Near Optimal Algorithm for Service Network Design

Hong-Hsu YEN
Department of Information Management, National Taiwan University
Taipei, 106, Taiwan, R.O.C.

and

Frank Yeong-Sung LIN
Department of Information Management, National Taiwan University
Taipei, 106, Taiwan, R.O.C.

and

Kai LIN
Department of Business Administration, Chung Kuo Institute of Technology
Taipei, 116, Taiwan, R.O.C.

ABSTRACT

Good service network design is important to service providers in modern business environment. It will improve customer satisfaction by quick customer response, and in the same time to minimize the total service center installation cost. In this paper, we consider the service network design problem where the objective is to minimize the total facility cost of service centers and service cost in order to satisfy the service time requirement and the commodity requirement of each user. The total facility cost includes the installation cost for sales and commodity. This service network design problem is the facility location problem, however by including the QoS constraints of service time and commodity flow requirement, this problem is more difficult than traditional facility location problem. We take the approach of mathematical programming in conjunction with optimization-based algorithms to solve the problem. We formulate the problem as a combinatorial optimization problem where the objective function is to minimize the total network deployment cost subject to the aforementioned QoS constraints. The integrality constraints associated with the problem formulation make it difficult to develop efficient and effective solution procedures. Lagrangean relaxation in conjunction with a number of optimization-based heuristics are proposed to solve this problem. From the computational experiments, the error gap between the lower bound and the upper bound are all within 38% in minutes of CPU time for network size up to 350 nodes.

Keywords: Service Network Design, Facility Location Problem, P-median Problem, Optimization, Lagrangean Relaxation, QR/ECR.

1. INTRODUCTION

In service industry, Quick Response/Efficient Consumer Response (QR/ECR) to customer has become a strategic necessity to companies and even to industry. But how to design a sophisticated service network, which consists of the planning of sales force and commodity, with the minimum deployment and operation cost subject to stringent servicing time constraints is a common challenge faced by service network designers and managers. This kind of service network design problem could be classified as the well-know facility location problem.

Location problem could be classified into two categories, uncapacitated location problem and capacitated location problem. When the admission facility locations are finite and known in advance, this is a discrete location problem. Krarup shows that even the discrete uncapacitated location problem is a NP-hard problem [5]. A number of researches have addressed the facility location problem [3, 4, 5, 8, 11, 12]. However, most of these researches only consider the travel distance as the only performance criteria between the service center and the user nodes. Carreras [3] model the P-median problem with the minimum catchment area concept and propose the tabu search algorithm to solve this problem. However, capacity constraint is not considered in [3]. Mazzola [8] model the multiproduct capacitated facility location problem in which the demand for a number of different product families must be supplied from a set of facility sites, and each site offers a choice of facility types exhibiting different capacities. Mazzola propose the Lagrangean relaxation based solution procedures to solve this problem. Hinojosa [4] model the multiperiod two-echelon multicommodity capacitated plant location problem and solve with the

Lagrangean relaxation technique. However, the service time constraint is not modeled in [4, 8].

Facility location problem is so generic that a lot of real world applications could be classified as the facility location problem. Intensive research has been conducted to address the facility location problem with different solution techniques in different application. Matsutomi [7] deal with the location problem of emergency service facility for a dispersed population in public service planning. Solution procedures based on the fuzzy decision problems. However, the response time requirement is not enforced in the constraint such that its application is limited. Weinmann [10] model the circuit design problem as the facility location problem. M. J. Kim [6] develop mathematical models for planning the fixed part of PCS network considering the hard handoff and solving with simulated annealing. Tutschku [9] model the transmitter location problem in the cellular mobile communication systems with the Maximal Coverage Location Problem, which is well-known in modeling and solving facility location problems. Al-Fawzan [2] model the Internet server location problem and present the tabu search algorithm to solve this problem.

In this paper, for the first time, we model the service network problem as the facility location problem where the objective is to minimize the service center installation cost and the service cost in order to meet the service time requirement for the user. The service center installation cost includes the sales force and warehouse installation cost.

This paper is organized as follows. In Section 2, mathematical formulation of the service network design is proposed. In Section 3, the dual approach for the service network design based on the Lagrangean relaxation is presented. In Section 4, the primal heuristics are developed to get the primal feasible solutions from the Lagrangean relaxation problem. In Section 5, the computational results are reported. In Section 6, the concluding remarks are presented.

