Aggregation Points Planning for Software-Defined Network Based Smart Grid Communications

Shaowei Wang *† and Xinxin Huang†
* State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
† School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China
E-mail: wangsw@nju.edu.cn, huangxx@smail.nju.edu.cn

***Abstract*—Smart grid is characterized by a large number of smart meters (SMs) that exchange huge amounts of data with control center, where an effective communication network is required to guarantee reliable data transmission. In this paper, we introduce software defined network (SDN) technology to the smart grid, which decouples the control plane from the data plane so as to satisfy the communication requirements in the mart grid effectively. Aggregation points (APs) are employed in the data plane to process and forward data between SMs and control center. A general mathematical model is formulated to plan the APs, where we try to minimize the total deployment cost, including the opening expenditure of the APs, the connection cost between SMs and APs, and the connection cost between APs and control center. We present a 5-approximation algorithm to address the generated NP-hard problem, which yields performance-guaranteed solutions. Three representative scenarios are investigated to verity the efficiency of our proposal. Numerical results show that our proposed algorithm has great advantages over other heuristic ones.**

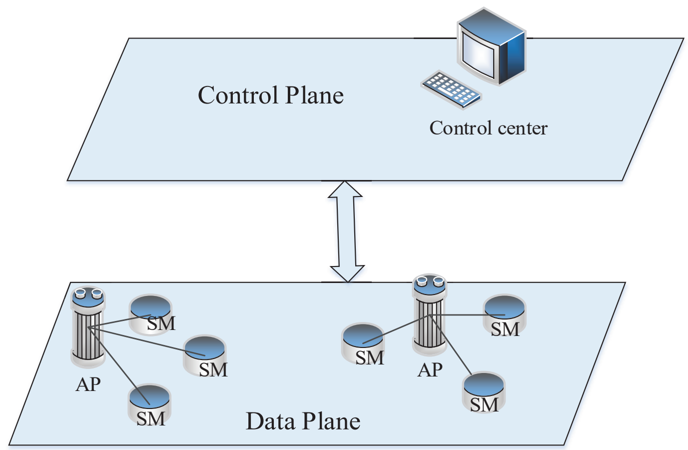


Fig. 1. Architecture of SDN-based smart grid communication.

I. Introduction

As an evolution of conventional power systems, smart grid employs advanced information and communication technologies to construct an intelligent bi-direction electricity and communication network [1]. That is, a smart grid not only delivers electricity from suppliers to consumers, but also uses a two-way flow of data communications to exchange information between consumers and suppliers to provide various services, such as integrating renewable distributed energy resources, applying demand side management and dealing with increased energy trade [2, 3].

Software defined networking (SDN), which decouples the control plane from the data plane, is a promising method that can meet the requirements of the smart grid communication [4–9]. SDN can be applied for developing self-configuring intelligent electric devices for the substation communication network [10]. In [11], multicast transmission of measurement unit data is handled by an SDN enable network to adapt different data rates requirements. Investigations in [12] indicate that SDN can enhance data exchange and distribute resources in the smart grid efficiently. A multi-layer SDN for the smart grid is proposed in [13], where the southbound application program interface is employed to the communication between the data and the control plane.

In this paper, we introduce a novel SDN-based smart grid communication architecture, as shown in Fig.1. Smart meters (SMs) in the data plane gather energy consumption and control information from residential and industrial consumers [14]. Recall that there are a huge number of SMs in the smart grid, it is not practical to send these data to the control center by deploying a direct link between each SM and the control center. We suggest that multiple aggregation points (APs) should be installed in the data plane, and each of them serves its surrounding SMs to assist the information exchanging between SMs and the control center. The data generated by SMs, such as demand response information, will be sent to the control center via these APs [15, 16]. Meanwhile, the other way of data flow, such as the control information for load management and price broadcasting, can be also sent from the control center to SMs via these APs.

