Solving hub location problems for networks

Richard Smith
Department of Mathematics, Maricopa College, Phoenix, Arizona, USA
Email: ric2187495@maricopa.edu

Chenwen Zheng
Computer Science, National University of Singapore, Singapore
Email: chenw_zheng@sohu.com, chenw_zheng@hotmail.com

Frederic Launois
Department of Computer Science, Universite Lille 1, 59655 Villeneuve d'Ascq, France
Email:flaunois@univ-lille1.fr

Abstract: - The hub location problem is important in the selection of technological networks, such as computer networks, cellular networks, or wireless sensor networks. These modern communication networks must be dynamically set prior to changes in external conditions, since the nodes deplete their batteries and go out of service. For this reason, it is necessary to update the available data in order to determine which nodes can be used as hubs. The dynamic location problem requires a shorter solution time, even though the quality of the solution may not be ideal. Heuristic methods are used for their simplicity and are easy to package in the firmware. The central aim of this work is to design a heuristic method that will obtain an good feasible solution in a reasonable amount of time. The methodology proposed for the heuristic method consists of obtaining the optimum solution of the relaxed problem followed by rounding this solution to a 0 or 1 value. The strategy developed for rounding the calculations is to first use a measure, called an attractive force, for each node and then to define those nodes more attractive as hubs. Finally, an integer model is solved to assign the nodes to the selected hubs. An interesting result is that the hubs selected by the optimal solution of the relaxed problem are always between the nodes that have a major attractive force. The heuristic algorithm is well established for problems with 10, 20, 25, 50 and 100 nodes. In each case, mixing two levels of difficulty we obtain four problems.

Key-Words: - Hubs, location, networks, integer programming, heuristics.

1 Introduction

Some applications of network designs require an optimal facility location. The problem of designing networks of communications considers the ideal location of communication equipment and the locations of transmission lines. Consider, for example, a network in which the nodes are computers and the edges are physical lines or wireless signal transmission.

For this case, it is necessary to decide how many and which of the computers will be connected to each hub. Another example is the location of antennas for cellular telephones, in which the goal is to optimise signal traffic for the user. In this paper, we discuss a subclass of problems in network design, where the nodes act as consolidation points for flows between them in the network. This type of facility is known as a hub.

Hubs are places where commodities such as air passengers or communications are concentrated. Thus, all of the traffic that is exchanged between nodes must be routed through one o more hubs. In such problems, the hubs are completely interconnected. There are several different variants of hub location problems. The exact number of hubs required the hub and node locations, and the use of simple or multiple connections are a few examples of variables that must be determined. The hubs can have capacity, cost, or location constraints. Campbell [1] provided a comprehensive review of hub location problems.

In this work, we focus on a particular variant of hub location problem, known as the simple hub location problem with node capacity restriction. We will refer to this problem as

Capacitated Single Allocation Hub Location Problem, denoted by CSAHLP.

The choice of the problem is motivated by a digital wireless equipment network design application. Consider a network of cellular antennas with the ability to transmit flows between them. The average traffic between antennas (nodes) is known. The packages of information - signals that route across one or two antennas - are assigned to hubs. For more details see the figure 1.

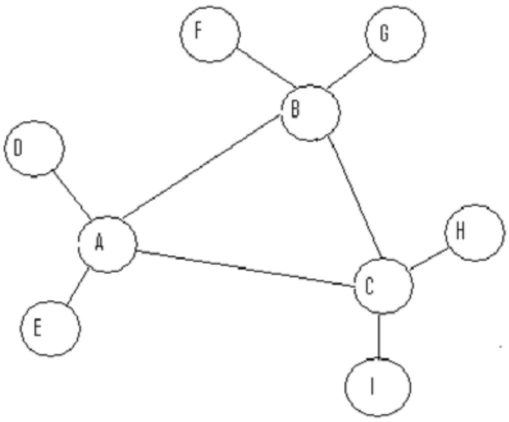


Figure 1

Given the limited concentration capacity of the antennas, it is necessary to introduce restrictions on the node capacity. In addition, the quantity of antennas required is unknown, since this quantity must be kept variable to allow the model to determine the optimum number.

