,\end{cases}$$

4. $\alpha_{uv} + \sum_{(i,j)\in\zeta}\gamma_{ij}^k = 0$

for all cycles ζ for which (u,v) or $(v,u)\in\zeta$, for all $k\in L$.

Set $v_{O(k)}^k = 0$ for all k and define:

$$\theta_{ij}^k = \begin{cases}\gamma_{ij}^k & \text{for all } k\in K\backslash L, \text{ for all } (i,j)\\ \gamma_{ij}^k & \text{for all } k\in L, \text{ for all } (i,j)\neq(u,v) \text{ or } (v,u)\\ \gamma_{ij}^k + \alpha_{uv} & \text{for all } k\in L \text{ and for } (i,j)=(u,v) \text{ or } (v,u).\end{cases}$$

We can now find unique multipliers, using θ_{ij}^k as arc lengths, so that $v_j^k - v_i^k = \theta_{ij}^k$ for all (i, j) and for all k. Multiplying the flow conservation constraint for node i, commodity k by multiplier v_i^k, and adding, we obtain

$$\sum_{k\in K}\sum_{\{i,j\}\in K}\{(v_j^k - v_i^k)f_{ij}^k + (v_i^k - v_j^k)f_{ji}^k\} = \sum_{k\in K} v_{D(k)}^k d_k$$

or

$$\sum_{k\in K}\sum_{\{i,j\}\in K}\{\gamma_{ij}^k f_{ij}^k + \gamma_{ij}^k f_{ji}^k\} + \sum_{k\in L}\alpha_{uv}(f_{uv}^k + f_{vu}^k)$$

$$= \text{some constant, say } \Theta.$$

Thus, inequality (8) is equivalent to

$$\Theta - \sum_{k\in L}\alpha_{uv}(f_{uv}^k + f_{vu}^k) + \alpha_{uv}x_{uv} + r_L\alpha_{uv}y_{uv} \geq \delta$$

which proves the theorem.

Theorem 3. *The capacity inequality* (2), *the upper bound inequalities* (4), *the arc residual capacity inequalities* (11), *and the nonnegativity constraints describe the convex hull of the set of feasible solutions to* **P(LAG)**.

Proof. Since the proof of this result is similar to the proof of a more special result given by Magnanti, Mirchandani and Vachani, we do not provide the details.

Let **P(LPR)** denote the linear program obtained by appending all the upper bound constraints (4) and the arc residual capacity inequalities (11) to the linear programming relaxation of **TFLP**. Clearly, the optimal solution to **P(LPR)** provides a lower bound on the cost of the optimal solution to **TFLP**.

Theorem 4. *The lower bound provided by the optimal solution to* **P(LPR)** *is equal to the lower bound obtained from the Lagrangian relaxation approach for solving* **TFLP** *in which we relax constraint* (1).

Proof. Theorem 3 establishes that inequalities (2), (4), and (11) and the nonnegativity constraints describe the convex hull of the set of feasible solutions to **P(LAG)**. Thus, we can replace constraints (3) of **P(LAG)** by (11) and the corresponding nonnegativity constraints and obtain an equivalent Lagrangian subproblem. This new (equivalent) subproblem satisfies the integrality property (Geoffrion) and, hence, it provides a Lagrangian lower bound equal to that obtained from solving **P(LPR)**.

3.3. 3-Partition Inequalities

As discussed in Section 2, for some problem instances neither the Lagrangian approach nor the use of the cutset inequalities is sufficient to provide a good lower bound for the **TFLP**. In this subsection, we identify another class of valid inequalities that can be used in conjunction with the cutset inequalities (6) and the arc residual inequalities (11) to strengthen the linear programming relaxation of **TFLP**.

One way to view the cutset inequality is in terms of network aggregation: We aggregate the network into two "aggregate" nodes S and T and write the inequality as a valid inequality for the resulting 2-node network. Building upon this idea, Magnanti, Mirchandani and Vachani have described an aggregate 3-node (3-partition) inequality for the single facility case. This inequality is useful for describing the convex hull of feasible solutions to the single-facility network loading problem. We describe two ways of generalizing this inequality for the 2-facility case. The 3-partition inequalities are motivated by the following consideration: Suppose that the formulation of the network loading problem consists of the flow conservation constraints, the capacity constraints, and the cutset inequalities. Then the linear programming relaxation of the loading problem on a 3-node network can produce a "half-integral solution" in **y**. For example, if $C = 24$, $d_{12} = d_{13} = d_{23} = 12$, $a_{12} = a_{13} = a_{23} = b_{12} = b_{13} = b_{23}$, then $y_{12} = y_{13} = y_{23} = 1/2$ and $x_{12} = x_{13} = x_{23} = 0$ is a nonintegral optimal solution to the linear programming relaxation of the problem. Notice that the arcs on which the solution **y** is half-integral form a cycle. This phenomenon occurs in larger networks for the same reason: A "half-cycle" satisfies the cutset constraints, but is cheaper than any other integral solution. The inequalities we present next are useful for cutting off such half-integral solutions. We will describe these inequalities for a network with three nodes; however, these results also apply to larger networks with three "aggregate" nodes.

