Two approaches for solving a continuous location problem: Stochastic geometry and Operational Research

Mathieu Trampont, **Christian Destré**
France Telecom division R&D, CORE/M2V/AOC
38-40 rue du Général Leclerc
92794 Issy les Moulineaux Cedex 9 FRANCE

Keywords: multi-source Weber problem, Stochastic Geometry, Large scale hierarchical network

1. Introduction

When a network operator has to decide on the location of network facilities on a territory, real data has to be considered in order to optimize the locations. The amount of data can be really huge: for example thousands of connected subscribers can be considered for a specific area. Such large scale problems require the use of efficient solution methods. However, location problems are generally difficult to solve exactly [i,ii]. Several ways exist to obtain good solutions in a limited time: approximation algorithms, heuristics, etc… But the size of the instances is still an issue and one way to circumvent it is to make some abstraction about real data.

An approach based on Stochastic Geometry [iii,iv] is currently studied at France Telecom R&D. The idea is to define realistic models for telecommunication network architectures capturing the spatial distribution of the network elements and having a limited number of parameters. The models are probabilistic and the parameters can be set in order to fit, on average, the characteristics of real data. The important point is that these models don't depend on the size of the instances. However, they can't be used to solve exactly specific instances.

In this article, we present our first results concerning the comparison between Stochastic Geometry and Operational Research approaches, and discuss the interest of a hybrid of both methods. We focus on a particular Weber problem in a hierarchical network which has been already studied with Stochastic Geometry [v]. We describe how we solve the same problem with O.R. methods. Then we compare the results between the two approaches.

In section 2, we describe with further details the location problem. In section 3, we recall the results of the Stochastic Geometry approach and we present our approach using O.R. Finally, in section 4, the results are presented and the first elements of a hybrid method are discussed.

2. Problem definition

For a geographical zone, the Public Switched Telephone Network (PSTN) connects the subscribers to a Wire Center Station (WCS). Telephone lines pass through several facilities from subscribers to reach the WCS. In order to simplify the model, we suppose that there is only one facility, called Service Area interface (SAI), between a subscriber and the WCS. Thus the model consists of a 3-level hierarchical network. The lower level corresponds to the subscribers, the intermediate level corresponds to the SAIs and the higher level corresponds to the unique WCS. Facilities of each level are located on the plane. We suppose that we know the subscribers and WCS locations. The objective is to decide on the number and the location of the SAIs in order to minimize the cost of serving all the subscribers. Several costs are considered:

- **Installation costs**: we consider for every SAI a constant installation cost, denoted by C.
- **Transportation costs:** the sum of capacity costs corresponding to the link medium, and infrastructure costs corresponding to the civil engineering necessary to establish the link

(e.g. trench digging). Both of them are proportional to the length of the link. We denote by A_1 and B_1 (respectively A_2 and B_2) the capacity and infrastructure costs per meter between subscribers and SAIs (respectively SAIs and WCS).

We denote by V_{y_i} the set of subscribers linked to SAI y_i, N_{y_i} the size of this set, and $\|x\text{-}y\|$ the Euclidean distance between two points x and y. The network architecture and the associated costs are described in Figure 1.