Nevertheless, there is a cost associated with antenna use, and it is a limiting factor. To complete the communication between antennas, the model must decide what node will be use like a hub. Additionally, it must assign the connections of nodes to hubs and account for the cost of transmission in each case. In this work, the cost of transmission from a node origin to a node hub is referred as the compilation cost.

The cost of transmission between hub nodes is called the transmission cost, and the cost of transmission between a hub node and a target node is called the distribution cost. The problem Capacitated Multiple Allocation with p Hubs Multilocation Problem, denoted by CMApHMP is a variant with capacity restrictions and multiple node-to-hub assignments.

CSAHLP has its origin in CSApHMP-type problems, but eliminates restrictions on the number of hubs. The work by O'Kelly [2] and [3] presents a quadratic problem formulation without capacity restrictions called Uncapacitated Single Allocation Hub Location Problem, denoted by USAHLP, which is characterised by the use of only a few variables. Since this formulation is not convex, it is difficult to solve. Campbell [4] has presented a linear form of this problem and added some capacity constraints. This model was analysed by Ernst & Krishnamoorthy [5] and named CSAHLP-C.

There are several recent works on the location problem of simple hub assignments worth mentioning here: Marcus Randall discussed solutions for the CSAHLP problem using meta-heuristic ant colonies. Chen [6] developed a heuristic for CSAHLP and compared it with simulated annealing (SA) to obtaining better results. Costa [7] produced a unique and interesting approach for CSAHLP problems.

This approach does not use capacity constraints on the flow, but instead uses a second objective function to minimise the CPU time. In this paper, we work with the CSAHLP-N model, which corresponds to the model reformulated by Ernst & Krishnamoorthy [8] and based on the work by Campbell [1] [4]. The original model from Campbell CSAHLP-C, is modified by changing the main variable *X*, which shows the percentage of flow from an origin node to a final node, to the variable Y. This new variable represents the amount of the flow from the origin and the route through the hubs.

Some additional equations were modified from the original formulation. The most important advantage of the new CSAHLP-N model with respect to the CSAHLP-C model is the reduction in the problem size and the associated decrease in CPU time. In 2008 Chen [12] developed a heuristic algorithm based in simulated annealing. Computational results indicate that the presented heuristic outperforms a simulated annealing method from the literature.

An Ant Colony System Hybridized with a Genetic Algorithm for the Capacitated Hub Location Problem was presented by Sun et al. [12], they deal with a capacitated asymmetric allocation hub location problem (CAAHLP). As the CAAHLP has impractically demanding for the large-sized problem, a solution method based on combined ant colony optimization algorithm and genetic algorithm was developed which solve hub location problem and node allocation problem respectively.

A review of the state of the art published by Alumur et al. [13] showed an increasing interest in the world for improve the powerful of the algorithms to solve big problems in few second.

In 2010 Contreras et al. [14] presented the Tree of Hubs Location Problem. They propose an integer programming formulation for the problem and present some families of valid inequalities that reinforce the formulation and we give an exact separation procedure for them.

2 Mathematical Formulation

The following is the formulation used in this paper of the p-hub Problem version CSAHLP-N.

2.1. Formulation CSAHLP-N

Minimise

subject to:

The variables of the model are:

: flow per unit time from node *i* through hubs *k* y and *l*;

: a decision variable, where a value of 1 indicates the optimal route for nodes *i* and *k* and 0 applies to any other case;

The data of the model are:

: distance between nodes *i* and k; :flow capacity of the node *k*;

: unitary recollection cost;
: unitary transportation cost;
: unitary distribution cost;
: material flow from node *i*;
: material flow to node *i*;
: cost of using node *k* as a hub;
: flow of material from node *i* to node *j*.

Equation (1) is used to constrain the decision variable ; equation (2) is used to select node hubs for each flow. Equation (3) restricts the use of nodes to their capacity limitations. Equation (4) is the difference equation for node *i* on node *k*, where the demand and the supply of the nodes are determined by the location . Equation (5) defines as binary, and Equation (6) defines as a real positive variable that includes 0.

