Gateway placement for sensor networks based on distributed systems

Technical report

Miklós Krész, László Hajdu, Balázs Dávid, Andrej Brodnik

September 2022

1 Background

The objective of this section is to introduce the structure of the distributed system based sensor networks. These networks are organized on two levels. The first one is the level of sensors where the nodes are the sensors measuring the environment. Sensors are connected through which they can send data to each other. The second level of the network is the level of gateways. They are forming a distributed system which means that they should be able to store the data in a distributed way. Data from the sensors are travelling to the gateways, each sensor is arranged to a gateway node through a path determined by the routing. All the collected data are shared among the gateway nodes. Therefore the process of the data collection is the following: 1. the sensors are measuring the environmental properties 2. each sensor is sending the measured data to one gateway 3. the gateway received the data will distribute it over the distributed gateway network.

For the gateway placement the sensor network is given and our goal is to identify the gateway nodes in an optimal way. We suppose that the number k of gateways is given. For the communication cost we can consider an arbitrary positive weight on the graph, but in several cases the unit cost for each link between the nodes (for all sensor-sensor, sensor-gateway and gateway-gateway links) will be assumed. Nevertheless, in the latter cases as well, all the specified problems in the following section can be generalized for non-uniform communication costs as well: e.g. the simplest generalization is using different costs for the links between gateways and for the rest of the sensor network.

2 Problem definition

For the rest of this study we are fixing a weighted directed connected (edge)weighted graph $G = (V, E, W)$ representing the sensor network highlighted in the previous section. (if the network is not connected, then each subproblem defined on the connected components need to be solved separately) The weights can be also dynamic as the communication cost between gateway nodes can be different from the cost between sensor nodes (or sensor-gateway communication). By the above fact the generalization in the case of certain conditions is not straightforward.

2.1 Objectives

Our goal is to identify an optimal placement of the gateways in graph G with respect to some cost function defined on the network. We assume that the number k of gateways is determined in advance. In different versions of the problem, because of the characteristics of distributed systems we need to ensure for all gateways (or certain percentage of gateways) to have all the data measured by sensors. We also assume that the system is operating by synchronous steps by data communication executed through edges connecting adjacent nodes.

In the model we suppose that the routing from each sensor is arranged to a unique gateway. In practice multipath routing is possible [GUO and ZHANG, 2011], but it is considered in a stochastic manner (one routing will be ending at a single gateway, but the choice during the routing can be stochastic). A good review is available in the thesis [de Araujo Marques Leão, 2018] In the following we will distinguish several versions of the problem and in certain cases multipath routing has no influence on the model, while in other cases it could be handled by a stochastic model only with an expected value for each gateway having all data. This stochastic feature would cause problem in ensuring the required level of data sharing among the gateways, hence in this study we will mainly consider deterministic models.

First we suppose that the set S of k gateways is selected. A path p from a sensor node v to a gateway node w is called an *S-routing path* if p contains a gateway node u such that $p[v, u]$ is running outside from S, while $p[u, w]$ is running inside S. In that case $p[v, u]$ is called the *external prefix* of p and u is referred as the *S-entrance* of p. If p is equal to its external prefix, then it is called an *external S-routing path*. An *S-routing system* $\mathcal{R}$ is a set of S-routing paths such that each sensor node v is the endpoint of a path in $\mathcal{R}$ with all paths starting from v having the same external prefix. We will call $\mathcal{R}$ a *weak S-routing system* if each path of $\mathcal{R}$ is external. On the other had, $\mathcal{R}$ is called a *strong S-routing system*, if it has the following property: for each sensor node v and gateway node w if v and w are connected by a path of $\mathcal{R}$, then v is connected by a path of $\mathcal{R}$ with all gateway nodes being in the component (of the subgraph induced by S) containing w. We will define problems for both weak and strong routing systems with appropriate cost functions. In each case the cost $c(S)$ of a gateway placement S is the cost of the weak (respectively, strong) S-routing system with minimum cost.

We are considering two type of problems.
Type 1 problem: A gateway set S is given and determine $c(S)$ and the S-routing system with cost $c(S)$.
Type 2 problem: Determine the feasible gateway set S with minimum cost $c(S)$.