Let 1, 2, and 3 be the three nodes of the network. Let d_{12}, d_{13} and d_{23} denote the demands between nodes 1 and 2, 1 and 3, and 2 and 3, respectively. Furthermore, if i, j and k are distinct elements of $\{1, 2, 3\}$, define $r_{ij} = d_{ij} \bmod(C)$ and $r_i = (d_{ij} + d_{ik}) \bmod(C)$.

Proposition 3. *Let* $\underline{r} = min(r_1, r_2, r_3)$. *Then the following inequality is valid for* $Conv$(**TFLP**) *defined on a 3-node, 3-arc network:*

$$x_{12} + x_{13} + x_{23} + \underline{r}(y_{12} + y_{13} + y_{23}) \geq \left\lceil \frac{\underline{r}\left(\left\lceil \frac{d_{12}+d_{13}}{C} \right\rceil + \left\lceil \frac{d_{12}+d_{23}}{C} \right\rceil + \left\lceil \frac{d_{13}+d_{23}}{C} \right\rceil \right)}{2} \right\rceil. \tag{12}$$

Proof. Since $r \leq r_1$, the cutset inequality with $S = \{1\}$ and $T = \{2, 3\}$ implies

$$x_{12} + x_{13} + \underline{r}(y_{12} + y_{13}) \geq \underline{r} \left\lceil \frac{d_{12} + d_{13}}{C} \right\rceil.$$

We can similarly obtain the corresponding inequalities for nodes 2 and 3. Adding these three inequalities, dividing by 2 and using integrality arguments to round up the right-hand side gives the desired result.

Proposition 4. *Consider the* **TFLP** *on a* 3*-node,* 3*-arc network. Then all feasible solutions satisfy the inequality*

$$2(x_{12} + x_{13} + x_{23}) + (r_{12} + r_{13} + r_{23})(y_{12} + y_{13} + y_{23}) \geq (r_{12} + r_{13} + r_{23})\left(\left\lfloor \frac{d_{12}}{C} \right\rfloor + \left\lfloor \frac{d_{13}}{C} \right\rfloor + \left\lfloor \frac{d_{23}}{C} \right\rfloor + 2 \right) \tag{13}$$

if and only if

1. *None of the remainders r_{12}, r_{13} and r_{23} equal C.*
2. *The remainders satisfy the triangle inequality; that is,*

$r_{12} + r_{13} \geq r_{23}$, $r_{12} + r_{23} \geq r_{13}$ *and* $r_{13} + r_{23} \geq r_{12}$.

3. *If $\max(d_{12}, d_{13}, d_{23}) > C$, then $r_{12} + r_{13} + r_{23} \leq 2C$.*

Proof. **(Necessity)**

1. Suppose that $r_{12} = C$, $r_{13} < C$ and $r_{23} < C$. Then the feasible solution

$$y_{12} = \frac{d_{12}}{C},\ y_{13} = \left\lfloor \frac{d_{13}}{C} \right\rfloor,\ y_{23} = \left\lfloor \frac{d_{23}}{C} \right\rfloor,$$
$$x_{12} = 0,\ x_{13} = r_{13},\ \text{and } x_{23} = r_{23}$$

violates inequality (13).

2. Suppose that $r_{12} + r_{13} < r_{23} < C$. Then the feasible solution

$$y_{12} = \left\lfloor \frac{d_{12}}{C} \right\rfloor,\ y_{13} = \left\lfloor \frac{d_{13}}{C} \right\rfloor,\ y_{23} = \left\lfloor \frac{d_{23}}{C} \right\rfloor,$$
$$x_{12} = r_{12},\ x_{13} = r_{13},\ \text{and } x_{23} = 0$$

violates inequality (13).

3. Suppose that $d_{12} > C$ and $r_{12} + r_{13} + r_{23} > 2C$. Then the feasible solution

$$y_{12} = \left\lfloor \frac{d_{12}}{C} \right\rfloor - 1,\ y_{13} = \left\lfloor \frac{d_{13}}{C} \right\rfloor,\ y_{23} = \left\lfloor \frac{d_{23}}{C} \right\rfloor,$$
$$x_{12} = C + r_{12},\ x_{13} = r_{13} \text{ and } x_{23} = r_{23}$$

violates inequality (13).

(Sufficiency.) Suppose that $y_{12} + y_{13} + y_{23} \geq \lfloor d_{12}/C \rfloor + \lfloor d_{13}/C \rfloor + \lfloor d_{23}/C \rfloor + 2$. Then inequality (13) is clearly satisfied. So assume that $y_{12} + y_{13} + y_{23} = \lfloor d_{12}/C \rfloor + \lfloor d_{13}/C \rfloor + \lfloor d_{23}/C \rfloor + 2 - s$ for some integer s, $1 \leq s \leq \lfloor d_{12}/C \rfloor + \lfloor d_{13}/C \rfloor + \lfloor d_{23}/C \rfloor + 2$. If $s = 1$, we can assume (by symmetry) that $y_{12} \geq \lceil d_{12}/C \rceil$. Since $y_{13} + y_{23} \leq \lfloor d_{13}/C \rfloor + \lfloor d_{23}/C \rfloor$, a cutset argument implies that $x_{13} + x_{23} \geq r_{13} + r_{23}$. Substituting this inequality in inequality (13) and using condition 2 proves the validity of inequality (13).