The proposed model has an integer variable called and a real variable called that correspond to the flow from node *i* to hubs *k* and *l*, respectively. indicates that node *i* is connected by node *k*, and indicates that node *i* is a hub. This formulation therefore corresponds to an integer mixed problem (MIP).

2.2. Relaxing the formulation CSAHLP-N

If the variables Z_{ik} are relaxed so that they take real values between zero and one, this model becomes a problem of linear programming (PPL). In this case, the variable Z_{ii} measures the degree to which node *i* has the potential to be a hub. In this paper, it is assumed that this potential is a measure of the attraction of the node to be used. For this reason, it is designated as the "attractive force".

Now, we focus on the relaxed problem CSAHLP-N, letting P_R) be the relaxed problem.

We have:

*P_R) M*inimise

With the following:

(1), (2), (3), (4) and (6) of the original problem expressed in 2.1

2.3.- Formulation CSAHLP-C

Minimise

Subject to

Skorin-Kapov et al. [10] note that constraints (3) is very weak. Hence, to obtain useful lower bounds from the LP relaxations, they replace (3) by the pair of constraints :

The new formulation is denoted as CSAHLP-LP. Unfortunately these constraints make the formulation very large. In practice terms, it means that solving the LP relaxation for problems with more than 20 nodes becomes too slow.

