Serdica J. Computing **5** (2011), 261–272

Serdica
Journal of Computing

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

AN ELECTROMAGNETISM METAHEURISTIC FOR THE UNCAPACITATED MULTIPLE ALLOCATION HUB LOCATION PROBLEM*

Vladimir Filipović

ABSTRACT. In this article, the results achieved by applying an electromagnetism (EM) inspired metaheuristic to the uncapacitated multiple allocation hub location problem (UMAHLP) are discussed. An appropriate objective function which natively conform with the problem, 1-swap local search and scaling technique conduce to good overall performance. Computational tests demonstrate the reliability of this method, since the EM-inspired metaheuristic reaches all optimal/best known solutions for UMAHLP, except one, in a reasonable time.

1. Introduction. The past four decades have witnessed an explosive growth in the field of network-based facility location modeling. The multitude of applications in practice is a major reason for the great interest in that field.

ACM Computing Classification System (1998): G.1.6.

Key words: Electromagnetism metaheuristic, hub location, combinatorial optimization.

*This research was partially supported by the Serbian Ministry of Education and Science under project 174010.

Computer and telecommunication networks, DHL-like services and postal networks, as well as transport systems can be analyzed as a hub network. All those systems contain a set of facilities (locations) that interact with each other, and with a given distance and transportation cost. Instead of serving every user from its assigned facility with a direct link, the hub network allows transportation via specified hub facilities.

Hubs are facilities that serve as switching points in telecommunications and transportation networks. Hub networks arise where there is traffic demand from many origin nodes to many destination nodes and connecting all origin-destination pairs by direct links is not practical and/or economical. It is often assumed that hubs are connected by a complete network and the routing cost between hubs is discounted. Flows from many origins to many destinations are consolidated at hubs and routed together to benefit from economies of scale. This is the case when traffic between origin and destination node must be routed via one or more hubs, i.e., direct communication between two non-hub nodes is forbidden. By using switching points in the network and increasing transportation between them the capacity of the network can also be used more efficiently. This strategy also provides lower transportation cost per unit.

There are various model formulations proposed for the problem of choosing a subset of hubs in the given network. They involve capacity restrictions on the hubs, fixed cost, a predetermined number of hubs and other aspects. Two allocation schemes in the network can be assumed: single allocation and multiple allocation.

In the single allocation hub location problem each node must be assigned to exactly one hub node so that all the transport from (to) each node goes only through its hub. The multiple allocation scheme allows each facility to communicate with more than one hub node. If the number of switching centers is fixed to p, we are dealing with p-hub problems. Capacitated versions of hub problems also exist in the literature, but the nature of capacities is different. The flows between hubs or between hubs and non-hubs can be limited. There are also variants of capacitated hub problems that involve limits on the flow into the hub node, through the node or fixed costs on hubs.

In this paper the uncapacitated multiple allocation hub location problem (UMAHLP) is considered. In this case:

- No capacities on the nodes are imposed, so overall communication in each hub and nonhub node is unbounded;

- The number of hubs is not fixed;

- Each non-hub node may be assigned to more than one hub (multiple allocation scheme), i.e., communication from one non-hub node to others can be via different hubs;
- An hub is located with certain expenses (fixed costs) for establishing it. Fixed costs must be paid for every hub and they participate in the objective function.

The objective of UMAHLP is to choose a set of hubs and allocate non-hub nodes to the hubs, so that the sum of total transportation cost and fixed costs is minimized. UMAHLP is known to be NP-hard in the general case. In a special case when the matrix of flows is sparse, the problem is solvable in polynomial time, but this situation is almost non-existent in practical applications.

2. Previous work. A detailed review of all hub location problems and their classification is out of this paper's scope but it can be found in [1, 5]. In the sequel only previous work related only to the UMAHLP will be described.

The first formulation of this problem is given in [4]. Dual ascent techniques within a Branch-and-Bound scheme on small instances of up to 25 nodes are given in [11]. A similar approach is used in [14], with tighter lower bounds and improved upper bounds. The results are presented on instances with up to 40 nodes.

A quadratic integer formulation of the UMAHLP based on the idea of multi-commodity network flows was introduced in [10]. This formulation showed to be suitable for using a Branch-and-Bound procedure. The authors present results on their own randomly generated instances with up to 80 vertices.