Next assume that $s \geq 2$. Then the aggregate capacity demand inequality implies that $x_{12} + x_{13} + x_{23} \geq C(s - 2) + r_{12} + r_{13} + r_{23}$. Substituting for the left-hand side of inequality (13) and using condition 3 proves the result. (If $\max(d_{12}, d_{13}, d_{23}) < C$, then $s = 2$, $y_{12} = y_{13} = y_{23} = 0$ and $2(x_{12} + x_{13} + x_{23})$ is at least as large as the right-hand side of $2(r_{12} + r_{13} + r_{23})$; otherwise, we use condition 3.)

When implemented in our computational study along with the cutset inequality, but without the arc residual capacity inequality, these valid inequalities were modestly effective in reducing the integrality gap. With both the cutset and arc residual capacity inequalities included, the effect of adding the 3-partition inequalities on the integrality gap was less pronounced.

4. COMPUTATIONAL STUDY

This section describes the results of a computational study designed to test the effectiveness of the valid inequalities identified in the previous section. We used these inequalities in a cutting plane procedure with two main phases. (For a discussion of this general approach, see Van Roy and Wolsey 1984 and Hoffman and Padberg 1985). We tested the algorithm on a total of 126 problems. Our test problems, although randomly generated, were based upon information provided by GTE Laboratories and are representative of the size, cost, and demand structures arising in real problems.

4.1. Phase I

During phase I, we start with the linear programming relaxation of **TFLP** and, given a fractional solution for the current linear programming formulation, identify a valid inequality that this solution violates and append it to the current formulation. We then solve the enlarged formulation and continue this procedure until we either find an integer solution, which must be an optimal solution to **TFLP**, or cannot identify any inequalities violated by the current fractional solution. We used the USER subroutine of LINDO, on the VAX 6640 and 8820 computers, to automate the generation and addition of the facet inequalities.

The separation problem for the cutset inequalities (i.e., identifying a cutset inequality that is violated by a given fractional solution to the problem) for the single commodity case can be solved as a max flow problem (see Mirchandani). However, solving the separation problem in the multicommodity case is difficult because of the structure of the cutset inequalities and a polynomially bounded algorithm does not seem evident. (In fact, the separation problem might well be NP-hard.) We could not readily identify an efficient method to solve the separation problem for the arc residual capacity inequalities or the 3-partition inequalities. Hence, we used heuristics to identify violated inequalities for all three classes of inequalities. The heuristics are similar and use a partial enumeration scheme. We describe the heuristic used for identifying violated cutset inequalities and since the heuristics for the other two inequalities are similar, we omit their details. The heuristic for the cutset inequality first carries out an exhaustive search of cutsets defined by sets S with small cardinality. It then uses a "growth" strategy, starting from a single node as S and sequentially building S, to identify violated inequalities. In stating this algorithm, we assume that the nodes are numbered, perhaps, in the order in which they are generated. (Refer to subsection 4.3 for details on problem generation.)

Heuristic for Identifying Violated Cutset Inequalities

STEP 1. Compute $X_{S,T} + rY_{S,T} - r\lceil D_{S,T}/C \rceil$ for all S with $|S| = 1$. If this enumeration identifies a violated inequality, select the one with the minimum value for $X_{S,T} + rY_{S,T} - r\lceil D_{S,T}/C \rceil$ and return. Otherwise, repeat the process for all S with $|S| = 2$. If no violated inequality is found, proceed to Step 2.

STEP 2. Compute $X_{S,T} + rY_{S,T} - r\lceil D_{S,T}/C \rceil$ for $S = \{1, 2, 3\}$. If this cutset inequality is violated, return. Otherwise, continue checking sequentially for all $|S| = 3$, choosing S in lexicographical order of increasing node indices. Add the first violated inequality found and return. If this search does not identify a violated inequality, repeat the process for $|S| = 4$ and $|S| = 5$. If no violated inequality is found, proceed to Step 3a.

STEP 3a. Initialize:

$D_i :=$ the total demand originating or terminating at node i;
$Z_i :=$ the total current capacity incident to node i (i.e., $\sum_{j \in N} (x_{ij} + Cy_{ij})$); and
$S := \{i^* : i^* = \operatorname{argmax}_{i \in N} D_i/Z_i\}$.

STEP 3b. If $|S| \leq 5$, go to Step 3c. Otherwise, check if the current fractional solution violates the cutset inequality defined by S. If yes, add this inequality. Return.

STEP 3c. If $|S| = |N| - 6$, (print "violated inequality cannot be identified") stop. Otherwise, add node $j^* := \operatorname{argmin}_{j \in N} d_{i^*j}/(x_{i^*j} + Cy_{i^*j})$ to S. Go to Step 3b.

