Aggregation Points Planning in Smart Grid Communication System

Xinxin Huang and Shaowei Wang

***Abstract*—Aggregation Point (AP) in neighborhood area network (NAN) plays a key role on smart grid communication. The AP stores and forwards data stream between home area network (HAN) and wide area network (WAN) in the smart grid. In this letter, we study how to plan the APs in the smart grid to minimize the total deployment cost of the opening APs and the connecting cost between the APs and the HANs while considering the capacity of the APs and the traffic demands of the HANs. Our general formulation leads to an NP-hard problem and an approximation algorithm is developed to address it. Numerical results show that our proposed algorithm has great advantages over other heuristic methods. Furthermore, our proposal gives a performance-guaranteed planning scheme of the APs for smart grid communication.**

***Index Terms*—Aggregation point, approximation algorithm, planning, smart grid communications.**

I. Introduction

SMART GRID is an enhancement power grid which is capable of delivering energy in more efficient ways [1], [2]. In contrast with traditional electricity grid, which is essentially a one-way transmission system from energy provider to users, smart grid uses a two-way flow of electricity and information to widely distributed automated energy delivery network. As a result, a well-designed communication network infrastructure is of great importance for the implementation of the smart grid.

Smart grid communication system is an interconnected network which generally employs a three-layer infrastructure that is shown in Fig. 1: home area network (HAN), neighborhood area network (NAN) and wide area network (WAN) [3]. The HANs always distribute randomly around the whole area, collecting information from a variety of appliances in large scale, while the WAN usually locates in the network data center of electricity companies [4] as shown in Fig. 1. As for the NANs, each of them serves multiple HANs by deploying an aggregation point (AP) to receive information from those HANs. All the data generated by the HANs, such as demand response information of the appliances, will be transmitted to the control center via the APs. Meanwhile, the other way of data flow, such as the control information for the purpose of load management and price broadcasting, is also transmitted from the control center to the HANs via the APs. For both cases, the APs work as a relay between the HANs and the WAN. Therefore, the APs are the core infrastructure for smart grid communication and should be designed carefully to meet the requirements of the scalability and the reliability of smart grid.

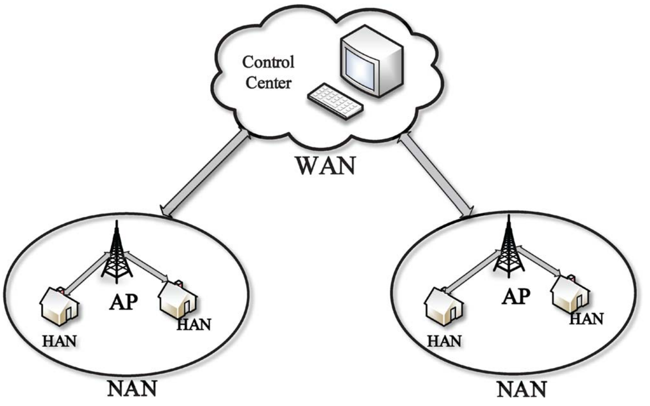

Fig. 1. Architecture of smart grid communication.

The deployment of the APs generally includes the installation of APs and the physical links between the APs and the HANs. In most cases, the connection cost between an AP and the HANs served by it is expensive. As a result, the APs should be planned in a cost-efficient way to reduce the capital expenditure of the smart grid communication system. Fiber optic communication and power line communication (PLC) are promising for the links between the APs and the HANs. The former can provide high speed data transmission, as well as good security guarantee. The latter is also an attractive scheme since the PLC has existed in a power grid [5], [6].

In this letter, we investigate the APs planning in smart grid. We show that the planning task leads to a facility localization problem [7], [8] and introduce an efficient algorithm to solve it. Numerical results show our proposal performs better than other heuristic methods, such as genetic algorithm (GA) and Tabu search (TS). Furthermore, our proposed method is an approximation algorithm that is performance-guaranteed, which sheds some insights on deploying a cost-efficient smart grid communication system.