A mixed integer linear programming (MILP) formulation for UMAHLP as well as two other similar hub location problems are given in [2]. The experimental results are presented for instances up to 50 nodes. In [13] a polyhedral structure of set packing problem is used to tighten the MILP formulation of UMAHLP. Benders decomposition is also used for solving UMAHLP to optimality [7]. It is able to solve some large instances of up to 200 nodes, considered out of reach of other exact methods in a reasonable time.

A dual-ascent heuristic method embedded into the Branch-and-Bound algorithm is proposed in [6]. This approach was effective on instances of up to 120 nodes, i.e., the dual ascent heuristic reaches up to 70% optimal solutions, which significantly reduces the time for the BnB method to verify optimality in those cases and to resolve the optimal solutions in other 30% cases.

In [12] a genetic algorithm (GA) for solving UMAHLP was proposed that uses binary encoding and genetic operators adapted to the problem. The computational results on standard ORLIB instances with up to 200 nodes are presented. The results show that the GA approach quickly reaches all optimal solutions that are known so far and also gives results on large-scale instances of 200 nodes that were unsolved before. This GA approach is further improved by adding a local search procedure based on 1-interchange ([8, 15]).

3. Mathematical formulation. Various formulations of UMAHLP arise in the literature and one mixed integer linear programming formulation [6] is used in this paper.

Consider a set $I = \{1, \dots, n\}$ of n distinct nodes in the network, where origin/destination or potential hub location is represented by each node. The distance from node i to node j is C_{ij}, and triangle inequality may be assumed [5]. The demand from location i to j is denoted as W_{ij}. Decision variables y_k and x_{ijkm} are used in the formulation as follows:

1. $y_k = 1$ if a hub is located at node k, 0 if not;
2. x_{ijkm} is the fraction of flow W_{ij} from node i that is collected at hub k, and distributed by hub m to node j.

Each path from demand to destination node consists of three components: transfer from an origin to the first hub, transfer between the hubs and distribution from the last hub to the destination location. Parameters χ and δ denote unit costs for collection and distribution, while α is unit cost for hub-to-hub transportation. According to the hub definition, the unit cost for hub-to-hub transportation is less than 1, so the discount factor for transport between hubs, represented by $1 - \alpha$, must be positive. The value f_k denotes the fixed cost of establishing hub k ($y_k = 1$). The objective function is the sum of the total flow cost and the total cost of location hubs. Using the notation mentioned above, the problem can be written as:

$$\text{(1)} \qquad \min \sum_{i,j,k,m} W_{ij} \cdot (\chi \cdot C_{ik} + \alpha \cdot C_{km} + \delta \cdot C_{mj}) \cdot x_{ijkm} + \sum_{k} f_k \cdot y_k$$

subject to

$$\text{(2)} \qquad \sum_{k,m} x_{ijkm} = 1 \quad \text{for every } i, j$$

$$\sum_{m} x_{ijkm} + \sum_{m,m \neq k} x_{ijmk} \leq y_k \quad \text{for every } i,j,k \tag{3}$$

$$y_k \in \{0,1\} \quad \text{for every } k \tag{4}$$

$$x_{ijkm} \geq 0 \quad \text{for every } i,j,k,m \tag{5}$$

The sum of the origin-hub, hub-hub and hub-destination flow costs multiplied by the factors χ, α and δ respectively and the sum of fixed costs for establishing hubs is minimized by the objective function (1). Constraint (2) specifies that all the flow is sent between every pair of nodes, while constraint (3) ensures that flow is only sent via opened hubs. Constraints (4) and (5) reflect the binary and/or non-negative representation of decision variables. The fact that $x_{ijkm} \leq 1$ is implied by constraint (2), and it is omitted.

4. Proposed EM method. Electromagnetism as an optimization heuristic was proposed in [3]. This method is a population-based algorithm that can solve nonlinear optimization problems. Details about the convergence of the method are provided in the cited paper. In the following text each member p_k, $k = 1 \dots m$ of the population maintained by the algorithm will be referred as an EM point (or solution point), and the population itself will be referred as the set of points (or solution set).