As can be seen from Fig.1, an AP receives data from its serving SMs and send the data to the control center. Since the SMs always generate a large amount of data, the AP needs to compress the data before transmitting them for the consideration of transmission efficiency. The compression ratio usually varies from $1/2$ to $2/3$, depending on the correlation of data [14]. In other words, the APs not only receive and forward information, but also process the raw data originated from the SMs. Thus planing APs is an important task for designing such an SDN-based smart grid communication system. The APs planning includes the installation of APs, the physical links between each AP and its serving SMs, and the links between the APs and the control center.

To reduce the capital expenditure of the smart grid communication system, APs should be planned in a cost-efficient way. Obviously, the key to achieve this goal is to minimize the number of APs while considering the links cost between APs and SMs, as well as the link cost between each AP and the control center. Since power line communication (PLC) always exists in the power systems, it is promising to use the PLC to provide the links between APs and SMs. As for the link between APs and the control center, it could be wired and wireless communication as discussed in [17]. In this paper, we employ the PLC to provide the communication link between APs and SMs. The link cost will b discussed in detail in the following section. To simplify analysis, the communication cost between each AP and the control center is a given value. We try to minimize the sum cost of the installation of APs, the links between APs and SMs, and the links between APs and the control center.

First, we give a brief survey of the related works. In [18], a general mathematical model is given to plan the APs which locate in the neighborhood area network of the smart grid. However, the channel characteristic of PLC is not considered in this work. In [19], the placement of data aggregation services in smart grid communication network is investigated, where the authors developed a minimum-cost-forwarding-based asynchronous distributed algorithm to find the optimal placement for the data aggregation service tree with the minimal cost of in-network processing. The controller placement problem in SDN is discussed in [20], where the objective is to determine the number and the locations of controllers, as well as the connection cost among the controllers and the switches. The relay placement problem in smart grid is studied in [21], where the optimization task is to place the minimum number of relays so that all intelligent devices in the smart grid can communicate with each other via these relays under the constraint of the coverage range of each relay. A 4.5-approximation algorithm is introduced. The relay placement problem in [21] is different from the APs planning problem investigated in this paper. The APs in the data plane need not to communicate with each other. The APs in our considered SDN-based sart grid communication system work as the base stations in the cellular systems for mobile communications. In [22], an optimal placement of data aggregation points in advanced metering infrastructures for smart grid is proposed. The objective is to minimize the total installation and transmission cost. It yields an integer program problem and the authors proposed K-means algorithm to solve it efficiently.

In this paper, we investigate the APs planning problem in the SDN-based smart grid communication system, which has not been well discussed in the literature as far as the authors have known. We show that the planning task leads to a facility localization problem [23, 24] and introduce an efficient approximation algorithm to solve it. Three scenarios with different SM densities are considered to verify our proposed algorithm. Numerical results show our proposal performs better than other heuristic methods, such as genetic algorithm (GA) [25] and Tabu search (TS) [26]. Furthermore, our proposed algorithm is performance-guaranteed, shedding some insights on deploying a cost-efficient SDN-based smart grid communication system.

The rest of this paper is organized as follows. In Section II, we illustrate the PLC channel characteristics and formulate the APs planning problem in the SDN-based smart grid system. In Section III, we introduce a 5-approximation algorithm to address the formulated optimization task. Numerical results and discussions are given in Section IV. In Section V, we conclude this paper and point out future researches.

II. System Model and Problem Formulation

Consider an SDN-based smart grid communication system with M SMs, denoted by $\mathcal{M} = \{1, 2, \ldots, M\}$. There are M_c control centers in the control plane. Here we only consider the case of $M_c = 1$ to simplify analysis. Multiple APs are installed in the data plane to serve the SMs. That is, an SM must be connected with one and only on AP. Denote $\mathcal{N} = \{1, 2, \ldots, N\}$ as a set of candidate sites where the APs can be installed. PLC is available for the communication links between the APs and the SMs. Each AP is connected to one of the control centers by the way of wired or wireless communication [17]. The number of SMs is calculated as follows:

$$M = \rho_{SM} r^2,$$

where r is the radius of the area of interest and ρ_{SM} is the density of SMs in the area. The value of ρ_{SM} is depend on the area of interest being rural, urban or sub-urban [22]. To make the rest of this paper easy to follow, we give a summary of the notations frequently used in this paper in Table I.