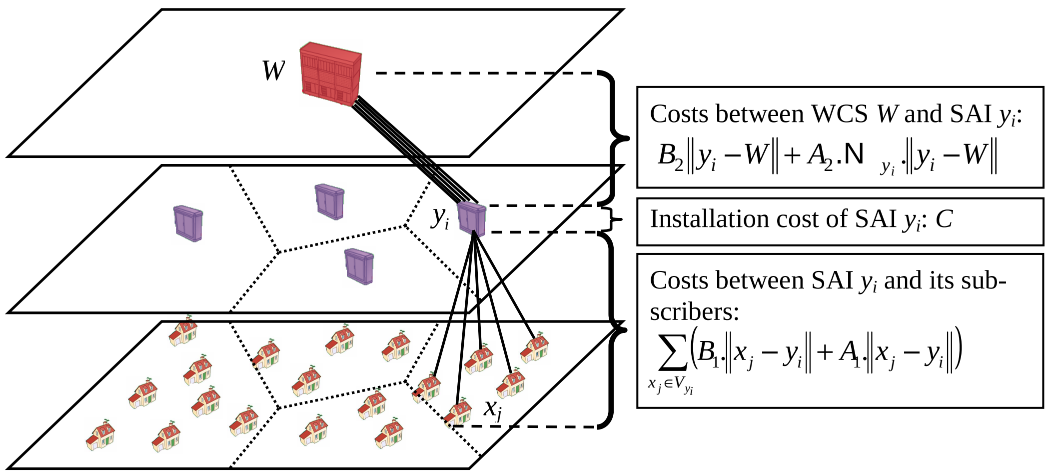

Figure 1: a 3-level hierarchical network and the associated costs

3. Different solution methods

Stochastic Geometry

Stochastic Geometry models the equipment distribution on the plane by using spatial point processes. These processes are defined according to a distribution law and parameters such as the intensity: i.e., the number of facilities per unit square.

Baccelli *et al.* [v] make the *strong* assumption that the different elements of the network are homogeneously distributed on the plane. That corresponds to using homogeneous Poisson point processes to model the different layers of facilities and subscribers. These processes generate points randomly on the plane in such way that their distribution is homogeneous in average. The only parameter for this kind of point processes is the intensity, denoted by a scalar λ. For the WCS layer, the process intensity is given in order to have one WCS in the considered geographical zone. The intensity of the subscribers' process is also known. The objective of Baccelli *et al.* is to minimize the expected cost by finding the optimal parameter of the point process modeling the SAIs. This optimal parameter, denoted λ^*, is obtained by formula. Thus a whole layer of the network is considered only through the law of one point process. The difficulty induced by the huge amount of data is circumvented. But this approach is not accurate to solve exactly specific instances. If we consider a set of subscribers on a plane, we have to compute the corresponding point process intensity. Since the method considers only this parameter, the exact locations of the subscribers aren't considered anymore and we can't solve exactly the instance.

We use Baccelli *et al.*'s results to find λ^* and then solutions are generated according to the law of the SAIs point process. Then the best solution is picked up. This process is done in polynomial time and a large number of potential solutions can be generated.

Operational Research approach

Another way to solve our specific problem is to use Operational Research methods. Our problem is similar to a known continuous location problem: the Multi-Source Weber Problem with Unknown Facility Number. We will further refer to it as the MSWPUFN. This problem is NP-Hard [i]. Since the number m of facilities to locate is part of the problem variables, it is difficult to model the problem with mathematical programming. Cooper proposed to split the objective function in two, with on one side the installation costs and on the other side the transportation costs:

$$F(m) = C.m + g_m * \qquad (1)$$

In this function, g_m* is the optimal value of the following Weber Multi-source problem with m facilities:

$$\begin{aligned} &\text{Min } \mathrm{g_m}(y, Z) = \sum_{i=1}^{\mathrm{m}} \sum_{j=1}^{\mathrm{n}} \left[z_{ij} . \left((\mathrm{A}_1 + \mathrm{B}_1) . \mathrm{d}(y_i, x_j) + \mathrm{A}_2 . \mathrm{d}(y_i, W) \right) \right] \\ &\qquad\qquad + \sum_{i=1}^{\mathrm{m}} \left(\mathrm{B}_2 . \mathrm{d}(y_i, W) \right) \\ &\text{s.c}: \sum_{i=1}^{\mathrm{m}} z_{ij} = 1 \qquad \forall j = 1...\mathrm{n} \\ &\qquad z_{ij} \in \{0;1\} \qquad \forall i = 1...\mathrm{m}, \forall j = 1...\mathrm{n} \\ &\qquad y_i \in \mathrm{IR}^2 \qquad \forall i = 1...\mathrm{m} \end{aligned} \qquad (2)$$

Where n is the number of subscribers, d(y_i,x_j) is the Euclidean distance from point y_i to point x_j, y is the vector of the SAIs, and Z is the matrix of the decision variables z_{ij} which are 1 if and only if the subscriber x_j is linked to the SAI y_i or else 0. The only constraint makes sure each subscriber is connected to one and only one SAI.

Even with a known number of SAIs, the cost function is non-convex and presents a lot of local minima. Thus, the problem complexity and the large size of the considered instances make it impossible to solve exactly in a limited time.