This heuristic adds one violated cutset inequality per iteration in increasing order of $|S|$. If the heuristic cannot identify a violated cutset inequality, we use similar heuristics to first search for violations of the arc residual capacity inequality and then for violations of the 3-partition inequality. For the arc residual capacity inequality, we check for all violated inequalities with the cardinality of the commodity set (L in expression 11) equal to 1 or 2. Since we found that the linear program solutions to large problems violate many of these inequalities, we added five such violated inequalities per iteration.

4.2. Phase II

In phase II, we used branch-and-bound starting with the fractional solution generated by phase I. Because the version of LINDO that we were using was not capable of solving general integer programs, we implemented this phase of the algorithm on an IBM 4381 computer using MPSX/370 version 2.0.

We used two methods to establish upper bounds for the problem. The first method used a Lagrangian based heuristic for solving the problem (see Vachani). The second method used a *bootstrapping* approach for the more difficult demand topologies (see subsection 4.4.1). The bootstrapping approach started from the solution to the linear programming relaxation by fixing some integer variables using a naive rounding strategy. Next, it carried out branch-and-bound on the remaining (smaller) set of variables to obtain a feasible solution. Phase II of our algorithm subsequently used the better of the two heuristic solutions as an upper bound for pruning the branch-and-bound tree.

4.3. Computational Study Design

We tested the algorithm on three different network sizes: with $|N| = 6$, 10, and 15 nodes. We randomly generated the ordinates and abscissae—uniformly distributed on a unit square—of these nodes. Given these points, we constructed the underlying backbone network. For the 6-node problems, we assumed a fully connected topology. To avoid an explosive growth in the number of variables, we assumed that the 10- and 15-node networks were sparse. For sparse networks, we chose a targeted nodal degree for each node to be equal to 3 or 5 with a probability of 0.3 and 4 with a probability of 0.4. Starting from node 1, we sequentially cycled through the nodes 1, 2, ..., $|N|$. At stage i, we determined node i's closest neighbor (in terms of Euclidean distance) with unsatisfied degree requirements. We added an arc between this pair of nodes with a probability of 0.80 and repeated the process until either i) the topology satisfied node i's degree requirements, or ii) we had considered all the nodes with unsatisfied degree requirements once. In case ii, we identified node i's closest neighbor, say node j, satisfying the property that the current topology did not include arc $\{i, j\}$, and we added this arc. Consequently, this step would cause us to exceed node j's degree requirement if it had already been satisfied.

Next, we determined the LC and HC costs. These costs consist of a fixed cost component, and a variable cost component which is a linear function of the arc length. We generated three different kinds of demand topologies as follows. We assumed that the probability of nonzero demand between any pair of nodes is 0.5 for the 10-node networks and 0.2 for the 15-node networks. For those pairs of nodes with nonzero demand, we chose the value of the demand in one of three different ways: i) uniformly distributed for all pairs of nodes; ii) uniformly distributed with a higher mean between *a* central node and all the other nodes as compared to the demand between pairs of nodes from the remaining set; and iii) uniformly distributed with a higher mean between *two* central nodes and all the other nodes as compared to the demand between pairs of nodes from the remaining set. We refer to these three demand patterns as complete (C), one (O) and two (T), respectively. For each case, we generated two different levels—low and high—of average demand. We identify a problem category by a combination of demand pattern (C, O or T), average demand (H for high and L for low), and the number of nodes (6, 10, or 15).

Table I summarizes the problem categories and data. Further details concerning the exact expressions used in the calculations of the demand and cost data and the test problems are available from the authors.

The total number of constraints (including (1), (2), (6), (11), (12) and (13)) varies from problem instance to problem instance and is about 25 million for category CL10 problems.

4.4. Computational Results

In this subsection, we report our computational results on the 126 test problems. These problems are distributed over 15 problem categories; we tested 6–10 problems in each problem category so that we might determine how the methodology works "on the average." Our results show that our methodology reduces the integrality gap from the one provided by the original linear programming formulation by 65–80% for the 6 and 10-node problems and approximately 55% for the 15-node problems; the average integrality gap after the completion of our cutting plane procedure is 8.13%; and the approach can solve problems with up to 10 nodes to optimality in a reasonable amount of time. Furthermore, the results identify ways to strengthen the linear programming formulation, a priori, and identify network topologies for which the **TFLP** is more difficult to solve. Finally, we compare computationally the Lagrangian and the cutting plane based approaches.

Table I
Problem Data

Problem Category	Average Number of Nodes/Arcs/ Commodities	Number of Problem Instances	Average Number of General Integer/Continuous Variables
CL6	6/15/15	10	30/450
OL6	6/15/15	10	30/450
TL6	6/15/15	10	30/450
CL10	10/22/20	10	44/880
OL10	10/22/22	10	44/968
TL10	10/21/21	10	42/882
CH10	10/23/22	10	46/1,012
OH10	10/21/22	10	42/924
TH10	10/21/22	10	42/924
CL15	15/29/21	6	58/1,218
OL15	15/30/22	6	60/1,320
TL15	15/31/21	6	62/1,302
CH15	15/33/21	6	66/1,386
OH15	15/32/18	6	64/1,152
TH15	15/34/21	6	68/1,428

4.4.1. Aggregate Results

We will use the following acronyms to denote the respective solutions in this discussion:

- LP: Linear programming model (with the flow conservation and capacity constraints).
- LPC: Cutset inequalities + 3-partition inequalities + linear programming model.
- LPR: Arc residual capacity inequalities + linear programming model.
- LPA: All inequalities of Section 3 + linear programming model.
- LLB: Lagrangian lower bound.
- BES: The best integer solution obtained.