For both problems the cost function is relevant and for Type 2 problems we have to identify the constraints for the feasible solutions. We will discuss these questions in the following subsections. For solving Type 2 problem we should be able to give solution for Type 1 as well. Generally Type 1 problems can be handled with some standard path algorithms in polynomial time, on the other hand Type 2 problems are hard. We will discuss about the basic approach of solution methodologies in later Sections.

2.2 Feasible solutions

Concerning feasible solutions we distinguish 3 cases.

a) For each gateway we expect to know all sensor information. In this case the gateway subnetwork needs to be connected as the subgraph induced by the gateway set: each set of nodes of size k forming a connected induced subgraph is a feasible solution.

b) There is en expected value q as a lower bound for the number of gateways to receive each sensor information. Each component of the subgraph induced by a feasible solution S connected to sensor nodes potentially by an S-routing system needs to have size at least q. In the special (and practical) case of $q > k/2$, only the largest component of S has role in the feasible solution: each sensor node needs to be potentially connected by S-routing to this component and its size needs to be at least q. It is important to note that this special case is not equivalent with *Case (a)*: Indeed, the "dummy" gateway nodes in the smaller components have no role in a feasible solution, however the S-routing paths are not allowed to cross them.

c) There is no lower bound for the number of gateways reached by sensor information, i.e. $q = 0$. In this case each subgraph of G is a feasible solution.

Finally we note that each of the above cases can be defined for both weak and the strong routing systems, by which we distinguish 6 subcases altogether.

2.3 Cost function

The objective function is defined by the combination of two cost functions described below.

(i) The first cost function is defined as gateway maximization. For each component of the feasible solution S, the product of its size and that of the number of sensor nodes connected by routing is considered and the sum of these products over all components are taken. This cost function is considered in *Cases b) and c)* of feasible solutions only (in *Case a)* it does not make sense). Also, at this cost function the weights of the edges have no influence on the solution.

(ii) The goal of the second cost function is the minimization of the distance between the gateway set and the sensor nodes connected by routing. The distance is defined as the graph theoretic distance through the optimal S-routing system. We distinguish two subcases:

[(ii/1)] Minimizing the largest distance from a sensor node to the gateway set.

[(ii/2)] Minimizing the sum of the distances from all the sensor nodes to the gateway set.

These above functions can be used in different ways: a linear combination as a single objective function, bilevel optimization or multiobjective optimization (e.g. as Pareto optimum).

3 State of the art

For the 2.2 Case b) there is no literature, and Case c) is the S-routing without connectivity constraint, thus we are making the review with concentrating on Case a) (gateway subgraph is connected) which is in certain case will refer to literature without connectivity constraints (implicitly Case c)). By the above reasons as it was outlined in Section 2.2, in cost function we will concentrate on Case (ii) only. For the rest of the Section for Case (ii/1) we will use "center problems", while we will refer to Case (ii/2) as "median problems". The reason is that they are the generalizations of the k-center and k-median problems. [Tansel et al., 1983]

First we consider the "weak versions". In this case we are facing the connected k-center and connected k-median problems. Interestingly, these problems started to be studied in recent years only and resolved for special cases only such as for block graphs, cactus graphs and complete multilayer graphs. [Bai et al., 2016], [Bai et al., 2021], [Nguyen et al., 2022] The linear combination of these two problems lead to the so-called centdian problem [Ben-Moshe et al., 2012], and it is solved also for these special cases only with the connectivity constraint [Kang et al., 2018].