2. MATHEMATICAL FORMULATION

O	The set of candidate locations for service centers
I	The set of users
A_k	The set of candidate sales capacity configurations for service center at location k
B_k	The set of candidate commodity capacity configurations for service center at location k
R_i	The service time requirement for user I
F_i	The goods demand for user I
d_{ik}	The physical distance between user i and service center k
$j_k(C_k)$	The installation cost for service center with sales capacity C_k
$\Psi_k(G_k)$	The installation cost for service center with goods capacity G_k
$D_{ik}(d_{ik})$	The service cost from service center k to user i, which is a function of d_{ik}
$T_{ik}(d_{ik})$	The service time from service center k to user i, which is a function of d_{ik}

And the decision variables are depicted as follows.

x_{ik}	1 when user $i \in I$ is assigned to service center k and 0 otherwise
C_k	The sales capacity assignment for service center at location k
G_k	The goods capacity assignment for service center at location k

$$\min \quad Z_{IP} = \sum_{k\in O} j_k(C_k) + \sum_{k\in O}\sum_{i\in I} x_{ik} D_{ik}(d_{ik}) + \sum_{k\in O} \Psi_k(G_k) \qquad \text{(IP)}$$

subject to:

$$\sum_{k\in O} x_{ik} T_{ik}(d_{ik}) \le R_i \qquad \forall i \in I \qquad (1.1)$$

$$\sum_{k\in O} x_{ik} = 1 \qquad \forall i \in I \qquad (1.2)$$

$$x_{ik} = 0 \text{ or } 1 \qquad \forall i \in I, k \in O \qquad (1.3)$$

$$\sum_{i\in I} x_{ik} \le C_k \qquad \forall k \in O \qquad (1.4)$$

$$C_k \in A_k \qquad \forall k \in O \qquad (1.5)$$

$$G_k \in B_k \qquad \forall k \in O \qquad (1.6)$$

$$\sum_{i\in I} F_i x_{ik} \le G_k \qquad \forall k \in O\,. \qquad (1.7)$$

The objective function is to minimize the total installation cost of service center and the total servicing cost. The first term and the third term of the objective function are the sales forces and warehouse installation cost for the service centers. Constraint (1.1) enforces the service time constraint for each user. Constraint (1.2) and (1.3) enforce that each user could only be serviced by one service center. Constraint (1.4) is the sales capacity constraint for each service center. Constraint (1.5) specifies the candidate sales capacity set for each service center. Constraint (1.6) specifies the candidate goods capacity set for each service center. Constraint (1.7) is the goods capacity constraint for each service center.

3. LAGRANGEAN RELAXATION

Constraints (1.4) and (1.7) of (IP) were dualized to obtain the (LR).

$$\min \quad Z_D(a,b) = \sum_{k\in O} j_k(C_k) + \sum_{k\in O}\sum_{i\in I} x_{ik} D_{ik}(d_{ik}) + \sum_{k\in O} \Psi_k(G_k)$$

$$+ \sum_{k\in O} a_k \Big(\sum_{i\in I} x_{ik} - C_k\Big) + \sum_{k\in O} b_k \Big(\sum_{i\in I} F_i x_{ik} - G_k\Big) \qquad \text{(LR)}$$

subject to:

$$\sum_{k\in O} x_{ik} T_{ik}(d_{ik}) \le R_i \qquad \forall i \in I \qquad (2.1)$$

$$\sum_{k\in O} x_{ik} = 1 \qquad \forall i \in I \qquad (2.2)$$

$$x_{ik} = 0 \text{ or } 1 \qquad \forall i \in I, k \in O \qquad (2.3)$$

$$C_k \in A_k \qquad \forall k \in O \qquad (2.4)$$

$$G_k \in B_k \qquad \forall k \in O. \qquad (2.5)$$

After proper rearrangement, we can decompose (LR) into three independent subproblems.

Subproblem 1: for x_{ik}:

$$\min \quad \sum_{k\in O}\sum_{i\in I} \left[a_k + D_{ik}(d_{ik}) + b_k F_i\right] x_{ik} \qquad \text{(SUB1)}$$

subject to: (2.1), (2.2) and (2.3).

Subproblem 2 : for C_k :

$$\min \quad \sum_{k\in O} \left[j_k(C_k) - a_k C_k\right] \qquad \text{(SUB2)}$$

subject to: (2.4).

Subproblem 3 : for G_k :

$$\min \quad \sum_{k\in O} \left[\Psi_k(G_k) - b_k G_k\right] \qquad \text{(SUB3)}$$

subject to: (2.5).

(SUB1) could be decomposed into $|I|$ independent subproblems. For each independent subproblem,
Subproblem 1-1: for x_{ik}:

$$\min \quad \sum_{k\in O} \left[a_k + D_{ik}(d_{ik}) + b_k F_i\right] x_{ik} \qquad \text{(SUB1.1)}$$

subject to:

$$\sum_{k\in O} x_{ik} T_{ik}(d_{ik}) \le R_i \qquad (2.6)$$

$$\sum_{k\in O} x_{ik} = 1 \qquad (2.7)$$

$$x_{ik} = 0 \text{ or } 1 \qquad \forall k \in O. \qquad (2.8)$$

(SUB1.1) could be optimally solved by the following algorithm,

(1) Identify the service centers that meet the service time constraints of x_{ik}.
(2) Among the service centers identified in the previous step, identify the lowest cost with respect to $a_k + D_{ik}(d_{ik}) + b_k F_i$, and set the associated x_{ik} to one. Set other x_{ik} to zero.