3. Heuristic attraction force algorithm (HAFA)

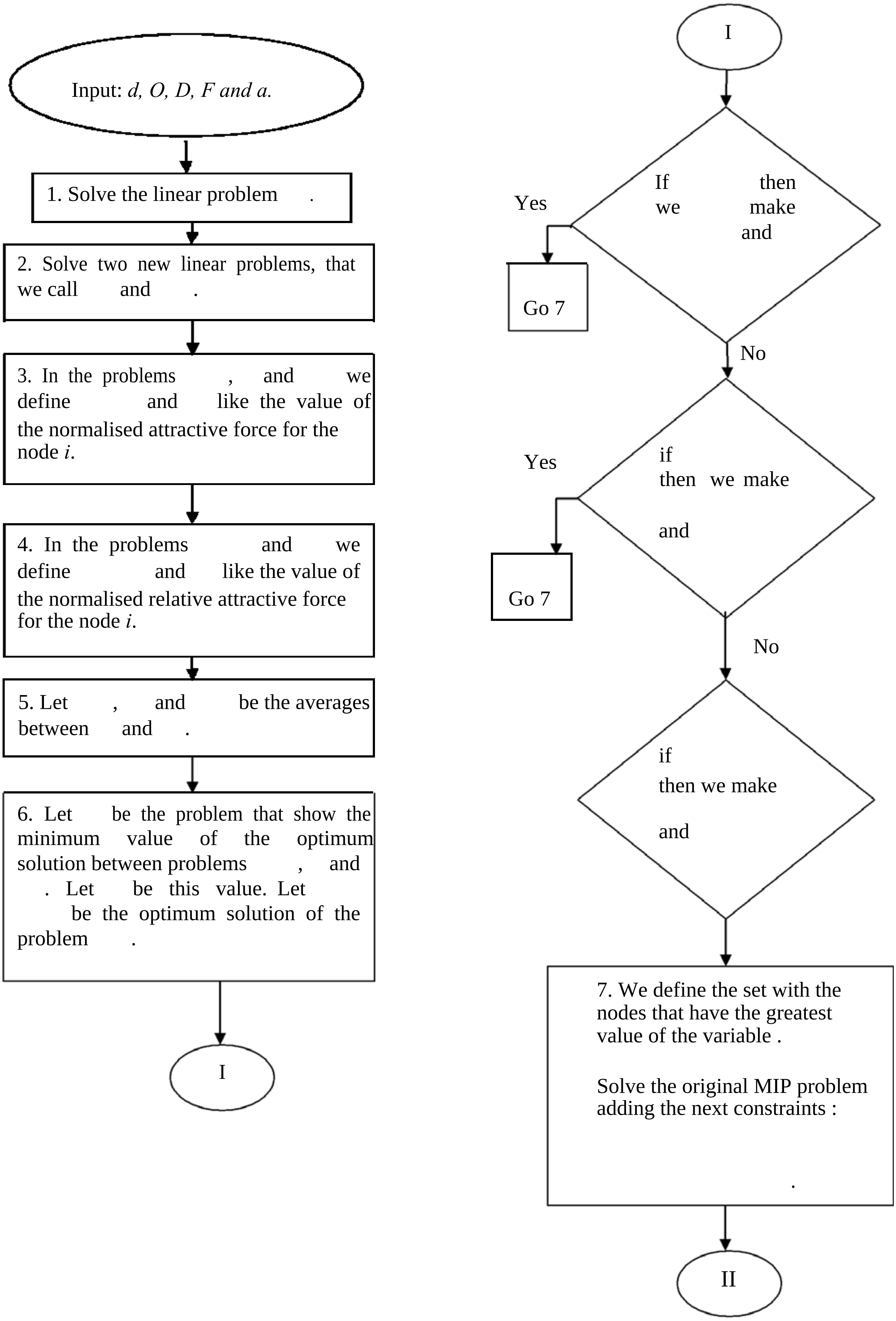
3.1 Attraction force of a node

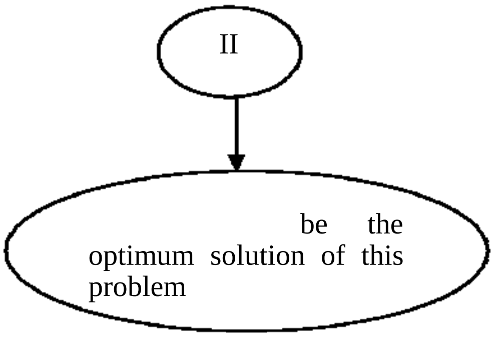
The attraction force of a node is a measure of the proportion of the total flow that would be assigned to this node if it was defined as a hub. This measure is calculated in the different iterations of the heuristic algorithm with results of the relaxed problems. The intuitive idea, it is that the node that receives more flow is a good candidate to be a hub. In addition we used other intuitive idea, it is that a good solution is which minimize of numbers of hubs that attend to all the flows. For consider this fact we divided the first measure by the number of hubs. And then both measures are used for build the attractive force of every node.

The attraction force of a node ( ), is the average between the pure attraction force and relative attraction force. The pure attraction force is the sum of all proportions of the flow that would be assigned to this node using of the optimum solution of the relaxed problem. The normalised attraction force is the pure attraction force divided by the numbers of hubs used. The flow chart and pseudo code of the heuristic is shown below. The inputs are the classic input of the hub and spoke problem. The outputs correspond to the optimum solution.

3.2. Heuristic Algorithm

3.2.1. Flow Chart of the HAFA


Input: *d, O, D, F and a.*
1. Solve the linear problem .
2. Solve two new linear problems, that we call and .
3. In the problems , and we define and like the value of the normalised attractive force for the node *i*.
4. In the problems and we define and like the value of the normalised relative attractive force for the node *i*.
5. Let , and be the averages between and .
6. Let be the problem that show the minimum value of the optimum solution between problems , and . Let be this value. Let be the optimum solution of the problem .
I
I
If then we make and
Yes
Go 7
No
if then we make and
Yes
Go 7
No
if then we make and
7. We define the set with the nodes that have the greatest value of the variable .
Solve the original MIP problem adding the next constraints :
.
II



3.2.2. Heuristic algorithm

HAFA ROUTINE
INPUT: *W, d, F, α, χ, δ, Γ*
OUTPUT:

1. Solve the linear problem . Let $Z^0 \in \Re^{n \times n}$ be the Z matrix of the optimum solution of this problem and let N^0 be the number of hubs a priori, which we will improve in the heuristic algorithm.

   Let be the matrix obtained of the optimum solution of the problem.

2. Then we solve two new linear problems, that we call and respectively. The problem is formed with the problem adding the constraint

   Doing this we force that the model assigns nodes.

   The problem is formed with the problem adding the constraint

   And now we force the model in order that it assigns nodes.