We used three performance measures for our analysis:

i. Percentage gap := (BEST − LPA)/BEST,
ii. Percentage improvement := (LPA − LP)/LP, and
iii. Percentage gap reduction := (LPA − LP) ÷ (BEST − LP).

Figure 1 presents these performance measures for the problem categories that we tested.

These results indicate that the inequalities under investigation are effective in reducing the integrality gap, especially for 6- and 10-node problems. The average

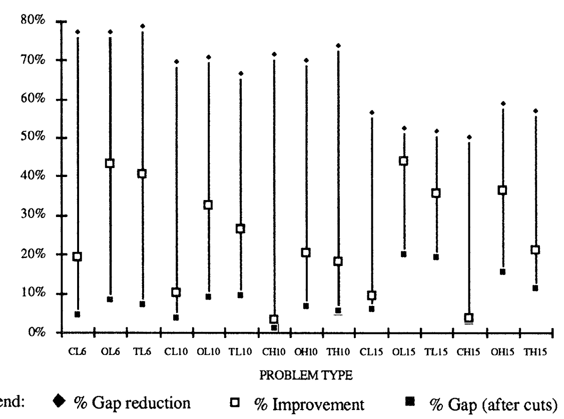

Figure 1. Average performance measures.

percentage gap is high for the O15 and T15 problem categories. We suspect that these larger gaps are attributable to a weak upper bound. However, because we did not run the branch-and-bound algorithm on these problem categories, we cannot substantiate this statement. Furthermore, we observe that the complete demand (C) topologies have the smallest percentage gaps (see Figure 1) and seem to be the easiest to solve, and that the one and two central node (O and T) topologies are more difficult. Moreover, as the demand level increases, the percentage gap becomes smaller.

The percentage improvement, on the other hand, is the lowest for the complete demand topologies (see Figure 1 again). We can explain this apparent anomaly by observing that the linear programming relaxation of the original formulation generates a smaller integrality gap for these problem categories. We also observe that the average reduction in the integrality gap does not seem to depend on the demand pattern; this figure is between 65% and 80% for the nine problem categories with up to 10 nodes (problems for which we found the optimal solution) and approximately 55% for the 15-node problems. (Notice that although we have not reported these data directly, our results show that the gap between the optimal objective values of the linear programming relaxation and the integer programming version of the original problem formulation **TFLP** is as high as 43%.)

In addition to testing the effectiveness of the polyhedral approach for solving the **TFLP**, this computational study was designed to identify possible ways to improve the formulation of the **TFLP** a priori. We collected information on the reduction of the gap after the addition of each cut. Figure 2 presents the cumulative improvement after the addition of each cut and the cumulative time taken up to that stage for a typical problem. From this figure, we observe that the "cumulative percentage improvement" for the cuts exhibit a "tailing effect." However, the improvements do jump on occasion: often when the method identifies a new class of inequality. Notice further that the cumulative time grows slowly in the beginning stages of the algorithm, but the slope tends to increase as the algorithm proceeds and the linear programs become larger. Observe that we achieve about 90% of the improvement in the integrality gap in about 50% of the total solution time.

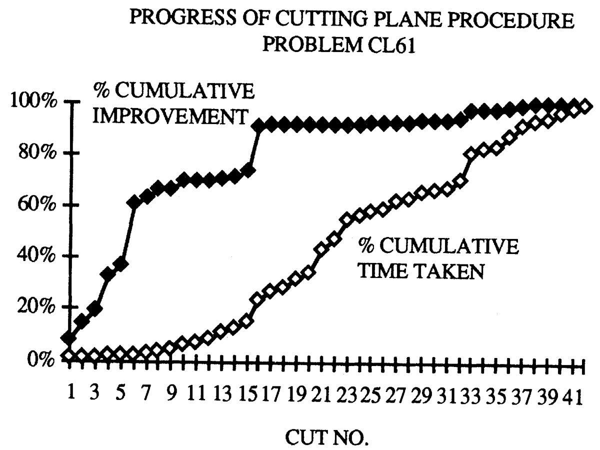

Figure 2. Progress of cutting plane procedure.

These observations lead us to conclude that we might try the effect of randomizing the order in which we select the class of inequality to be considered, and we might terminate the cutting plane procedure after we have added a predetermined number of violated inequalities.

We observed from the timing information that the method spends most of its time solving the linear program: thus, the time for facet-based optimization could be reduced by adding, say, 3 to 5 cutset inequalities simultaneously. This implementation would, however, defeat our objective of checking the progress of the algorithm at each step. This study's developmental nature prompted us to focus less attention on the algorithm's timing and to concentrate more on testing the method's effectiveness in reducing the integrality gap.