II. System Model and Problem Formulation

Consider a certain area that should be served by a smart grid. Optical fiber cable is available for the links between the APs and HANs. Denote $\mathcal{N} = \{1, 2, \ldots, N\}$ as a set of candidate sites where the APs can be installed. $\mathcal{M} = \{1, 2, \ldots, M\}$ is the set of the HANs. Each AP $n \in \mathcal{N}$ requires an installation cost of c_n and equips with a capacity of w_n. Each HAN $m \in \mathcal{M}$ requests a traffic demand d_m. When an AP is selected for opening, the cost of connecting the HAN m with the AP n by optical fiber is u_{nm}, which is related to their distance c_{nm}, that is, $u_{nm} = \phi(c_{nm})$. For simplification, the connection cost u_{nm}

Manuscript received January 8, 2015; revised May 27, 2015; accepted May 27, 2015. Date of publication June 4, 2015; date of current version August 10, 2015. The associate editor coordinating the review of this paper and approving it for publication was H. Mohsenian-Rad. *(Corresponding author: Shaowei Wang.)*

The authors are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: huangxx@smail.nju.edu.cn; wangsw@nju.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2015.2441722

is assumed to be linear to c_{nm}. $H(n, m)$ is the attenuation of powerline cables between the AP n and the HAN m and can be expressed as

$$H(n, m) = ae^{-\varphi(c_{nm})},$$

where a is a weighting factor. $\varphi(c_{nm})$ is an attenuation factor that is linear to the distance c_{nm} between AP n and HAN m [9].

Define x_n as an index variable which indicates whether the AP n is selected or not,

$$x_n = \begin{cases} 1 & \text{if AP } n \text{ is selected,} \\ 0 & \text{otherwise.} \end{cases}$$

Let z_{nm} be an index variable to show whether the HAN m is assigned to the AP n or not, that is,

$$z_{nm} = \begin{cases} 1 & \text{if HAN } m \text{ is assigned to AP } n, \\ 0 & \text{otherwise.} \end{cases}$$

Our planning objective is to select a subset from the candidate sites to install APs to satisfy the traffic demands of all HANs with the minimum cost that consists of the expenditures of the APs and the fiber optic cable cost between the APs and the HANs. Mathematically, the optimization problem can be written as follows:

$$\begin{aligned} \min_{x_n, z_{nm}} \quad & \sum_{n\in\mathcal{N}} c_n x_n + \sum_{n\in\mathcal{N}} \sum_{m\in\mathcal{M}} u_{nm} z_{nm} \\ s.t. \quad & C_1 : \sum_{n\in\mathcal{N}} \sum_{m\in\mathcal{M}} d_m z_{nm} \geq \sum_{m\in\mathcal{M}} d_m, \\ & C_2 : \sum_{m\in\mathcal{M}} d_m z_{nm} \leq w_n x_n, \ \forall n \in \mathcal{N}, \\ & C_3 : H(n, m) \geq QoS \times z_{nm}, \ \forall n \in \mathcal{N}, \ \forall m \in \mathcal{M}, \\ & C_4 : x_n \geq z_{nm}, \ \forall n \in \mathcal{N}, \ \forall m \in \mathcal{M}, \\ & C_5 : \sum_{n\in\mathcal{N}} z_{nm} = 1, \ \forall m \in \mathcal{M}, \\ & C_6 : x_n \in \{0, 1\}, \ \forall n \in \mathcal{N}, \\ & C_7 : z_{nm} \in \{0, 1\}, \ \forall n \in \mathcal{N}, \ m \in \mathcal{M}. \end{aligned} \tag{1}$$

C_1 ensures that the total traffic demand of the HANs should be satisfied. C_2 means that the sum traffic demand provided by an AP cannot exceed its capacity. C_3 ensures the communication link reliability between the APs and the HANs. The QoS is defined as the minimum attenuation level. C_4 means that a HAN can be covered by an AP that has been selected to open. C_5 ensures a HAN can be covered by only one AP. It is easy to verify that (1) defines an integer programming problem. More specifically, it is equivalent to facility location problem [7].