For solving the general case of the "weak version", we can consider another approach by submodular maximization received quite significant attention in the last 20 years. In their pioneer paper Nemhauser et al [Nemhauser et al., 1978] proved that nonnegative monotone submodular functions with matroid constraints can be approximated with $1 - 1/e$. This result was successfully applied by Kempe et al for influence maximization [Kempe et al., 2003] and later several applications were identified such as Machine learning methods among others. For a review see [Krause and Golovin, 2014]. Though the center and median problems can be easily transformed to appropriate submodular maximization, with connectivity constraint they are losing the matroid property and generally we can consider it as an optimization on the so-called k-systems [Feldman et al., 2011]. Unfortunately with this framework we can guarantee the trivial $1/k$ approximation only. On the other hand, specializing the question for submodular maximization on graphs with connectivity constraints an $1/\sqrt{k}$ approximation can be achieved. [Kuo et al., 2013]. It has several applications, typically concentrating on some sensor network problems, however all of them are different from our questions [Zhang and Vorobeychik, 2016], [Yu et al.,] A special case of the so-called h-hop submodular functions were also considered recently, however our problems do not have this property. [Xu et al., 2022]

If we focus to the "strong version" of our problem we arrive to the connected facility location problem. More exactly our problem is similar to the "rent-or-buy" problem. In that case for each edge we have a "rent" and a "buy" cost. In our setting we can distinguish the "gateway cost" (when two gateways are connected) and the "sensor cost" (when two sensors or a sensor with a gateway are connected). However, these problems do not require for the "gateways" to be connected, but searching the minimum cost Steiner tree for them as "gateway cost". [Bardossy and Raghavan, 2010], [Gollowitzer and Ljubić, 2011], [Eisenbrand et al., 2008], [Sun et al., 2021] In our strong version we could consider "rent-or-buy" connected problems (center, median and centdian). In that case it can be easily seen that the cost on the gateway subgraph is represented by the minimum spanning tree gateway cost.

Finally, if we consider the stochastic version of the problem we can incorporate the multipath routing [GUO and ZHANG, 2011] and the decentralized routing such as the Gossip protocol [Jelasity, 2011]. In that case we can consider the optimal design problem with respect to routing protocol types. Instead of distance in that case the weights are representing transmission probabilities between nodes. Together with a diffusion (or influence or infection) model the protocol can be described. For a unified concept of diffusion models see e.g. [Srivastava et al., 2015]. For P2P networks the gossip protocol model was widely studied, it can be transformed to diffusion model, such as in [Dong et al., 2015], where a similar questions to our problem was studied. Nevertheless, it is not a straightforward question what should be the objective in this stochastic case. In facility location the stochastic property is generally represented on the side of demands instead of the transmission, [Bieniek, 2015]. Though in a recent study ([Turke et al., 2021]) the transportation cost is also defined in a stochastic manner, but only by a finite number of scenarios. In that paper an excellent literature review about stochastic facility location problems can be also found. Related to stochastic models the approach of robust optimization also attracted a significant interest in recent years [An et al., 2014].

In our case the median type problem would be equivalent to the expected value of the communication cost, the center problem on the other hand would minimize the expected value of the number of iterations. [Saidi and Mohtashemi, 2012], [Robin et al., 2022]

4 Research questions

Based on the previous session some open problems are suggested and summarized below.

(1) Concerning weak S-routing systems as outlined in the previous section the main basic problems are the connected k-median and k-center problems.These problems were studied in the previous years, but with the focus on block and cactus graphs. It is a natural question to consider other special cases (e.g. planar graphs) on one hand, and on the other hand to study the general problem. Since the k-median and k-center objective functions are submodular, thus currently the best approximation ratio is provided by the [Kuo et al., 2013].

    It can be a a good objective to find a methodology by which a better approximation ratio can be guaranteed (at least with certain structures like sparse graphs) for the general case of submodular functions or to find a method with better approximation ratio for specifically the connected k-median and k-center problems. For these problems as a conjecture a potential approach could be to have an approximation ratio with $1/p$,where p is the longest distance or the average distance among nodes.In the case of highly clustered graphs it would be a better approximation. A natural approach could be that instead of k iterations we are running for a larger set (like $p\dot{k}$) and proving the existence of a connected subgraph among the gateway nodes.