Brimberg *et al.* [vi] propose a multi-phase heuristic for this problem:

- First, to obtain a good initial allocation of all the subscribers (i.e. each subscriber is associated to a SAI) and an estimation of the optimal number of facilities, the discrete version of the MSWPUFN is solved. The discrete problem is an Uncapacitated Facility Location Problem (UFLP) with a set of potential facilities equal to the set of subscribers.
- The initial solution is then improved by Cooper's LOC-ALLOC heuristic [vii] which alternate between location and reallocation phases. Given its subscribers, the location of one SAI is a simple Weber Problem and can be solved easily by a Weiszfeld procedure (see [ii]) which corresponds to a gradient descent method. When the coordinates of every SAI are known, it is easy to verify whether each subscriber is connected to its best SAI (the SAI with the best associated cost is not always the closest one).
- Optionally, local search on the number of facilities to locate can be done. It is possible either to remove or add some SAI in the obtained solution, or to solve the problem anew with fixed values of m close to the value taken in the UFLP solution.

This heuristic can be adapted to solve our specific problem. The difference is that we have a few more terms in our cost function due to our specific transportation costs and to the hierarchical structure of our data. But the discrete analogue to our problem is also a UFLP in which the potential sites for the SAIs are the subscribers plus the WCS. The UFLP is a well known NP-Hard discrete location-allocation problem and several algorithms exist to solve it. But for instances with more than a thousand subscribers the solution often lasts several minutes [viii,ix]. In the implementation of the heuristic, we use the dual ascent procedure of Körkel [viii] to solve exactly the UFLP. The discrete analogue of our problem solved, we repeat successively the continuous location phase and the reallocation phase until stabilization. During the continuous location phase, finding the optimal coordinates of one SAI corresponds to solving a Weber Problem. With (u_i,v_i), (a_j,b_j), and (c,d) denoting respectively the coordinates of the SAI *i*, the subscriber *j* and the WCS, the problem is the following:

$$\text{Min } f(u_i, v_i) = \sum_{j=1}^{N_i} \left[(A_1 + B_1).\left((u_i - a_j)^2 + (v_i - b_j)^2\right)^{\frac{1}{2}} \right] + (N_i.A_2 + B_2).\left((u_i - c)^2 + (v_i - d)^2\right)^{\frac{1}{2}} \tag{3}$$

In these sub-problems some "weights" corresponding to the costs parameters are given to the subscribers and the WCS. The objective is to minimize the weighted distances between the SAI and both its subscribers and the WCS. The solution is found using a Weiszfeld procedure.

This heuristic relies on the similarities between the discrete and continuous problem to find a good local optimum, but nothing is done to explore other local optima.

4. Comparison of results

In order to compare the results, we fix the cost parameters and the intensities for the subscribers and WCS. All these parameters are based on operational data. The solving is performed for each approach as follows:

- For the Stochastic Geometry approach, we compute λ^*, the intensity of the Poisson point process of the SAIs. Then we use this point process to generate 500 solutions.
- For the Operational Research approach, a set of 10 instances (set of subscribers + local WCS) is generated by the Poisson point processes with the fixed parameters above. Then we apply the multi-phase heuristic on each instance.

		OR method			Stochastic Geometry		
Instances	Number of subscribers	Number of SAIs	Solution cost	CPU Time (in seconds)	Number of SAIs for best solution	Best cost	Mean cost
PC_0	1244	71	3412792	318	83	4962003	5248274
PC_1	1318	75	3557940	483	83	5104809	5422906
PC_2	1224	70	3363414	347	83	4893366	5176161
PC_3	1280	72	3479381	498	83	5045818	5325201
PC_4	1245	71	3402685	299	66	4929280	5229313
PC_5	1189	72	3299309	306	79	4834918	5096051
PC_6	1300	73	3455140	332	83	5028733	5340640
PC_7	1239	69	3367153	145	83	4890102	5213842
PC_8	1203	66	3358448	138	83	4875952	5155920
PC_9	1300	85	3544953	634	78	5107145	5386558

Table 1: Computational results

The heuristic gives a specific solution for each of the 10 instances. On the other hand, the 500 generated solutions are valid for *any* instance. For each instance used by O.R., we compute the costs of the 500 generated solutions and we compare them to the solution cost of the heuristic. A good lower bound on the solutions would be useful to assess the optimality gap. But for the time being, our only reference is the cost obtained with the multi-phase heuristic.