We also aggregated the percentage improvement by cutset inequality class when $|S| = 1$ and $|S| = 2$ for the three network sizes. Table II presents this information. We observe that adding, a priori, all cutset inequalities (a polynomial number) for $|S|$ equal to 1 or 2 can be quite effective in strengthening the formulation. As the size of the problem grows, the impact of adding only these inequalities, though still considerable, seems to be less pronounced.

Note that the *total* decrease in the integrality gap due to the addition of these inequalities occurs in two stages: before and after the addition of the arc residual capacity and the 3-partition inequalities. Therefore, the actual improvement in the integrality gap, if we were to include the cutset inequalities of cardinality 1 and 2 in advance, would be slightly lower than that suggested by the last column.

4.4.2. Computational Comparison of the Polyhedral and Lagrangian Methods

This section compares polyhedral methodology with the Lagrangian-based approach. Admittedly, such a comparison would depend on the problem class that we are investigating, the inequalities identified and implemented for the polyhedral approach, and the implementation of the Lagrangian approach. Nonetheless, this comparison could be useful in an algorithm development process; we

Table II
Average Gap Reduction by Inequality Type

Problem Category (No. of Nodes)	Cutset Inequalities (Percent)		
	$\lvert S\rvert = 1$	$\lvert S\rvert = 2$	$\lvert S\rvert = 1$ and $\lvert S\rvert = 2$
6	51.9	21.4	73.3
10	26.8	20.0	46.8
15	30.9	15.6	46.5

believe that this is the first study of its kind in this respect.

In Theorem 4, we proved that the duality gaps for the following two problems are equal: the Lagrangian problem that dualizes the flow conservation constraints, and problem **TFLP** extended by adding all the arc residual capacity inequalities. However, in practice it is difficult to obtain the optimal solutions to both these problems. First, although the literature suggests a number of strategies for using the Lagrangian approach, the most commonly implemented strategy, subgradient optimization, does not necessarily converge to the optimal solution. On the other hand, adding the arc residual capacity inequalities, a priori, would increase the size of the linear program exponentially. Instead, we added only a small subset of these possible inequalities: all those violated inequalities with the cardinality of the commodity set equal to 1 or 2. Thus, for both approaches, we obtained lower bounds to the actual solution values.

Figure 3 compares the average integrality gaps that we obtained using these two approaches. In this figure, LAG refers to the integrality gap that we obtained using the Lagrangian approach. ARC, CUT and ALL refer to the integrality gaps obtained using (a) only the (1 and 2 commodity set) arc residual capacity inequalities, (b) the cutset and the three-partition inequalities, and (c) all the inequalities of Section 3 in the cutting plane procedure. The LAG and the ARC gaps are fairly close to each other (although the LAG gaps are slightly higher for the 15-node problems), suggesting that as the underlying network becomes larger, the polyhedral approach seems to provide better lower bounds.

On the VAX 8820, the Lagrangian approach required approximately 20–40 seconds to solve 6-node problems, 2 to 4 minutes to solve 10-node problems, and 3–6 minutes to solve 15-node problems; this time also includes

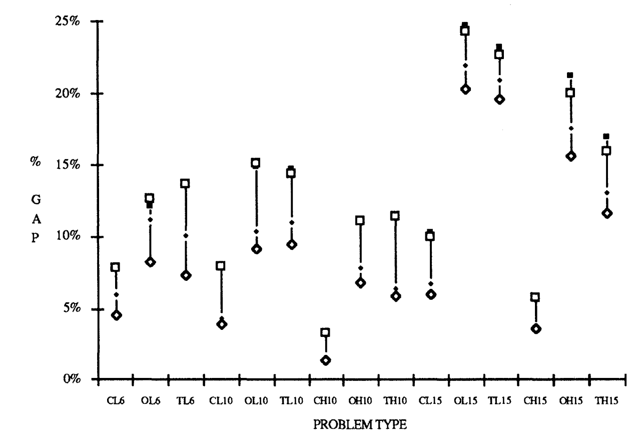

Figure 3. Percentage gap comparison for different approaches. (If the symbols corresponding to LAG and ARC (or CUT and ALL) overlap, we show only □ (or ◇).

the time for determining the heuristic feasible solution. We implemented the polyhedral methodology on the VAX 6440 and the VAX 8820 machines. On these machines, phase I of the procedure required 2–4 seconds to solve 6-node problems, 4–50 seconds to solve 10-node problems, and 14–350 seconds to solve 15-node problems when using only arc residual capacity inequalities. When we included all the inequalities in the cutting plane procedure, the polyhedral procedure required 32–41 seconds for the 6-node problems, 37–601 seconds for the 10-node problems, and 147–12,880 seconds for the 15-node problems. The wide variation (*across* problem categories) in the timings, especially for the 15-node problems, corresponds to the variation in the integrality gaps: Problem categories with smaller gaps ran faster. This information reinforces our earlier conclusion that some problem categories are much easier to solve than others.