   Let and be the matrix obtained of the optimum solution of the problems and respectively.

3. In the problems , and we define and like the value of the normalised attractive force for the node *i*.

4. In the problems and we define and like the value of the normalised relative attractive force for the node *i*.

5. Let , and be the averages between and :

   We will call these forces normalized attractive forces average.

6. Let be the problem that show the minimum value of the optimum solution between problems , and .

Let be this value. Let be the optimum solution of the problem .

If then we make and , if then we make and and if then we make and .

Then we define the set with the nodes that have the greatest value of the variable .

Solve the original MIP problem adding the next constraints :

Let be the optimum solution of this problem.

END SUBROUTINE

The complexity of the algorithm is bounded by a term with three components, associated with the three problems solved. First, a linear problem (*LP*) is solved. Second, three linear problems that fix the number of hubs are solved. Finally, three *MIP* problems (*MIP*) with fixed numbers of hubs are solved. Then the number of iterations of the heuristic algorithm proposed is on the order of $4 * PL + 3 * MIP$.

Although mixed integer programs are not polynomial problems, in practical terms, the computational complexity and CPU time is reduced fixing the nodes that will be assigned like hub.

3.3 A bound for the integrality gap

Theorem:

Let be the optimum solution of the integer problem and let be the optimum value of the relaxed problem. Then we define the integrality gap.

The integrality gap of the **CSAHLP-N** is:

Then:

Proof:

In the optimal solution of the relaxed problem the cost of transport is equal to the cost of operation of the hubs. It is:

The left side of the equality is the cost of transport and the right size correspond to the operation cost.

Then the value of the objective function in the optimal solution is:

With

Bounding lower:

Bounding upper

Bounding upper again:

and finally bounding by the N_0 be the number of hubs a priori force we have:

Then :

It is possible that this bound was not a tight bound, but it shows that the gap decreases when grows. And like is related to the attraction force of all nodes, we can establish that the gap decreases when the attraction force of all nodes grows.

Many authors relates the integrality gap with the complexity of solve the integer linear programming problem. Then we can conjecture that the attraction force of all nodes, is a measure of the complexity of solve the integer linear programming problem.

4. Computational experiments

To prove the heuristic algorithm, a data set denoted "AP data set" was used. This data set belongs to the public library in the OR-library posted by Beasley [9]. The authors Ernest & Krishnamoorthy and posted this collection of data. This data include capacity restrictions and costs on nodes. We show the characteristics of every instance of the test problem used.

The name of each instance consists of the number of nodes followed by two characters: the first is the type of cost and the second is the capacity type. The characters LL designates low cost and relaxed capacity, the characters LT designates low costs and tight capacity, the characters TL designates high costs and relaxed capacity while the characters TT represents high cost and tight capacity.

Table 1 :Results of the HAFA for CSAHLP-N

Code Problem	CPU Time hr:min:sec	Objective Function	Hubs Selected	Error %
10LL	00:00:00.3	224250.05	4,5,10	0.00
10LT	00:00:00.3	250992.26	1;4;5;10	0.00
10TL	00:00:00.3	263399.94	4;5;10	0.00
10TT	00:00:00.3	263399.94	4;5;10	0.00
20LL	00:00:00.7	250022.7 (1)	7;14;	6.53
20LT	00:00:01.0	259,755.2	6;10;14 (2)	2.46
20TL	00:00:00.7	271892.7 (1)	7:19	2.82
20TT	00:00:05.7 (3)	296035.4	1;10;19	0.00
25LL	00:00:15.3 (3)	238978.2	8;18	0.00
25LT	00:00:26.6	286200.6	9;14;16;25 (2)	3.56
25TL	00:00:01.5	310317.9	9;23	0.00
25TT	00:00:20.7	352070.1	9;12;14 (2)	1.06
40LL	00:00:20.9	242167.2	14;29 (2)	0.09
40LT	00:06:55.5 (3)	272218.3	14;26;30	0.00
40TL	00:03:36.5	298919.0	14;19;	0.00
40TT	00:05:27.0 (3)	354874.1	14;19;40	0.00
50LL	00:00:35.9	238573.1	15;35; (1)	0.02
50LT	00:00:24.9	273382.8	6;26;32;46	0.18
50TL	00:00:34.1	331511.6	3;27;45	3.92
50TT	00:13:13.4	424125.2	12;25;26	1.60
100LL	00:14:52.3	246714.0	29;64;73	0.00
100LT	00:29:12.0	265,932.7	29;71;95	3.62
100TL	00:16:44.3	362,950.1	44;52	0.00
100TT	00:40:56.1	465,213.4	5;34;86;95	0.00