A. PLC Channel Model

PLC uses the existing power line infrastructure to provide the communication links between the APs and the SMs in the SDN-based smart grid [27, 28]. The channel characteristics of PLC can be subdivided into transfer function H_f and additive noise N_f. We assume that signal propagation only takes place along a direct path between one AP and one SM. The loss of power line cable causes an attenuation $A(f, dis_{nm})$ that increases with frequency f and distance dis_{nm} between AP n and SM m. $B(f, dis_{nm})$ is the delay portion which is also related to f and dis_{nm}. Therefore, the transfer function H_f can be expressed as

$$H_f = g \cdot A(f, dis_{nm}) \cdot B(f, dis_{nm}),$$

where g is a weighting factor representing the product of the reflection and transmission factor along the path. It is not more than one,

$$g \leq 1.$$

The attenuation of power line cable can be characterized by

$$A(f, dis_{nm}) = e^{(a_0 + a_1 f) \cdot dis_{nm}},$$

where a_0 and a_1 are cable parameters discussed in [29].

The delay τ of the path between AP n and SM m

$$\tau = \frac{dis_{nm}\varepsilon_r}{v_0},$$

can be calculated from the dielectric constant ε_r of the insulating material, the speed of light v_0 and the distance dis_{nm} between AP n and SM m. Therefore, the delay portion $B(f, dis_{nm})$ can be expressed as

$$B(f, dis_{nm}) = e^{-2\pi jf\cdot\tau}.$$

Thus the transfer function H_f can be expressed as

$$H_f = g \cdot e^{(a_0+a_1 f)\cdot dis_{nm}} \cdot e^{-2\pi jf\cdot\tau}.$$

Another channel characteristic is the background noise of PLC. The power spectral density of the background noise is a decreasing function of frequency. It can be calculated as

$$N_f = 10^{K-3.95\cdot 10^{-5} f},$$

where the value of K varies with the transmitter or receiver locations [30]. Once an AP is selected, its location is also fixed. Then the PLC channel noise is a constant value for a given frequency.

B. Mathematical Model

The optimization objective of our APs planning problem is to install the minimum number of APs in the candidate sites $\mathcal{N}$ while considering the traffic demands of all the SMs and all the links cost discussed above. Assume that each AP $n \in \mathcal{N}$ requires an installation cost of c_n^{AP}. The connection cost between an AP and the control center is P. When AP n is selected for opening, the connecting cost between SM m and AP n is c_{nm}, which is related to the distance dis_{nm}:

$$c_{nm} = \phi(dis_{nm}).$$

We assume the connection cost c_{nm} is linear to the distance dis_{nm} for simplicity. However, tt is intuitive to extend our results to other convex cost functions.

Define x_n as an index variable that indicates whether AP n is selected for opening or not:

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

Let y_{nm} be an index variable to show whether SM m is connected to AP n or not:

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

SM $m \in \mathcal{M}$ requests a traffic demand d_m and AP $n \in \mathcal{N}$ is equipped with a capacity of w_n. The APs installed in the area of interest should satisfy all the traffic demands of SMs in this area. Meanwhile, the traffic demands of SMs served by an AP cannot exceed the capacity limitation of this AP.

TABLE I
NOTATIONS

Symbol	Semantics
$\mathcal{N}$	Candidate sites set for placing APs
$\mathcal{M}$	SMs set
M_c	Number of control centers
c_n^{AP}	Installation cost of AP n
x_n	Index variable showing that AP n is selected or not
y_{nm}	Index variable indicating whether SM m is assigned to AP n or not
d_m	Traffic demand of SM m
dis_{nm}	Distance between SM m and AP n
ρ_{SM}	SM density
c_{nm}	Link cost between SM m and AP n
P	Link cost between AP and control center
w_n	Capacity of AP n
OPT	Optimal value
c_n'	Per unit cost of distance to AP n
N_k	Cluster centered around SM k
k	Center of N_k
C	Set of cluster center
F	Set of APs that could be selected
F_m	APs in F that fractionally serve SM m
B_m	Set of unclustered APs
$n^*(m)$	AP in A_m nearest to m
A_m	Set of APs not in B_m