As an alternative, PLC can also be used to provide the links between the APs and the HANs. The difference is that there is no deployment cost between an AP and the HANs served by it since the PLC usually exists in a power grid. Thus the planning problem for the PLC case is as follows:

$$\begin{aligned} \min_{x_n, z_{nm}} \quad & \sum_{n\in\mathcal{N}} c_n x_n \\ s.t. \quad & C1 \sim C7. \end{aligned} \tag{2}$$

Intuitively, (2) can be seen as a special case of (1), where $u_{nm} = 0$.

III. PROPOSED ALGORITHM

We introduce an algorithm that is based on the primal-dual method. Our proposed algorithm can produce performance-guaranteed solutions. The relaxation form of (1) can be written as follows:

$$\begin{aligned} \min_{x_n, z_{nm}} \quad & \sum_{n\in\mathcal{N}} c_n x_n + \sum_{n\in\mathcal{N}} \sum_{m\in\mathcal{M}} u_{nm} z_{nm} \\ s.t. \quad & C1 \sim C5, \\ & C_6 : x_n \leq 1, \ \forall n \in \mathcal{N}, \\ & C_7 : z_{nm} \geq 1, \ \forall n \in \mathcal{N}, \ m \in \mathcal{M}. \end{aligned} \tag{3}$$

(3) is the linear programming relaxation of (1), where the integrality constraints C_6 and C_7 of (1) are relaxed to rational values between 0 and 1. The dual form of (3) is as follows:

$$\begin{aligned} \max & \sum_{m\in\mathcal{M}} \alpha_m \\ s.t. \quad & C_1 : \alpha_m \leq u_{nm} + \beta_{nm} + d_m \gamma_n + H(n, m) \rho_n, \\ & \qquad \forall n \in \mathcal{N}, \ m \in \mathcal{M}, \\ & C_2 : \sum_{m\in\mathcal{M}} \beta_{nm} \leq c_n + \delta_n - w_n \gamma_n - QoS \rho_n, \ \forall n \in \mathcal{N}, \\ & C_3 : \alpha_m, \beta_{nm}, \gamma_n, \delta_n, \ \rho_n \geq 0, \ \forall n \in \mathcal{N}, \ m \in \mathcal{M}. \end{aligned} \tag{4}$$

$\alpha_m, \beta_{nm}, \gamma_n, \delta_n, \rho_n$ are dual variables. Intuitively, we can take α_m as the total cost of HAN m including the connecting cost and part of deployment cost of the AP which serves HAN m.

A special case of (1) is that there is only one HAN, which plays an important role for designing an approximation algorithm for (1). Assume that only HAN $k \in \mathcal{M}$ exists, (1) can be described as follows:

$$\begin{aligned} \min_{v_n, g_n} \quad & \sum_{n} c_n v_n + \sum_{n} c'_n g_n \\ s.t. \quad & C_1 : \sum_{n\in L_k} g_n \geq D_k, \\ & C_2 : g_n \leq w_n v_n, \ \forall n \in L_k, \\ & C_3 : H_n \geq QoS g_n, \ \forall n \in L_k, \\ & C_4 : v_n \leq 1, \ \forall n \in L_k, \\ & C_5 : v_n, g_n \geq 0, \ \forall n \in L_k, \end{aligned} \tag{5}$$

TABLE I
GREEDY ALGORITHM FOR THE SINGLE HAN PROBLEM

Algorithm: Greedy Algorithm

```
Initialize $g_n = v_n = 0$;
if $D_k \neq 0$
  for $n$ in increasing order of $(\frac{c_n}{w_n} + c'_n)$;
    $g_n = \min(w_n, D_k)$;
    $v_n = \frac{g_n}{w_n}$;
    $D_k = D_k - g_n$;
  end for
end if
return $(g, v)$
```