In the EM method, a charge is associated to each EM point in the solution set. The charge is calculated as a function of its own and other points, objective functions. Every point has an impact on others through charge, and its exact value is given by Coulombs Law. This means that the power of connection between two points will be proportional to the product of charges and reciprocal to the distance between them. In other words, the points with a higher charge will move other points in their direction more strongly. Beside that, the best EM point will stay unchanged. The proposed EM program for solving UMAHLP is given by the following pseudo-code:

During the initialization part of the algorithm, EM points are created and their location in the solution space is randomly selected from the interval $[0,1]$. Therefore, the EM point p_k represents an n-dimensional vector of real valued coordinates with taking values from the interval $[0,1]$. When the EM points (potential solutions) are created, their location in the solution space is randomly selected. In the context of the UMAHLP problem, coordinate i of the EM point represents whether the corresponding node i is a hub node or non-hub node.

Algorithm 1: EM pseudo code

```
Algorithm EM
data_input()
initialization()
while iteration < max_iteration do
    foreach point p_k in solution_set do
        calculate_objective_value(p_k)
        1_swap_local_search(p_k)
        scale_solution(p_k)
    calculate_charge_and_forces()
    apply_forces()
    if same_solution_unchanged_max_number_of_times() then
        stop()
```

The decision variable y_i in formulas 1 to 5 is a binary variable, so its value is obtained by rounding the value of the corresponding i coordinate of EM point p_k. These real values are mapped to the binary solution vector by using a threshold value, which is set to 0.5 in the proposed algorithm. So, decision variable y_i is set to 1 if $p_{k,i} > 0.5$ and set to 0 if that is not the case.

The quality of the particular EM point is measured by calculating its objective value. Since users can be assigned only to opened hub facilities, only the array y_k directly obtained from coordinates of the EM point is sufficient for successfull calculation of EM point's objective value. In other words, UMAHLP problem sets no limit on capacities, so EM algorithm is designed on the way that calculation of values xijkm is executed within the evaluation of objective function. For a fixed set of hubs (y_k) the modified version of the well-known Floyd-Warsall algorithm, described in [9], is used to obtain shortest paths. After finding shortest paths between all pair of nodes, the evaluation of objective function is relatively straightforward. It is done by summing shortest distances multiplied with flows and corresponding χ, α, and δ parameters, and adding the fixed cost f_k of established hubs ($y_k = 1$).

The pseudo-code of the EM algorithm indicates that, in each iteration, it tries to improve each EM point with a 1-swap local search procedure. The design of the 1-swap local search procedure in the proposed algorithm is as follows: the procedure tries to swap one element of the array of decision variables y_i with its complement value and recalculate the objective function. In other words, during the local search procedure, the node at position i becomes 'promoted' to hub if it was not hub originally, or becomes 'demoted' to non-hub if it was a hub before. After that swap, the objective value of the new EM point is compared to the objective value of the original one. The proposed local search procedure uses first improvement strategy, which means that when improvement is detected, improvement is immediately applied and the local search continues. If for each node the swap produces an objective value greater or equal than the original one, the local search ends with no improvement.

The scaling procedure, which is introduced in the proposed algorithm, has influence on balancing between intensification and diversification of the search process. The scaling procedure is performed after 1-swap local search, and its main goal is to transform the vector in such manner that the intensification of the search is increased. Let $\lambda \in (0,1)$ denotes the scale factor, p_k is the k th point in the solution set and $\bar{p_k}$ is the vector of decision variables y_i, $i = 1, \ldots, n$ after the local search performed on the kth EM point. The position of p_k^{new}, which represents the new (scaled) EM point, is given by the following formula:

$$p_k^{new} = \lambda \cdot \bar{p_k} + (1 - \lambda) \cdot p_k \tag{6}$$

Choosing the appropriate value of scale factor λ is significant for governing the search process. In the extremal case, when λ is close to 1, the search process will likely stick to a local optimum. Another extremal case, when λ is equal to 0, obviously represents a no-scaling situation.

During calculation of charges, potential solutions (EM points) are being evaluated according the following formula:

$$q_i = \exp\left(-n \frac{f(p_i) - f(p^{best})}{\sum\limits_{k=1}^{m} f(p_k) - f(p^{best})} \right), \tag{7}$$

where n is the dimension space, $f(p)$ is the objective function's value for point p in the solution space and p^{best} is the best solution in the solution space after local search and scaling.