Mathematically, our APs planning optimization problem can be written as follows:

$$\begin{aligned} \min_{x_n, y_{nm}} \; & \sum_{n\in\mathcal{N}} c_n^{AP} x_n + \sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}} c_{nm} y_{nm} + \sum_{n\in\mathcal{N}} x_n P \\ s.t. \quad & C_1 : \sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}} d_m y_{nm} \geq \sum_{m\in\mathcal{M}} d_m, \\ & C_2 : \sum_{m\in\mathcal{M}} d_m y_{nm} \leq w_n x_n, \forall n \in \mathcal{N}, \\ & C_3 : \mid H_f \mid \geq Q \times y_{nm}, \forall n \in \mathcal{N}, \forall m \in \mathcal{M}, \\ & C_4 : x_n \geq y_{nm}, \forall n \in \mathcal{N}, \forall m \in \mathcal{M}, \\ & C_5 : \sum_{n\in\mathcal{N}} y_{nm} = 1, \forall m \in \mathcal{M}, \\ & C_6 : x_n \in \{0,1\}, \forall n \in \mathcal{N}, \\ & C_7 : y_{nm} \in \{0,1\}, \forall n \in \mathcal{N}, m \in \mathcal{M}, \end{aligned} \tag{1}$$

where

$$Q = Q_a \cdot Q_b.$$

Q_a and Q_b are the attenuation limit and the delay limit of the links between APs and SMs, respectively. The first term of the objective function accounts for the installation cost of APs. The second term is the link costs between SMs and APs, and the third term is the link costs between APs and the control center. C_1 guarantees that the total traffic demands of SMs should be satisfied. C_2 means that the sum traffic demand provided by an AP cannot exceed its capacity limitation. C_3 is the attenuation and relay requirements between APs and SMs. C_4 means that an SM should be served by an AP that has been selected to open. C_5 and C_7 ensure an SM can be served by only one AP. C_6 is intuitive.

III. Proposed Algorithm

Equation (1) defines an integer programming problem that is NP-hard. It falls into the general facility location problem class [23, 24]. The main difficulty of solving (1) lies in the integer constraints which usually generate exponential complexity if the optimal solution could be worked out. An intuitive way to tackle such kind of problems is the heuristic methods such as TS and GA, which can yield feasible solutions with reasonable complexity. However, these methods cannot give a provable gap between the produced solutions and the optimum of the original problem. In this paper, we introduce a 5-approximation algorithm to produce quality guaranteed solutions. The relaxation form of (1) can be written as follows:

$$\begin{aligned}
&\min_{x_n, y_{nm}} \sum_{n\in\mathcal{N}} c_n x_n + \sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}} c_{nm} y_{nm} \\
s.t.\quad & C_1 : \sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}} d_m y_{nm} \geq \sum_{m\in\mathcal{M}} d_m, \\
& C_2 : \sum_{m\in\mathcal{M}} d_m y_{nm} \leq w_n x_n, \forall n \in \mathcal{N}, \\
& C_3 : \mid H_f \mid \geq Q \times y_{nm}, \forall n \in \mathcal{N}, \forall m \in \mathcal{M}, \\
& C_4 : x_n \geq y_{nm}, \forall n \in \mathcal{N}, \forall m \in \mathcal{M}, \\
& C_5 : \sum_{n\in\mathcal{N}} y_{nm} = 1, \forall m \in \mathcal{M}, \\
& C_6 : 0 \leq x_n \leq 1, \forall n \in \mathcal{N}, \\
& C_7 : 0 \leq y_{nm} \leq 1, \forall n \in \mathcal{N}, m \in \mathcal{M}.
\end{aligned} \tag{2}$$

where

$$c_n = c_n^{AP} + P.$$

Equation (2) 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 (2) can be written as follows:

$$\begin{aligned}
&\max \sum_{m\in\mathcal{M}} \alpha_m \\
s.t.\quad & C_1 : \alpha_m \leq c_{nm} + \beta_{nm} + d_m\gamma_n + H_f\rho_n, \\
& C_2 : \sum_{m\in\mathcal{M}} \beta_{nm} \leq c_n + \delta_n - w_n\gamma_n - Q\rho_n, \\
& C_3 : \alpha_m, \beta_{nm}, \gamma_n, \delta_n, \rho_n \geq 0.
\end{aligned} \tag{3}$$