TABLE II
CLUSTERING ALGORITHM

Step 1: Clustering

```
while $S \neq \emptyset$ do
  for $m$ in increasing order of $\alpha_m$
    $B_m = \{n \in F_m : n \notin \bigcup_{k\in C} N_k, c_{nm} \leq \min_{k\in C} c_{nk}\}$;
    $S = S \setminus m$;
    while $B_m \neq \emptyset$
      $k = m$;
      $N_k = B_m$;
      $C = C \cup \{k\}$;
    end while
  end for
end while
$U = F - \bigcup_{k\in C} N_k$;
for $n \in U$ do
  if $k = \arg\min_{k\in C} c_{nk}$
    $N_k = N_k \cup \{n\}$;
  end if
end for
```

TABLE III
ROUNDING ALGORITHM

Step 2: Rounding

```
for each $N_k$
  $L_k = \{n \in N_k : x_n < 1\}$;
  $D_k = \sum_{n\in L_k} \sum_m z_{nm}$;
  find $(g_n^{(k)}, v_n^{(k)})$ by greedy algorithm in Table I and
  calculate the value $O_k^*$;
end for
for each $N_k$
  while $0 < v_n^k < 1$
    $x_n = 1$;
  end while
end for
```

TABLE IV
PARAMETERS OF THE GA AND TS

GA		TS	
Population size	100	Tabu size (active)	3
Crossover probability	0.7	Tabu size (candidate)	7
Mutation probability	0.3	Iterations	300

where L_k is the set of the APs that fractionally serve HAN k. D_k is the total traffic demand served by these APs, and c'_n is the cost of connecting HAN k with AP n. v_n indicates whether AP n is open or not, and g_n is the traffic demand assigned to the AP n. H_n is the attenuation between the HAN k and the AP n. We can set $\widehat{v}_n = \frac{g_n}{w_n}$ and obtain a feasible solution with no greater cost. A Greedy algorithm for the special single HAN problem is shown in Table I, by which an optimal solution to (5) can be obtained.

Denote $F = \{n : x_n > 0\}$ as the set of APs and $F_m = \{n : n \in F, z_{nm} > 0\}$ as the set of APs in F that fractionally serve the HAN m, respectively. Our proposed approximation algorithm is divided into two steps: clustering and rounding.

Step 1 Clustering: We partition the APs with $x_n > 0$ into clusters, each of them will be centered around a HAN. We call this as a cluster center. Denote the cluster centered around the HAN k as N_k. N_k consists of the HAN k, the set of the APs assigned to it, and the fractional demands served by these APs.

Let C be the set of current cluster centers which is initially empty and S be the set of all HANs that could be chosen as cluster centers which is initially $\mathcal{M}$. For each HAN $m \notin C$, B_m represents the set of unclustered APs that are closer to it than to any other clustered center. To find all cluster centers, $m \in \mathcal{M}$ is ordered in increasing of α_m and the cluster N_k can be formed by B_m. The procedure of clustering is described in Table II.

Step 2 Rounding: In this step, we will decide which AP will be fully opened in each cluster. For each cluster obtained in the first step, it can be seen as the single HAN problem. Therefore, the cluster N_k can obtained by the Greedy algorithm shown in Table I. Then we have $L_k = \{n \in N_k : x_n < 1\}$ and $D_k = \sum_{n\in L_k} \sum_m d_m z_{nm}$, where the tuple of $(g_n^{(k)}, v_n^{(k)})$ is an optimal solution produced by the Greedy algorithm. While $0 < v_n^{(k)} < 1$, all APs are fully opened in L_k, i.e., $x_n = 1$ for $n \in L_k$. Piecing together the solutions for all clusters, x_n's are assigned by $\{0, 1\}$. Once all APs have been located, each HAN is assigned to the closest AP around it. The procedure of rounding is described in Table III. Basically, our proposed method is an approximation algorithm. We do not give the proof in detail of the approximation ratio because of room limitation. Similar proof can be found in [8]. The computational complexities of step 1 and step 2 are $O(N^2M)$ and $O(NM)$, respectively. We can conclude that the complexity of our proposed algorithm is bounded by $O(N^2M)$.