In Table 1 we present for each instance: the number of subscribers ; the cost of the best solution generated with the stochastic approach ; the mean cost of all the generated solutions ; the cost of the solution obtained with the multi-phase heuristic ; and the CPU time used by the heuristic (with IBM XSERIES_455 server 3GHz, RAM 10GB).

All solutions obtained by the multi-phase heuristic have a smaller cost than those of the stochastic approach. The average gain of the heuristic over the stochastic approach is about 31 %. However, the mean CPU Time used by the heuristic to solve one instance is 350 seconds, more than four times the 79 seconds used by the stochastic approach to process *all* the instances.

An explanation of the difference in costs lies in the distribution of the SAIs. We can see in Figure 2 that two SAIs can be very close in the best solution generated by Stochastic Geometry, which implies that one of such SAIs is hardly useful. On the contrary, the SAIs in the heuristic solution are more efficiently distributed. This could also explain that the number of SAIs in the best generated solutions exceeds the one in the heuristic solutions from about 11 % on average. But if you consider all the generated solutions, the mean number of SAIs differs only from about 3.5 % between the two approaches.

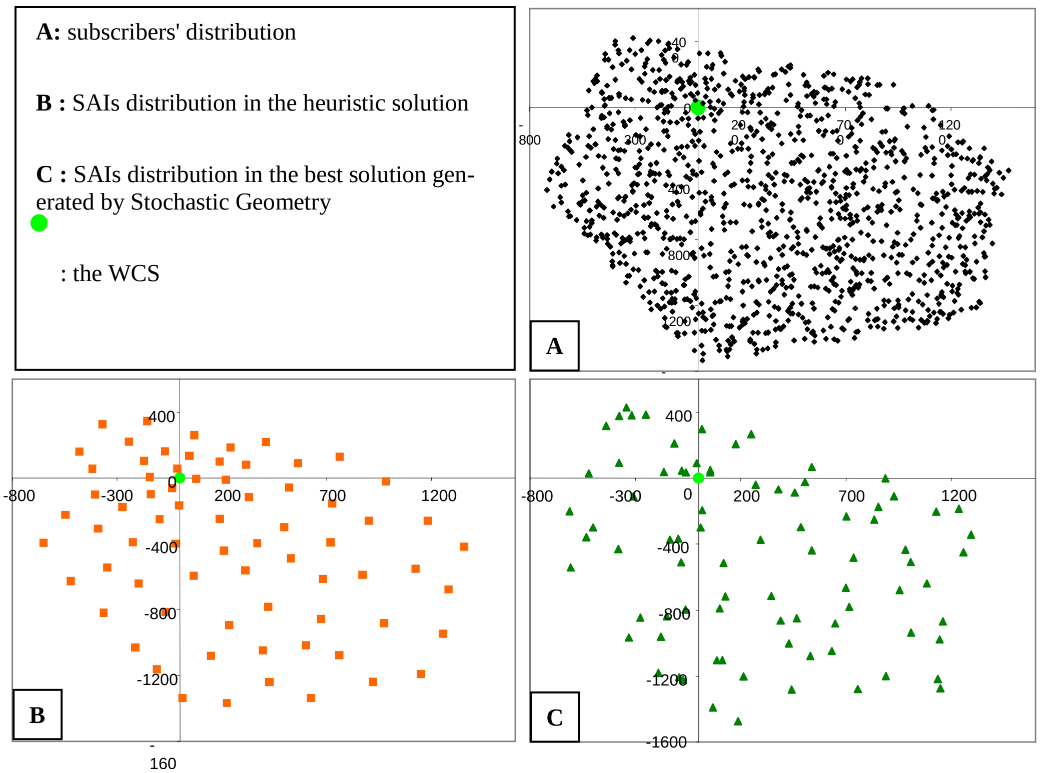

Figure 2: Distribution of the different network elements for the instance PC_0

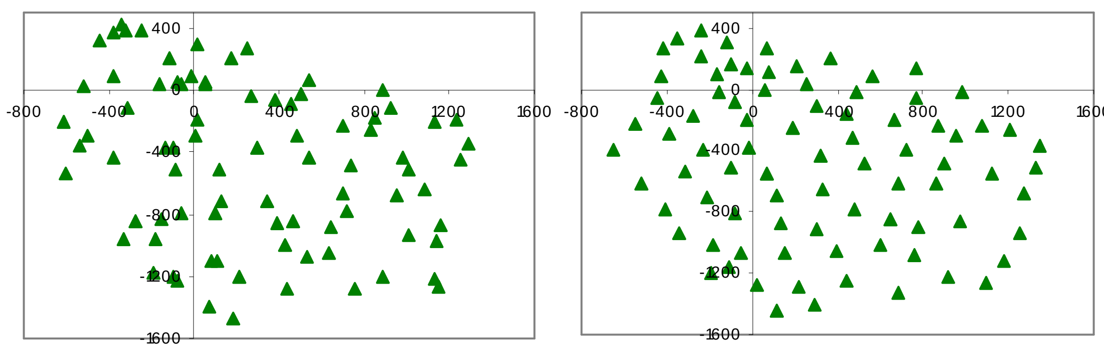

Figure 3 : SAIs distribution in the best generated solution for PC_0 before and after the use of LOC-ALLOC

It is also interesting to note that the inefficiencies in the SAIs distribution for the generated solutions can be partially corrected by LOC-ALLOC, as shown in Figure 3. So a better tradeoff between CPU time used and solution quality could be obtained with a hybrid method using generated solutions as initial solutions of the LOC-ALLOC heuristic.

5. Conclusions and future research

In this paper, we have discussed two approaches to solve a Multi-Source Weber problem for large instances of a 3-level hierarchical network. On the first hand, Stochastic Geometry uses probabilistic models to make some abstraction about real data. Only the density of clients and WCS are taken into account and a set of solutions is generated. This method is fast but, as expected, the quality can vary a lot between solutions. On the other hand, we have adapted a multi-phase heuristic to process all real data to obtain a specific solution. On average, this method is forty times slower than the previous one (with 500 generated solutions) but obtains solutions better by 31 %. But we have no information on how far is the optimum.