$\alpha_m, \beta_{nm}, \gamma_n, \delta_n, \rho_n$ are dual variables. The original problem (2) is convex and is equivalent to the dual problem (3). We can take α_m as the total cost of SM m including the connecting cost and the deployment cost of the AP that serves SM m. That is, α_m also includes part of the installation cost of this AP.

A special case of (2), for which there is only one SM, plays an important role for designing our proposed algorithm. Assume that only one SM $k \in \mathcal{M}$ exists, the special case can be described as follows:

$$\begin{aligned}
&\min_{v_n, g_n} \sum_n c_n v_n + \sum_n c_n' g_n \\
s.t.\quad & C_1 : \sum_{n\in L_k} g_n \geq D_k, \\
& C_2 : g_n \leq w_n v_n, \forall n \in L_k, \\
& C_3 : \mid H_f \mid \geq Q g_n, \forall n \in L_k, \\
& C_4 : v_n \leq 1, \forall n \in L_k, \\
& C_5 : v_n, g_n \geq 0, \forall n \in L_k,
\end{aligned} \tag{4}$$

where L_k is the set of APs that fractionally serves SM k. D_k is the total traffic demand served by these APs, and c_n' is the cost of connecting SM k with AP n; v_n indicates if AP n is open or not; g_n is the traffic demand assigned to AP n. Set $v_n = \frac{g_n}{w_n}$, and we can obtain a feasible solution with no greater cost. Therefore, we rewrite (4) into the following form:

$$\begin{aligned}
&\min_{g_n} \sum_n (\frac{c_n}{w_n} + c_n') g_n \\
s.t.\quad & C_1 : \sum_{n\in\mathcal{N}} g_n \geq D_k, \forall n \in L_k, \\
& C_2 : g_n \leq w_n, \forall n \in L_k, \\
& C_3 : g_n \geq 0, \forall n \in L_k.
\end{aligned} \tag{5}$$

TABLE II
Greedy algorithm for the single SM case

Algorithm: Greedy Algorithm
1: Initialize $g_n = v_n = 0$;
2: **for** n in increasing order of $(\frac{c_n}{w_n} + c_n')$;
3: **while** $D_k \neq 0$
4: $g_n = \min(w_n, D_k)$;
5: $v_n = \frac{g_n}{w_n}$;
6: $D_k = D_k - g_n$;
7: **end while**
8: **end for**
9: **return** (g, v)

We propose a Greedy algorithm for the single SM problem as shown in Table II, by which an optimal solution to (5) can be obtained. The solution generated the greedy algorithm has the following property: at most one of the v_n's is not zero; if it exists, it falls into the interval $0 \leq v_n \leq 1$.

By decomposing (2) into a collection of single SM case that can be solved independently, we can obtain the optimal fractional solution of (2). Then we introduce an efficient rounding technique to yield a feasible solution to (1) with quality guarantee. Stack **x** and **y** into a vector **x** = $\{x_1, x_2, ..., x_N\}$ and a vector **y** = $\{y_{11}, y_{12}, ..., y_{NM}\}$, respectively. Let **(x, y)** be the optimal solution to (2). Denote $F = \{n : x_n > 0\}$ as the set of APs in **(x, y)**, and $F_m = \{n : n \in F, y_{nm} > 0\}$ as the set of APs in F that fractionally serve SM m. Our proposed approximation algorithm can be divided into two step: clustering and rounding.