Both (SUB2) and (SUB3) could be decomposed into $|O|$ independent subproblems. For each independent subproblem, it could be optimally solved by exhaustively search since the candidate capacity configuration for each service center is limited.

From the arguments above that the algorithms developed for each subproblem could all be optimally solved, the weak Lagrangean Duality Theorem could be applied. That is, the lower bound from the dual Lagrangean formulation is a legitimate lower bound to the corresponding original problem [1]. By using the weak Lagrangean duality theorem (for any given set of nonnegative multipliers, the optimal objective function value of the corresponding Lagrangean relaxation problem is a lower bound on the optimal objective function value of the primal problem), $Z_D(a,b)$ is a lower bound on Z_{IP}. We construct the following dual problem to calculate the tightest lower bound and solve the dual problem by using the subgradient method.

$$Z_D = \max Z_D(a,b) \qquad \text{(D)}$$

subject to: $a, b \ge 0$.

Let the vector S be a subgradient of $Z_D(a,b)$ at (a,b). In iteration x of the subgradient optimization procedure, the multiplier vector $m^x=(a^x, b^x)$ is updated by $m^{x+1} = m^x + t^x S^x$, where $S^x(a,b) = \Big(\sum_{i\in I} x_{ik} - C_k, \sum_{i\in I} F_i x_{ik} - G_k\Big)$.

The step size t^x is determined by $d\frac{Z_{IP}{}^h - Z_D(m^x)}{\left\|S^x\right\|^2}$, where $Z_{IP}{}^h$ is an primal objective function value (an upper bound on optimal primal objective function value), and d is a constant ($0 \le d \le 2$).

4. GETTING PRIMAL FEASIBLE SOLUTIONS

The solutions to the dual problem (LR) are used to get the primal feasible solutions. There are three possible ways to get the primal feasible solution from the solution to the (LR). The first is starting from the solutions to the (SUB1). From the servicing assignment x_{ik}, we could determine minimum sales capacity configuration and goods capacity configuration for each service center. If the minimum capacity configuration is within the range of the capacity configurations in the candidate set, then we consider it as a feasible solution, otherwise it is not feasible solutions. The second and the third are starting from the solutions to the (SUB2) and (SUB3) respectively. However, it is not easy to determine the servicing assignment variable x_{ik} from the capacity assignment of (SUB2) and (SUB3).

5. COMPUTATIONAL EXPERIMENTS

The computational experiments for the service network design problem are performed. The algorithms developed in the above sections are coded in C++ and performed on a PC with INTEL™ PIII-800 CPU. The tested network contains the 250 user nodes and 20 potential service center nodes. Since each user node could be potential assigned to each potential service center node, the total number of links is 5000. The goods requirement for each user is randomly generated between two to fifteen. And the locations (x-axis and y-axis) for the user nodes and potential service centers are also randomly generated. The computational time is about one to two minutes in this kind of the network size.

The servicing time function $T_{ik}(d_{ik})$ and servicing cost function $D_{ik}(d_{ik})$ is assumed to be the linear function of Euclidean distance of the link. And the maximum allowable time requirement for each user is assumed to be a constant value, e.g. 1.0. The maximum number of iterations for the algorithms to solve (LR) is 1500, and the improvement counter is 30. The step size for the (LR) is initialized to be 2 and be half of its value when the objective value of the dual algorithm doesn't improve for 30 iterations.

Table 1 Comparison of solution quality obtained by various network sizes

Number of Users	Lower bound	Upper bound	Error Gap(%)	Maximum servicing time	R_i
100	1049.7	1374.7	30.9	0.357	2
150	1539.8	2122.1	37.8	0.539	2
200	2054.3	2617.8	27.4	0.301	2
250	2565.8	3275.9	27.6	0.318	2
300	3183.8	3651.6	14.6	0.436	2
325	3427.4	4048.5	18.1	0.428	2
350	3638.8	*	*	*	2

We perform two sets of computational experiments. In the first set of computational experiments, the choice of the R_i value is fixed (set to 2) so as to examine the solution quality of the service network design problem in different sizes of the network. Table 1 summarizes the results. The first column is the number of user nodes. The second column reports the lower bound of the proposed dual Lagrangean problem. The third column reports the upper bound of the proposed algorithm. The forth column reports the error gap between the lower bound and the upper bound. The fifth column reports the maximum servicing time among all users. The sixth column is the service time requirement (R_i). As can be seen in the forth column, the error gaps between the lower bound and the upper bound are decreasing when the network size is growing. Hence, the algorithms proposed in Section 3 and 4 are even better when the network size is growing. In other words, the proposed algorithms have a good scalability. And the error gaps are reasonably tight when the value of R_i is loose as compared to the maximum servicing time among all users. On the other hand, the * symbol in the last row indicates that the primal feasible solution cannot be found.