(2) For strong S-routing systems as outlined earlier, the related problem is the single sink rent-or-buy problem. The key difference is that in the rent-or-buy problem the connectivity of the subgraph of "gateway nodes" is not required . Therefore for the k-median case here we can define a new problem when the objective is to minimize the cost of the linear combination of the minimum cost spanning tree of subgraph S and the "$k - median$ cost" from the sensor nodes to S. We can call it e.g. strong rent-or-buy problem. For the k-center problem it is already a question what would be a meaningful objective function. Since it is a new problem it would be possible to start with special cases such as cactus graphs and block graphs. In this manner it might be useful to review first the solution methodology of the "weak versions" and try to extend the approach. [Kang et al., 2018], [Bai et al., 2016], [Bai et al., 2021]

    Again, it is also a good target to solve the problem for other special cases (like planar graphs) as well as the general case. For these goals it can be useful to study the solution techniques developed for the rent-or-buy problem [Bardossy and Raghavan, 2010], [Gollowitzer and Ljubić, 2011], [Sun et al., 2021], [Eisenbrand et al., 2008]

(3) For the diffusion type problems again we are focusing on questions which were not studied yet in this aspect. The "weak version" of our problem family without edge weights are meaningful to consider. Different diffusion models can be considered, but as starting point it is natural to consider gossip models as earlier studies provide knowledge about the characteristics of the process. As outlined, for the "median type" question is the minimization of communication cost (i.e. the number of calls):

finding S for which the number of expected calls on the sensor nodes is minimized with the objective that all node information need to arrive to S. For the "center type" problem the challenge is to minimize the number of expected iterations: finding S for which the expected number of iterations is minimized with the objective that all node information need to arrive to S. For this research goal it is relevant to study those results describing the expected communication cost and the iteration number for spreading the information on the whole network [Robin et al., 2022], [Saidi and Mohtashemi, 2012]. By the characteristics of our problems, however, research focusing on the expected cost of reaching a target vertex in a stochastic graph environment can be also very significant. [Akrida et al., 2020] Nevertheless,concerning the earlier mentioned special cases would be also beneficial to study. Moreover, as in this setting the "center-type" and "median-type" problems were not studied at all (the only related paper is [Blagojevic et al., 2012], but still a different approach), it also makes sense that they could be considered without the connectivity constraint as initial research questions.

(4) Finally, as mentioned in the previous section, robust optimisation became an important field in recent years. Concerning any of the above mentioned problems in this framework (again, even for special graphs) would be relevant. Without connectivity constraint the k-median was studied in [An et al., 2014].

References

[Akrida et al., 2020] Akrida, E. C., Mertzios, G. B., Nikoletseas, S., Raptopoulos, C., Spirakis, P. G., and Zamaraev, V. (2020). How fast can we reach a target vertex in stochastic temporal graphs? *Journal of Computer and System Sciences*, 114:65–83.

[An et al., 2014] An, Y., Zeng, B., Zhang, Y., and Zhao, L. (2014). Reliable p-median facility location problem: two-stage robust models and algorithms. *Transportation Research Part B-methodological*, 64:54–72.

[Bai et al., 2016] Bai, C., Kang, L., and Shan, E. (2016). The connected p-center problem on cactus graphs. *Theor. Comput. Sci.*, 749:59–65.

[Bai et al., 2021] Bai, C., Zhou, J., Liang, Z., and De Maio, C. (2021). The connected p-median problem on cactus graphs. *Intell. Neuroscience*, 2021.

[Bardossy and Raghavan, 2010] Bardossy, M. G. and Raghavan, S. (2010). Dual-based local search for the connected facility location and related problems. *INFORMS J. Comput.*, 22:584–602.

[Ben-Moshe et al., 2012] Ben-Moshe, B., Dvir, A., Segal, M., and Tamir, A. (2012). Centdian computation in cactus graphs. *J. Graph Algorithms Appl.*, 16:199–224.

[Bieniek, 2015] Bieniek, M. (2015). A note on the facility location problem with stochastic demands. *Omega*, 55(C):53–60.

[Blagojevic et al., 2012] Blagojevic, M., Geilen, M., Basten, T., and Hendriks, T. (2012). Fast sink placement for gossip-based wireless sensor networks. In *2012 IEEE 31st International Performance Computing and Communications Conference (IPCCC)*, pages 110–119.

[de Araujo Marques Leão, 2018] de Araujo Marques Leão, L. (2018). *Optimization of Communications in Multi-Sink Wireless Sensor Networks*. PhD thesis, Université Bourgogne Franche-Comté.