TABLE III
PROPOSED ALGORITHM FOR APS PLANNING PROBLEM

Step 1: Clustering

Clustering:
while $S \neq \varnothing$ **do**
  for m in increasing order of α_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 $B_m \neq \varnothing$
      $k = m$;
      $N_k = B_m$;
      $C = C \cup \{k\}$;
    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 IV
THE PROPOSED ALGORITHM FOR APS PLANNING PROBLEM

Step 2: Rounding

Rounding:
for each N_k
  $L_k = \{n \in N_k : x_n < 1\}$;
  $D_k = \sum_{n \in L_k} \sum_m y_{nm}$;
  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

Step 1: Clustering. Partition the APs with $x_n > 0$ into clusters and each of them can be centered around an SM. We call it as the cluster center. Denote the cluster centered around SM k as N_k. The cluster N_k consists of SM k, the set of APs assigned to it, and the fractional demands served by these APs, that is, $\sum_{n \in N_k} \sum_m y_{nm}$. At the clustering step, the following two properties are held:

(1) $\sum_{n \in N_k} x_n \geq \frac{1}{2}$;

(2) if the APs in N_k serve SM m, the center k is not too far away from SM m.

Let C be the set of current cluster centers, which is initially empty. For each SM $m \notin C$, B_m represents the set of unclustered APs that are closer to it than to any other clustered center, i.e.,

$$B_m = \{m \in F_m : n \in \bigcup_{k \in C} N_k, c_{nm} \leq \min_{k \in C} c_{nk}\}.$$

To find all cluster centers, $m \in \mathcal{M}$ is ordered in the increasing of α_m and the cluster N_k can be formed by B_m. The procedure of clustering is described in Table III.

Step 2: Rounding. In this step, we decide which AP will be fully opened in each cluster. For each cluster obtained in the first step, it can be tackled as the single SM problem. Therefore, the cluster N_k can be obtained by the greedy algorithm shown in Table II. Then we have $L_k = \{n \in N_k : x_n < 1\}$ and $D_k = \sum_{n \in L_k} \sum_m d_m y_{nm}$, where the tuple of $(g_n^{(k)}, v_n^{(k)})$ is an optimal solution produced by the algorithm in Table II. 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 fixed, each SM is assigned to the closest AP around it.

Lemma 4.1 and 4.2 show that the cost c_{nk} is bounded.

Lemma 4.1. *Consider SM m and AP $n \in A_m$. If AP n is assigned to the cluster N_k, then $c_{nk} \leq \alpha_m$.*

Proof: If SM m is a cluster center, we can get $c_{nm} \leq \alpha_m$ with the analysis of complementary slackness. It can be verified as follows. Denoe $C^{'}$ as the set of cluster centers when SM m is removed from S. Since $n \notin B_m$, we have

$$c_{nm} > \min_{k' \in C'} c_{nk'}.$$

Let $C^{''}$ be the set of cluster centers when AP n is assigned to N_k. So k is the cluster center closest to n among all cluster centers in $C^{''}$, that is:

$$c_{nk} = \min_{k' \in C''} c_{nk'}.$$

Hence,

$$c_{nk} \leq \alpha_m.$$

■

Lemma 4.2. *Consider SM m and AP $n \in F_m \setminus A_m$. If $m \in C$, then $c_{nk} \leq c_{nm}$; otherwise, $c_{nk} \leq c_{nm} + c_{n^*(m)m} + \alpha_m$.*

Proof: If SM m is a cluster center, when it is removed from S, we can construct the cluster N_k that is equal to the current set B_m, which is also $F_m \setminus A_m$. Then we can simply get $c_{nk} \leq c_{nm}$.