Notes: (1) The allocation is not optimal
(2) The hub assigned is not optimal
(3) CPU time becomes unfeasible

During the benchmarking, we solved two versions of the problem: CSAHLP-N version and CSAHLP_C version.

First we used the CSAHLP-N version and we solved problems containing up to 100 nodes. The obtained solution and the selected hubs correspond to the optimum solution for the instances from 10LL to 100LL. The instances that are not optimal are marked with an asterisk. In Table 1, we show results for the **CSAHLP-N**.

Then we used the CSAHLP-C version of the problem and we solved problems containing up to 25 nodes. Like we said before, this formulation is very large and solving problems with more than 25 nodes was impossible to solve.

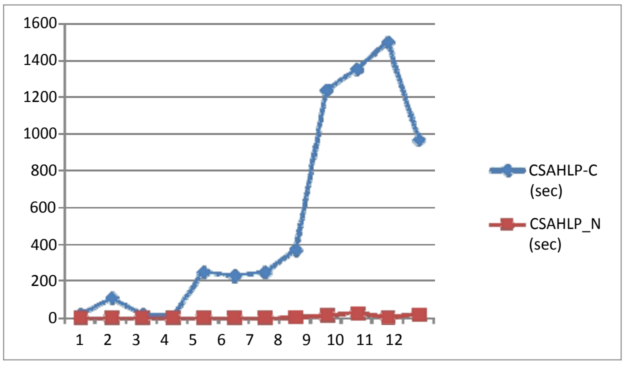
In Table 2, we show results for the **CSAHLP-C**.

Table 2: results of HAFA for CSAHLP-C

Problem name	CPU time hr:min:sec	Objective function	Hubs Selected	Error %
10LL	0:00:20	224250.05	4,5,10	0.00
10LT	00:01:49 *	250992.26	1;4;5;10	0.00
10TL	0:00:21	263399.94	4;5;10	0.00
10TT	0:00:18	263399.94	4;5;10	0.00
20LL	0:04:08	234690.96	7;14;	0.00
20LT	0:03:48	253517.40	10;14;	0.00
20TL	0:04:08	271128.18	7:19	0.00
20TT	0:06:11	296035.40	1;10;19	0.00
25LL	0:20:41	238977.95	8;18	0.00
25LT	0:22:37	279991.15	9;12;25 **	1.31
25TL	0:25:05	316327.51	14;23 **	1.94
25TT	0:16:14	359043.45	8,14,25 **	3.06

* An unfeasible case, ** The selected nodes are not optimal

Graphic N° 1:
Comparing CPU-time of CSAHLP-C vs CSAHLP-N in seconds



Looking at graphic1, we can see the Cpu-time is increasing strongly for the problems with formulation CSAHLP-C, while the problems with formulation CSAHLP-C increasing lowly. This exponentially growing in the Cpu-time in the formulation CSAHLP-C is caused by the explosive growing in the number of variables and in the number of constraints

In this formulation every flow between two nodes uses four indexes in the variable X. For this reason the numbers of variable grows exponentially with the numbers of nodes. So we didn't prove the set problem for 30 o more nodes with the formulation CSAHLP-C.

Conclusions

- This work developed a heuristic algorithm to find a solution for the CSAHLP problem. Two formulations were proved CSAHLP-C y CSAHLP-N. For the CSAHLP-C only three size of nodes were proved: 10, 20 and 25 nodes. For problems with more nodes the Cpu-time was very large. For the CSAHLP-N six size of nodes were proved: 10, 20, 25, 40, 50 and 100 nodes. The Cpu-time found are interesting and the gaps are few in the most of cases.