The Lagrangian and polyhedral approaches differ in three other important respects:

1. To improve the lower bounds using the Lagrangian relaxation approach, we might have to add new constraints to the original problem formulation. However, doing so can make the relaxed problem much more difficult to solve. Therefore, reducing the integrality gap becomes increasingly more difficult using the Lagrangian approach. On the other hand, the polyhedral approach offers an opportunity for continuous improvement through the identification and implementation of new facets and valid inequalities.
2. Unlike the Lagrangian approach, the polyhedral approach generates monotonically increasing lower bounds at every iteration.
3. As the problem size becomes larger, the size of the linear program to be solved for the polyhedral approach increases rapidly (especially for fully-connected networks) and this approach might become difficult to use in practice. On the other hand, the computational burden of the Lagrangian approach does not increase as rapidly with problem size.

Figure 3 also shows that the cutset inequalities are more effective in reducing the integrality gap than are the arc residual capacity inequalities across all problem categories. When both these inequalities are used together, the arc residual capacity inequalities seem to be more useful for the more difficult (i.e., the O and the T) problem categories.

To conclude this section, we note that the percentage gaps are still high for some problem categories, perhaps because the upper bounds are loose. Nevertheless, a further study of these network topologies might permit us to identify new valid inequalities and to improve the performance of cutting plane methods for these problems.

5. CONCLUSIONS

In this paper, we model and develop solution approaches for a capacitated network design problem that arises in the telecommunications industry. We study two solution approaches to the problem: a Lagrangian approach, and a cutting plane approach. One objective of this research is to compare the two approaches theoretically and computationally. We identified a set of inequalities, the arc residual capacity inequalities, that when appended to the original linear programming formulation guarantee a lower bound equal to the Lagrangian lower bound. In addition to the arc residual inequalities, we also identified two other classes of valid inequalities (the cutset and the 3-partition inequalities) for the underlying polyhedron. Our computational results show that these inequalities are quite effective in reducing the integrality gap for the problem. Using the results of the computational study, we also identified inequalities that might be added to the formulation, a priori, to reduce the integrality gap significantly without an enormous increase in the size of the linear program.

As noted in Section 1, for telecommunications applications, subscribers might have a choice of a third facility, DS3, with capacity equal to 28 DS1 facilities. In general, consider m facilities denoted by HC(1), HC(2), ..., HC(m). Let the capacities of these facilities be $\lambda^1 C$, $\lambda^2 C$, $\lambda^3 C$, ..., $\lambda^m C$ for some set of multipliers $\lambda^i \in Z^1_+$ and $\lambda^1 = 1$; the facilities are indexed so that $\lambda^j > \lambda^i$ if $j > i$. Let y^p_{ij} denote the number of facilities of type p installed on arc $\{i, j\}$. If x_{ij} denotes the number of LC facilities (with capacity 1) installed on arc $\{i, j\}$ and we define aggregate variables across an $\{S, T\}$ cutset as before, then it is possible to show that

$$X_{S,T} + r \sum_{p=1}^{m} \lambda^p Y^p_{S,T} \geq r \left\lceil \frac{D_{S,T}}{C} \right\rceil$$

(where $r = D_{S,T} \bmod(C)$ as earlier) is a valid inequality for the underlying multiple facility polyhedron. In fact, this inequality is facet defining under conditions similar to the conditions of Theorem 1. Thus, while we have discussed our results for the two-facility loading problem, they are applicable in more general settings.

In conclusion, we pose some questions related to this research. First, under what conditions would the proposed inequalities describe the convex hull of the feasible solutions to the capacitated network loading problem? Second, can we identify additional classes of facet inequalities for the problem that might help us in reducing the integrality gap further? Finally, can we extend the formulation for other problem classes so that we obtain a bound that theoretically competes with the bound obtained using Lagrangian relaxation approaches? The answers to these questions might be of value in further understanding the polyhedral structure of the capacitated network design model and other integer programming problems.

REFERENCES

Balakrishnan, A., T. L. Magnanti and P. Mirchandani. 1994. Designing Hierarchical Survivable Networks. Working Paper No. 734, Katz Graduate School of Business, University of Pittsburgh, Penn.

Balakrishnan, A., T. L. Magnanti and R. T. Wong. 1989. A Dual-Ascent Procedure for Large Scale Uncapacitated Network Design. *Opns. Res.* **37,** 716–740.

Balakrishnan, A., T. L. Magnanti and R. T. Wong. 1995a. A Decomposition Algorithm for Expanding Local Access Telecommunications Network Expansion Planning. *Opns. Res.* **43,** 58–76.

Balakrishnan, A., T. L. Magnanti and R. T. Wong. 1995b. Local Access Telecommunication Network Expansion: Modeling and Polyhedral Characterization (in preparation).

Barahona, F., M. Groetschel, M. Juenger and G. Reinelt. 1988. An Application of Combinatorial Optimization to Statistical Physics and Circuit Layout Design. *Opns. Res.* **36,** 493–513.

Barany, I., T. J. Van Roy and L. A. Wolsey. 1984. Uncapacitated Lot-Sizing: The Convex Hull of Solutions. *Math. Prog. Study* **22,** 32–43.

Fisher, M. L. 1981. The Lagrangian Relaxation Method for Solving Integer Programming Problems. *Mgmt. Sci.* **27,** 1–18.