IV. NUMERICAL RESULTS AND DISCUSSIONS

We compare the performance of our proposed approximation algorithm with other heuristic methods with reasonable complexity: GA proposed in [10] and TS developed in [11]. The parameters of the GA and the TS are shown in Table IV. Consider an area served by a smart grid communication system, where all HANs and candidate sites for installing APs locate randomly in the area. The number of candidate sites is set to 60. For the candidate site n, its capacity w_n is distributed uniformly within (600, 900). For HAN m, its traffic demand d_m is distributed uniformly within (20, 30) [12]. The weighting factor g is 1 and the QoS is set to 20 as proposed in [9]. Without loss

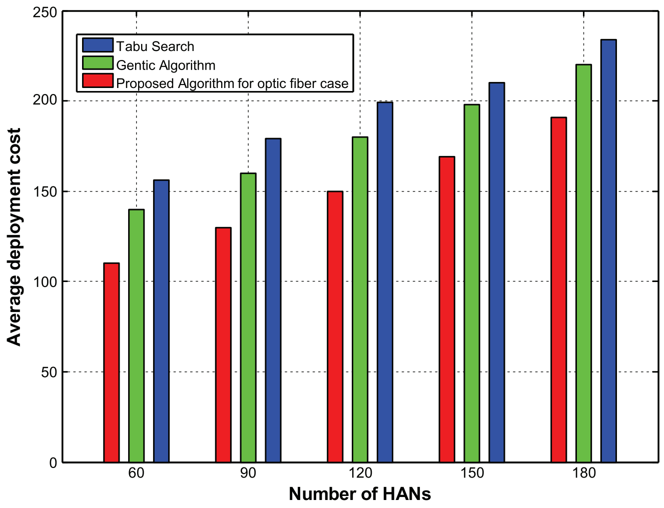

Fig. 2. Average deployment cost for the fiber optic case as the number of the HANs M changes. $N = 60$, $c_n = 9.3$, $\forall n \in \mathcal{N}$, $u_{nm} = 0.001c_{nm}$, $\forall n \in \mathcal{N}$, $m \in \mathcal{M}$.

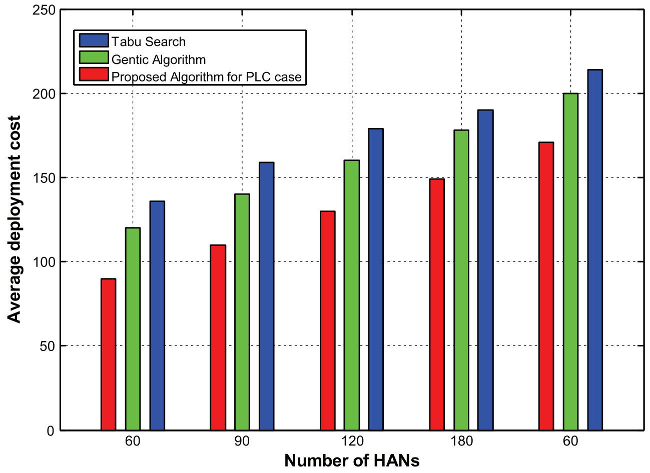

Fig. 3. Average deployment cost for the PLC case as the number of the HANs M changes. $N = 60$, $c_n = 9.3$, $\forall n \in \mathcal{N}$, $u_{nm} = 0$, $\forall n \in \mathcal{N}$, $m \in \mathcal{M}$.

of generality, we assume the installation costs of all APs are equal.

Figs. 2 and 3 show the average costs of deploying APs with fiber optic cable and PLC, respectively, as a function of the number of HANs. The cost of AP n is set to $c_n = 9.3$. The connection cost of the AP n and the HAN m for the fiber cable case is $u_{nm} = 0.001c_{nm}$, where c_{nm} is the distance between the HAN and the serving AP. As can be seen from the Figs. 2 and 3, our proposed algorithm outperforms the GA and the TS notably. In the fiber optic scenario, the total cost of our proposal is about 20% lower than that of the tabu TS, which is also the case for the PLC scenario as shown in Fig. 3. Obviously, when the number of HANs increases, the total deployment cost also increases.