- The heuristic is quick approaches the optimal solution for most types of problems. This heuristic is very simple because it only needs mathematical operations, meaning it can be packaged in firmware.

- This work introduces a new concept, called the attractive force, to the logistic problem. This is a measure of the likelihood of a node to become a candidate hub. This measure is made using a mix of characteristic nodes that are placed into the linear programming problem.

- There is a strong relation between the quality of the solution and the tight of cost and capacity constraints used. This suggests that there is some equilibrium between the costs of the objective function. The cost used must be equilibrated with the transport costs.

- This kind of heuristic is based on relaxing some restrictions and reformulating the original problem so it is easier to solve than the original. The decision variables of this problem are then transformed into new equations that restrict the solution space for the new problem. Again, we must take the decision variable and transform it to a new restriction for next problem. We can state that it takes a circular form to obtain the result, and in most cases, it is possible to find the optimal solution.

• A comparison of our heuristic algorithm with the meta-heuristic approach shows that our approach sometimes finds the optimal solution, while the optimal solution is never found in the meta-heuristic approach, only a quasi-optimal solution.

• The useful concept "Total Attractive Force" introduced measures the capacity necessary to cover the demand of flow at a minimal cost and was verified. This allows us to use the "Total Attractive Force" as a first approach to find the optimal number of hubs, since the cost function is quadratic and convex.

• This methodology can be extended to other problems with similar characteristics, such as location, set covering, and network problems. In the heuristic showed the resolution of a problem of linear programming is necessary, which is obtained when we relax the integer variables. This is done using the simplex algorithm, which is not very efficient for big problems. One of the future directions to research is reduce the times of resolution of the problem of linear programming, using some heuristic algorithm of more speedy convergence.

• Another area of improvement of the heuristic is the step seven. In this step a problem of linear integer programming is solved. This problem correspond to the original problem in which there have been fixed the variable corresponding to the values of the diagonal of the matrix Z, with ones or zeros.

References:

[1]Campbell, J. F., (1994a). Hub location and the p-hub median problem. Operations Research, 44(6):923–935.

[2]O'Kelly, M.E., (1986a). Activity levels at hub facilities in interacting networks. Geographical Analysis, 18(4):343–356.

[3] O'Kelly, M.E., (1986b). The location of interacting hub facilities. Transportation Science, 20(2):92– 106.

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

[5] Ernst, A.T., Krishnamoorthy, M., (1996). Solution algorithms for the capacitated single allocation hub location problem .Annals of Operations Research 86, 141-159.

[6] Chen, J ., Cao, X, Xiao, Y., and Sun, Y., (2008), Simulated annealing for optimization with wireless sensor and actuator networks Electronics Letters, 44, 20, 1208-1209.

[7]·Costa, M., Captivo, M.. E., Climaco, J., (2008) Capacitated single allocation hub location problem - A bi-criteria approach Computers and Operations research, 35 3671-3695.

[8] Ernst, A.T., Krishnamoorthy, M., (1999). Efficient algorithms for the uncapacitated single allocation p-hub median problem. Location Science 4, 139-154.

[9]Beasley, (1990), OR-Library distributing test problems by electronic mail. Journal of the operational research society, 41, 11, 1069-1072.

[10]Skorin-Kapov, D., Skorin-Kapov,J., O'Kelly, M. (1996). Tight linear programming relaxations of uncapacitated p-hub median problems. European Journal of Operational Research. 94, 3, 582-592.

[11]Chen, J., F., (2008). A Heuristic for the Capacitated Single Allocation Hub Location Problem. Lecture Notes in Electrical Engineering, Advances in Industrial Engineering and Operations Research, ISSN1876-1100 (Print) 1876-1119 (Online) Volume 5, 185-196.

[12]Sun, J.,U., Park, D., An Ant Colony System Hybridized with a Genetic Algorithm for the Capacitated Hub Location Problem, Lecture Notes in Computer Science Volume 7425, 2012, pp 173-181.

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

[14] Contreras, I., Fernández, E., Marín, A., (2010). The Tree of Hubs Location Problem, 202, 390-400.