Since the value for the maximum allowable servicing time (R_i) for each user have a significant impact on the solution of the service network design problem. In the second set of computational experiments, we try to examine the impact of the R_i value on the solution quality of service network design problem. Table 2 summarizes this result. In Table 2, the number of users is constant (250) and the service time requirement (R_i) is a variable to examine the impact of the R_i value on the solution quality of service network design problem. As could be seen from Table 2, the error gap remains the same under more stringent service time requirements. From Table 2, we could say that we have the stable solution quality under more and more stringent servicing time requirements.

Table 2 Comparison of solution quality obtained by various R_i

Number of Users	Lower bound	Upper bound	Error Gap(%)	Maximum Servicing Time	R_i
250	2565.8	3275.8	27.6	0.31	2
250	2565.8	3275.8	27.6	0.31	1.8
250	2565.8	3275.8	27.6	0.31	1.6
250	2565.8	3275.8	27.6	0.31	1.4
250	2565.8	3275.8	27.6	0.31	1.2
250	2565.8	3275.8	27.6	0.31	1.0
250	2565.8	3275.8	27.6	0.31	0.8
250	2565.6	3245.4	26.4	0.31	0.6
250	2565.8	3241.1	26.3	0.31	0.5
250	2565.9	3240.9	26.3	0.31	0.45
250	2565.9	3225.4	25.7	0.31	0.4
250	2565.9	3263.8	27.2	0.31	0.35
250	2565.9	*	*	*	0.3

6. CONCLUDING REMARKS

In this paper, we considered the problem of service center site selection and sales and goods capacity assignment problem with maximum allowable servicing time and capacity requirements. We formulate this problem as an integer programming problem. The discrete (integer constraints) property makes the problem difficult. We take an optimization-based approach by applying the Lagrangean relaxation technique in the algorithm development.

According to the first set of computational experiments, the error gaps are becoming smaller when the network size is growing. And the error gaps are reasonably tight when the value of R_i is loose as compared to the maximum servicing time among all users. On the other hand, from the second set of computational experiments, the solution quality is the same under more and more stringent servicing time requirements. Hence, the algorithms developed above are typically suitable for solving the large network and stringent time requirements environment for service network design problem.

7. REFERENCES

[1] R. K. Ahuja, T. L. Magnanti and J. B. Orlin, "Network Flows—Theory, Algorithms, and Applications", Prentice Hall, ISBN0-13-027765-7, 1993.

[2] M. A. Al-Fawzan and F. Hoymany, "Placement of network servers in a wide-area network", Computer Networks, 34, pp. 355-361, 2000.

[3] M. Carreras and D. Serra, "On optimal location with threshold requirements", Socio-Economic Planning Sciences, 33, pp. 91-103, 1999.

[4] Y. Hinojosa, J. Puerto and F. R. Fernandez, "A multiperiod two echelon multicommodity capacitated plant location problem", European J. of Operational Research, 123, pp. 271-291, 2000.

[5] J. Krarup and P. M. Pruzan, "The simple plant location problem: Survey and synthesis", Europena J. of Operation Research, 12, pp. 36-81, 1983.

[6] M. J. Kim and J. S. Kim, "The Facility Locations Problem for Minimizing CDMA Hard Handoffs", Proc. of IEEE GLOBECOM, pp. 1611 –1615, Vol.3, 1997.

[7] T. Matsutomi and H. Ishii, "An Emergency Service Facility Location Problem with Fuzzy Objective and Constraint", Proc. of IEEE International Conference on Fuzzy Systems, pp. 315 –322, 1992.

[8] J. B. Mazzola and A. W. Neebe, "Lagrangian-relaxation-based solution procedures for a multiproduct capacitated facility location problem with choice of facility type", European J. of Operation Research, 115, pp. 285-299, 1999.

[9] K. Tutschku, "Demand-based Radio Network Planning of Cellular Mobile Communication Systems", Proc. of IEEE INFOCOM, pp. 1054 –1061, Vol.3, 1998.

[10] U. Weinmann, O. Bringmann and W. Rosenstiel, "Device Selection for System Partitioning", Proc. of EURO-DAC, Design Automation Conference, pp. 2–7, 1995.

[11] R. E. Wendell, A. P. Hurter, Jr. and T. J. Lowe, "Efficient Points in Location Problems", AIIE Trans. Vol. 9, pp. 314-320, 1977.

[12] G. O. Wesolowsky and R. F. Love, "Rectangular Distance Location under the Minimax Optimality Criterion", Transportation Science, Vol. 6, No. 2, pp. 103-113, 1972.