[Dong et al., 2015] Dong, W., Yang, Y., and Zhang, W. (2015). On the selection of information sources for gossip spreading. *International Journal of Distributed Sensor Networks*, 11.

[Eisenbrand et al., 2008] Eisenbrand, F., Grandoni, F., Rothvoss, T., and Schäfer, G. (2008). Approximating connected facility location problems via random facility sampling and core detouring. In *SODA '08*.

[Feldman et al., 2011] Feldman, M., Naor, J., Schwartz, R., and Ward, J. (2011). Improved approximations for k-exchange systems - (extended abstract). In *ESA*.

[Gollowitzer and Ljubić, 2011] Gollowitzer, S. and Ljubić, I. (2011). Mip models for connected facility location: A theoretical and computational study. *Computers & Operations Research*, 38.

[GUO and ZHANG, 2011] GUO, W. and ZHANG, W. (2011). Routing between nodes and multiple gateways in wireless mesh sensor network. *Journal of Circuits, Systems and Computers*, 20(08):1477–1503.

[Jelasity, 2011] Jelasity, M. (2011). *Gossip*, pages 139–162. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Kang et al., 2018] Kang, L., Zhou, J., and Shan, E. (2018). Algorithms for connected p-centdian problem on block graphs. *J. Comb. Optim.*, 36(1):252–263.

[Kempe et al., 2003] Kempe, D., Kleinberg, J., and Tardos, E. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, page 137–146, New York, NY, USA. Association for Computing Machinery.

[Krause and Golovin, 2014] Krause, A. and Golovin, D. (2014). Submodular function maximization. In *Tractability*.

[Kuo et al., 2013] Kuo, T.-W., Lin, K. C.-J., and Tsai, M.-J. (2013). Maximizing submodular set function with connectivity constraint: Theory and application to networks. *IEEE/ACM Transactions on Networking*, 23:533–546.

[Nemhauser et al., 1978] Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14:265–294.

[Nguyen et al., 2022] Nguyen, K. T., Nhan, T. H. N., Teh, W. C., and Hung, N. T. (2022). The connected p-median problem on complete multi-layered graphs. *Discret. Math. Algorithms Appl.*, 14:2150118:1–2150118:13.

[Robin et al., 2022] Robin, F., Sericola, B., Anceaume, E., and Mocquard, Y. (2022). Stochastic analysis of rumor spreading with multiple pull operations. *Method. Comput. Appl. Prob.*, 24(3):2195–2211.

[Saidi and Mohtashemi, 2012] Saidi, A. G. and Mohtashemi, M. (2012). Minimum-cost first-push-then-pull gossip algorithm. *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2554–2559.

[Srivastava et al., 2015] Srivastava, A., Chelmis, C., and Prasanna, V. K. (2015). The unified model of social influence and its application in influence maximization. *Social Network Analysis and Mining*, 5:1–15.

[Sun et al., 2021] Sun, J., Sheng, H., Sun, Y., Du, D., and Zhang, X. (2021). Approximation algorithms for stochastic set cover and single sink rent-or-buy with submodular penalty. *J Comb Optim*, https://doi.org/10.1007/s10878-021-00753-x:649–694.

[Tansel et al., 1983] Tansel, B. , Francis, R. L., and Lowe, T. J. (1983). Location on networks: A survey. part i: The p-center and p-median problems.

[Turke et al., 2021] Turke, R., Sörensen, K., and Cuervo, D. P. (2021). A matheuristic for the stochastic facility location problem. *J. Heuristics*, 27:649–694.

[Xu et al., 2022] Xu, W., Peng, D., Liang, W., Jia, X., Xu, Z., Zhou, P., Wu, W., and Chen, X. (2022). Maximizing h-hop independently submodular functions under connectivity constraint. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, pages 1099–1108.

[Yu et al.,] Yu, N., Dai, H., Chen, G., Liu, A., Tian, B., and He, T. Connectivity-constrained placement of wireless chargers.

[Zhang and Vorobeychik, 2016] Zhang, H. and Vorobeychik, Y. (2016). Submodular optimization with routing constraints. In *AAAI*.