Finally, the total force produced by previously calculated charges is applied. The resulting force F_i on point i is the sum of force vectors induced by all other neighbor points on point i:

$$F_i = \sum_{j=1, j \neq i}^{m} F_{ij}, \text{ where}$$

(8)
$$F_{ij} = \begin{cases} \left(\frac{q_i q_j}{||p_j - p_i||^2} \right) \cdot (p_j - p_i), & f(p_j) < f(p_i) \\ \left(\frac{q_i q_j}{||p_j - p_i||^2} \right) \cdot (p_i - p_j), & f(p_j) \geq f(p_i) \end{cases},$$

where $\|p_i - p_j\|$ is the euclidean distance between EM points p_i and p_j. After calculation in (8), the value of F_i is normalized, so it represents only a direction in which a point is going to move. The actual step that is made is the product of a uniform random variable from domain $[0; maxstep]$ and the normalized value $\hat{F}_i$. The EM point with the best objective value is fixed, because all other EM points are moving toward it.

5. Experimental results. All computational results were carried out on an Intel 2.5 GHz single processor with 1GB of memory. The algorithm was coded in the C programming language and tested on AP ORLIB instances from the literature.

The finishing criterion of GA is the maximal number of iterations $N_{iter} = 100$. The scaling factor λ is set to 0.1. Since the results of EM is nondeterministic, the method was applied 20 times on each problem instance.

Table 1 summarizes the EM results on all AP instances and is organized as follows:

- the first three columns contain the instance name, the optimal solution if it is known and the best known solution from the literature if the optimal solution is not known;
- the best solution obtained by EM in 20 runs, named EM_{best}, is given in the fourth column;
- the average running time (t) used to reach the final EM solution for the first time is given in the fifth column, while the sixth and the seventh column (t_{tot} and $iter_{LS}$) show the average total running time and the average number of local search steps for finishing EM, respectively;

Table. 1. Experimental results

Inst.	Opt_{sol} or $Best_{known}$	EM_{best}	t (sec)	t_{tot} (sec)	$iter_{LS}$	gap_{avg} (%)	σ_{avg} (%)
10L	221032.734	opt	0.0031	0.0066	124.7	0.000	0.000
10T	257558.086	opt	0.0010	0.0059	99.5	1.278	1.261
20L	230385.454	opt	0.0031	0.0166	207.2	1.714	2.215
20T	266877.485	opt	0.0052	0.0182	181.2	0.809	1.392
25L	232406.746	opt	0.0073	0.0508	453.1	2.225	2.506
25T	292032.080	opt	0.0117	0.0309	261.3	1.607	1.890
40L	237114.749	opt	0.0292	0.1386	310.9	0.826	1.282
40T	293164.836	opt	0.0229	0.1145	325.1	1.093	2.249
50L	233905.303	opt	0.0791	0.3715	444.3	1.085	0.963
50T	296024.896	298147.281	0.0652	0.2292	323.2	0.948	0.349
60L	225042.310	opt	0.1722	0.4120	180.6	0.000	0.000
60T	243416.450	opt	0.1481	0.3277	114.8	0.358	1.557
70L	229874.500	opt	0.3190	0.7465	269.1	0.746	0.247
70T	249602.845	opt	0.2918	0.4964	130.7	0.532	0.916
80L	225166.922	opt	0.5549	1.1362	261.1	1.578	0.603
80T	268209.406	opt	0.4121	0.9239	130.9	0.000	0.000
90L	226857.465	opt	0.6708	2.3096	333.0	0.767	0.991
90T	277417.972	opt	0.5971	1.1817	124.2	0.000	0.000
100L	235097.228	opt	0.9878	2.5722	334.8	1.852	0.819
100T	305097.949	opt	0.8878	1.5004	128.2	0.000	0.000
110L	218661.965	opt	1.5066	4.2050	402.4	1.072	1.230
110T	223891.822	opt	1.5752	2.6364	176.1	0.479	0.527
120L	222238.922	opt	2.3347	5.2403	332.6	1.033	1.454
120T	229581.755	opt	2.6948	3.9519	162.4	0.000	0.000
130L	223814.109	b.k.	3.3614	9.0636	519.0	0.949	0.690
130T	230865.451	b.k.	3.4129	5.2910	188.9	0.411	0.609
200L	230204.343	b.k.	25.3535	48.4941	595.9	0.694	1.219
200T	268787.633	b.k.	28.9455	43.4441	500.3	0.835	0.571

- in the last two columns (gap_{avg} and σ_{avg}) contain information on the average solution quality: gap_{avg} is a percentage gap defined as $gap_{avg} = \frac{1}{20}\sum_{i=1}^{20} gap_i$, where $gap_i = 100 * \frac{EM_i - EM_{best}}{EM_{best}}$ and EM_i represents the EM solution obtained in the i th run, while σ_{avg} is the standard deviation of gap_i, $i = 1, 2, \ldots, 20$, obtained by the formula $\sigma_{avg} = \sqrt{\frac{1}{20}\sum_{i=1}^{20} (gap_i - gap_{avg})^2}$.