We have also seen that the use of the second part of the multi-heuristic can correct some disadvantages of the stochastic approach. We propose a combination of stochastic geometry and LOC-ALLOC, and our first results are encouraging. Unlike the multi-phase heuristic, such a hybrid method can explore different local optima depending on the initial solutions generated by stochastic geometry. The generated solutions also seem to give a good range for the number of SAIs. To strengthen the comparison between the different approaches, we are also trying other O.R. methods such as simulated annealing.

Finally, in this article, subscribers are modeled by a simple homogeneous model. This is the first (easy) step for comparing the two approaches. In reality, the road network influences the subscribers' location. Models taking into account the road network for both Stochastic Geometry and O.R. are currently studied, and developed.

References

[i] N. MEGIDDO and K.J. SUPOWIT, "On the Complexity of Some Common Geometric Location Problems", SIAM Journal on Computing, 13, 182-196, 1984.

[ii] Z. DREZNER and H. HAMACHER (Eds), Facility Location : Applications and Theory, p. 1-36, 2002.

[iii] F. BACELLI, M. KLEIN, M. LEBOURGES and S. ZUYEV, "Stochastic geometry and architecture of communication networks", J. Telecommunications Systems. 7, 209-227, 1997.

[iv] K. TCHOUMATCHENKO, "Modeling of Communication Networks Using Stochastic Geometry", Ph. D. thesis. University of Nice – Sophia Antipolis, 1999.

[v] F. BACCELLI and S. ZUYEV, "Poisson-Voronoi spanning trees with applications to the optimization of communication networs", Operations Research, 47(4), 619-631, 1999.

[vi] J. BRIMBERG, N. MLADENOVIC, S. SALHI, "The multi-source Weber problem with constant opening cost", *Journal of the Operational Research Society*, 55, 640-646, 2004

[vii] L. COOPER, "Location Allocation Problems", *Operations Research*, 11, 331-341, 1963

[viii] M. KÖRKEL, "On the exact solution of large-scale simple plant location problems", European Journal of Operational Research, 39, 157-173, 1989.

[ix] P. HANSEN, J. BRIMBERG, D. UROŠEVIĆ, N. MLADENOVIĆ, "Primal-Dual Variable Neighborhood Search for Bounded Heuristic and Exact Solution of the Simple Plant Location Problem", *Les Cahiers du GERAD*, Octobre 2003.