From Figs. 2 and 3 we can see that the approximation algorithm can produce promising solutions to the AP planning problem, comparing to other heuristic methods. Recall that our proposed algorithm has the worst case performance guarantee. It is difficult even impossible for other heuristic methods to provide such a guarantee. So we can conclude conservatively that our proposal is promising for applications in smart grid with requirements of scalability and reliability.

Also, as can be found in Figs. 2 and 3, the cost of employing fiber optic cable is much higher than that of the PLC. The reason is intuitive: The PLC requires no connection cost between the HAN and the AP serving it. So it is cost-effective to adopt PLC if the required data rate is not high. However, the fiber optic cable is preferred for the high data rate and security requirement scenario.

V. CONCLUSION

In this letter, we studied the minimum cost APs planning problem in the smart grid communication system. Two wired access technologies, fiber optic communication and PLC, are considered. The general problem formulation yields an NP-hard optimization task. We introduce an approximation algorithm that can yield solutions with the worst case performance guarantee. Numerical results show our proposed algorithm performs better than other representative heuristic ones. In future work, wireless technologies or hybrid networking model should be investigated, e.g., smart grid may employ wired technology for parts of the service area and wireless technology for the others, depending on the data rate requirement, the installation expenditure and the connection cost between the APs and the HANs.

REFERENCES

[1] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—The new and improved power grid: A survey," *IEEE Commun. Survey Tuts.*, vol. 14, no. 4, pp. 944–980, Dec. 2012.

[2] P. Samadi, H. Mohsenian-Rad, R. Schober, and V. Wong, "Advanced demand side management for the future smart grid using mechanism design," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1170–1180, Sep. 2012.

[3] W. Meng, R. Ma, and H.-H. Chen, "Smart grid neighborhood area networks: A survey," *IEEE Netw.*, vol. 28, no. 1, pp. 24–32, Jan. 2014.

[4] M. Ghamkhari and H. Mohsenian-Rad, "Energy and performance management of green data centers: A profit maximization approach," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 1017–1025, Jun. 2013.

[5] V. Gungor *et al.*, "Smart grid technologies: Communication technologies and standards," *IEEE Trans. Ind. Inf.*, vol. 7, no. 4, pp. 529–5239, Nov. 2011.

[6] S. Bush, S. Goel, and G. Simard, "IEEE vision for smart grid communications: 2030 and beyond roadmap," in *IEEE Visual Smart Grid Communication: 2030 Beyond Roadmap*. Piscataway, NJ, USA: IEEE Press, Sep. 2013, pp. 1–19.

[7] K. Jain and V. V. Vazirani, "Primal-dual approximation algorithms for metric facility location and k-median problems," in *Proc. IEEE FOCS*, 1999, pp. 2–13.

[8] R. Levi, D. B. Shmoys, and C. Swamy, "LP-based approximation algorithms for capacitated facility location," *Math. Program.*, vol. 131, no. 1/2, pp. 365–379, Feb. 2012.

[9] M. Zimmermann and K. Dostert, "A multipath model for the powerline channel," *IEEE Trans. Commun.*, vol. 50, no. 4, pp. 553–559, Apr. 2002.

[10] N. Weicker, G. Szabo, K. Weicker, and P. Widmayer, "Evolutionary multiobjective optimization for base station transmitter placement with frequency assignment," *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 189–203, Apr. 2003.

[11] C. Lee and H. Kang, "Cell planning with capacity expansion in mobile communications: A tabu search approach," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1678–1691, Sep. 2000.

[12] Y. Zhang *et al.*, "A multi-level communication architecture of smart grid based on congestion aware wireless mesh network," in *Proc. IEEE NAPS*, 2011, pp. 1–6.