Gavish, B., and A. Altinkemer. 1990. Backbone Network Design Tools With Economic Tradeoffs. *ORSA J. Comput.* **2,** 236–252.

Gavish, B., and I. Neuman. 1989. A System for Routing and Capacity Assignment in Computer Communication Networks. *IEEE Trans. Commun.* **37,** 360–366.

Gerla, M., and L. Kleinrock. 1977. On the Topological Design of Distributed Computer Networks. *IEEE Trans. Commun.* **25,** 48–60.

Geoffrion, A. M. 1974. Lagrangian Relaxation and Its Uses in Integer Programming. *Math. Prog. Study* **2,** 82–114.

Geoffrion, A. M., and R. McBride. 1978. Lagrangian Relaxation Applied to Capacitated Facility Location Problems. *AIIE Trans.* **10,** 40–47.

Graves, S. C. 1982. Using Lagrangian Techniques to Solve Hierarchical Production Planning Problems. *Mgmt. Sci.* **28,** 260–274.

Groetschel, M., M. Juenger and G. Reinelt. 1985a. A Cutting Plane Algorithm for the Linear Ordering Problem. *Opns. Res.* **32,** 1195–1220.

Groetschel, M., M. Juenger and G. Reinelt. 1985b. Facets for the Linear Ordering Problem. *Math. Prog.* **33,** 43–60.

Groetschel, M., L. Lovasz and A. Schrijver. 1981. The Ellipsoid Method and Its Consequences in Combinatorial Optimization. *Combinatorica* **1,** 169–197.

Groetschel, M., C. L. Monma and M. Stoer. 1992. Computational Results With a Cutting Plane Algorithm for Designing Communication Networks With Low-Connectivity Requirements. *Opns. Res.* **40,** 309–330.

Groetschel, M., and M. W. Padberg. 1985. Polyhedral Theory. In *The Traveling Salesman Problem*, E. Lawler et al. (eds.). John Wiley, New York.

Hoffman, K., and M. W. Padberg. 1985. LP-Based Combinatorial Problem Solving. *Anns. Opns. Res.* **4,** 145–195.

KARP, R. M., AND C. H. PAPADIMITRIOU. 1982. On the Linear Characterization of Combinatorial Optimization. *SIAM J. Comput.* **11,** 299–311.

LEBLANC, L., AND R. V. SIMMONS. 1989. Continuous Models for Capacity Design of Large Packet Switched Telecommunications Networks. *ORSA J. Comput.* **1,** 271–286.

LEUNG, J., T. L. MAGNANTI AND V. SINGHAL. 1990. Routing in Point-to-Point Delivery Systems. *Trans. Sci.* **24,** 245–260.

LEUNG, J., T. L. MAGNANTI AND R. VACHANI. 1989. Facets and Algorithms for Capacitated Lot Sizing. *Math. Prog.* **45,** 331–359.

MAGNANTI, T. L., P. MIRCHANDANI AND R. VACHANI. 1993. The Convex Hull of Two Core Capacitated Network Design Problems. *Math. Prog.* **60,** 233–250.

MAGNANTI, T. L., AND R. VACHANI. 1990. A Strong Cutting Plane Algorithm for Production Scheduling With Changeover Costs. *Opns. Res.* **38,** 456–473.

MAGNANTI, T. L., AND R. T. WONG. 1984. Network Design and Transportation Planning. *Trans. Sci.* **18,** 1–55.

MARUYAMA, K., K. FRATTA AND D. T. TANG. 1977. Heuristic Design Algorithm for Computer Communication Networks With Different Classes of Customers. *IBM J. R. D.* 360–369.

MIRCHANDANI, P. 1989. Polyhedral Structure of a Capacitated Network Design Problem With an Application to the Telecommunication Industry. Ph.D. Dissertation, MIT Sloan School of Management, Cambridge, Mass.

NEMHAUSER, G. L., AND L. A. WOLSEY. 1988. *Integer and Combinatorial Optimization.* Wiley-Interscience, New York.

NG, T., AND D. HOANG. 1987. Joint Optimization of Capacity and Flow Assignment in a Packet-Switched Communication Network. *IEEE Trans. Comput.* **35,** 202–209.

PADBERG, M. W., T. J. VAN ROY AND L. A. WOLSEY. 1983. Valid Inequalities for Fixed Charge Problems. *Opns. Res.* **33,** 842–861.

POCHET, Y. 1988. Valid Inequalities and Separation for Capacitated Economic Lot Sizing. *O. R. Letts.* **7,** 105–116.

PULLEYBLANK, W. 1989. Polyhedral Combinatorics. In *Optimization*, G. L. Nemhauser et al. (eds.). North-Holland, Amsterdam.

SCHRIJVER, A. 1986. *Theory of Linear and Integer Programming.* Wiley-Interscience, New York.

VACHANI, R. 1988. Design of Private Telecommunication Networks. TM 0027-07-88-446, GTE Laboratories Incorporated, Waltham, Mass.

VAN ROY, T. J., AND L. A. WOLSEY. 1984. Solving Mixed Integer Programs by Automatic Reformulation. *Opns. Res.* **35,** 45–57.