Optimal solutions in Table 1 are marked by *opt* and best known solutions by *b.k*.

The data from Table 1 show that the EM method reached an optimal solution (or best known solution for the last 4 instances) in all cases except one (instance 50T). The fact that EM reached 27 out of 28 optimal/best known solutions with a rather small average gap indicate that the EM approach can be reliably used in solving the UMAHLP.

6. Conclusions. In this article, an electromagnetism-inspired metaheuristic that solves the UMAHLP is introduced. The objective function natively conforms with the problem, while the 1-swap local search and the proposed scaling technique directs EM to promising search regions. Extensive computational experiments indicate that the proposed method is very powerful and that the medium-size and large-size UMAHLP instances can be solved in less than fifty seconds of running time for sizes attaining 200 nodes.

Hence, future work could also concentrate on the speed-up of the algorithm by taking advantage of parallel computation and on GA hybridization with exact methods.

REFERENCES

[1] Alumur S., B. Y. Kara. Network hub location problems: The state of the art. *European Journal of Operational Research*, **190** (2008), 1–21.

[2] Boland N., M. Krishnamoorty, A. T. Ernst, J. Ebery. Preprocessing and cutting for multiple allocation hub location problems. *European Journal of Operational Research*, **155** (2004), 638–653.

[3] Birbil S. I., S. C. Fang. An electromagnetism-like mechanism for global optimization. *Journal of Global Optimization*, **25** (2003), 263–282.

[4] Campbell J. F. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, **72** (1994), 387–405.

[5] Campbell J. F., A. Ernst, M. Krishnamoorthy. Hub location problems. In: Location theory: Applications and theory (Eds H. Hamacher, Z. Drezner), Springer-Verlag, Berlin-Heidelberg, 2002, 373–407.

[6] Cánovas L., S. García, A. Marin. Solving the uncapacitated multiple allocation hub location problem by means of a dual-ascent technique. *European Journal of Operational Research*, **179** (2007), No 3, 990–1007.

[7] De Camargo R. S., G. Miranda, H. P. Luna. Benders decomposition for the uncapacitated multiple allocation hub location problem. *Computers & Operations Research*, **35** (2008), 1047–1064.

[8] Djukić M. A hybrid genetic algorithm for solving the uncapacitated multiple allocation hub location problem. Master thesis, University of Belgrade, Faculty of Mathematics, 2010.

[9] Ernst A. T., M. Krishnamoorty. An exact solution approach based on shortest-paths for p-hub median problem. I*NFORMS Journal of Computing*, **10** (1998), 149–162.

[10] Abdinnour-Helm S., M. A. Venkataramanan. Solution approaches to hub location problems. *Annals of Operations Research*, **78** (1998), 31–50.

[11] Klincewicz J. G. A dual algorithm for the uncapacitated hub location problem. *Location Science*, **4** 1996, No 3, 173–184.

[12] Kratica J., Z. Stanimirović, D. Tošić, V. Filipović. Genetic Algorithm for Solving Uncapacitated Multiple Allocation Hub Location Problem. *Computing and Informatics*, **24** (2005), 415–426.

[13] Marín A., L. Cánovas, M. Landete. New formulations for the uncapacitated multiple allocation hub location problem. *European Journal of Operational Research*, **172** (2006), No 1, 274–292.

[14] Mayer G., B. Wagner. HubLocator: An exact solution method for the multiple allocation hub location problem. *Computers and Operations Research*, **29** (2002), 715–739.

[15] Stanimirović Z., M. Djukic, J. Kratica. A hybrid genetic algorithm for solving the uncapacitated multiple allocation hub location problem. In: Proceedings of SMTDA 2010, Chania, Crete, Greece.

Vladimir Filipović
Faculty of Mathematics
University of Belgrade
Studentski trg 16/IV
11 000 Belgrade, Serbia
e-mail: `vladaf@matf.bg.ac.rs`

Received June 6, 2011
Final Accepted August 11, 2011