Suppose $m \notin C$. Let $k^{'} \in C^{'}$ be the cluster center if AP $n^*(m)$ is removed from B_m, we have

$$c_{n^*(m)k'} < c_{n^*(m)m}.$$

Since

$$\alpha_{k'} \leq \alpha_m,$$

we also have

$$c_{n^*(m)k'} \leq \alpha_m.$$

Using Lemma 4.1., we have

$$\begin{aligned} c_{nk} &= \min_{k' \in C''} c_{nk''} \\ &\le c_{nm} + c_{n^*(m)m} + c_{n^*(m)k'} \\ &\le c_{nm} + c_{n^*(m)m} + \alpha_m . \end{aligned} \tag{6}$$

■

Let $k(n) \in C$ denote the cluster where AP n is assigned, we can obtain that $n \in N_{k(n)}$. $(g_n^{(k)}, v_n^{(k)})$ construct the optimal solution to (9) and the value is O_k^*. Define $R_1 = \{n : x_n < 1\}$ and $R_2 = \{n : x_n = 1\}$, we can obtain the following Lemma 4.3 which shows that the optimal solution to the single SM case problem is comparable to the OPT.

Lemma 4.3. *The optimal value* $O_k^* \le \sum_{n \in L_k} c_n v_n + \sum_m \sum_{n \in L_k} c_{nk} y_{nm}$, *and* $\sum_{k \in C} O_k^* \le \sum_{R_1} c_n v_n + \sum_m \sum_{R_1} c_{nk} y_{nm}$.

Proof: Set

$$\widehat{v}_n = x_n ,$$

and

$$\widehat{g}_n = \sum_m d_m y_{nm} ,$$

for all $n \in L_k$. The AP cost is at most

$$\sum_{n \in L_k} c_n \widehat{v}_n = \sum_{n \in L_k} c_n v_n .$$

The second bound follows from the first since the clusters N_k are disjoint. Then we can prove the lemma. ■

Let $\widehat{x}$ be the $0-1$ vector indicating which AP is open. That is, $\widehat{x}_n = 1$ if n is open, and 0 otherwise. Let $\widehat{x}^{(k)}$ be the portion of $\widehat{x}$ consisting of the APs in L_k, we can show that the cost of the solution is bounded by aggregating the bounds obtained for each partial solution.

Lemma 4.4. $\sum_{n \in L_k} c_n \widehat{x}_n^{(k)} + \sum_m \sum_{n \in L_k} c_{nm} \widehat{y}_{nm}^{(k)}$ *is at most* $O_k^* + 2 \sum_{n \in N_k} c_n x_n + \sum_m \sum_{n \in L_k} c_{nm} y_{nm} + \sum_m \sum_{n \in L_k} c_{nk} y_{nm}$.

Proof: First, it is intuitive that $\sum_{n \in N_k} x_n \ge \sum_{n \in N_k} y_{nk} \ge \frac{1}{2}$ since N_k is generated with this property in the first step. Since all APs have the same cost, the cost of the APs in N_k is

$$c \le c \cdot 2 \sum_{n \in N_k} x_n = 2 \sum_{n \in N_k} c_n x_n .$$

Therefore, the AP cost $\sum_{n \in L_k} c_n \widehat{x}_n^{(k)}$ is at most

$$\sum_{n \in L_k} c_n v_n^{(k)} + 2 \sum_{n \in N_k} c_n x_n .$$

The service cost $\sum_m \sum_{n \in L_k} c_{nm} \widehat{y}_{nm}^{(k)}$ can be

$$\sum_m \sum_{n \in L_k} c_{nk} y_{nm}^{(k)} + \sum_m \sum_{n \in L_k} c_{mk} \widehat{y}_{nm}^{(k)} .$$

So the service cost is

$$\sum_m \sum_{n \in L_k} c_{nm} \widehat{y}_{nm} \le \sum_{n \in L_k} c_n y_n^{(k)} + \sum_m \sum_{n \in L_k} (c_{nm} + c_{nk}) y_{nm} . \tag{7}$$

Recall that $O_k^* = \sum_{n \in L_k} (c_n v_n^{(k)} + c_n g_n^{(k)})$, the lemma is proved. ■

Next, we prove a bound of the opening APs cost and the related connecting cost.

Lemma 4.5. *The sum of opening APs cost and connecting cost is at most* $\sum_m \sum_{R_2} \alpha_m y_{nm} - \sum_n \delta_n$.

Proof: For AP n that $\delta_n > 0$, we have $x_n = 1$. Consider Equation (5-7), we have,

$$\begin{aligned} \sum_m \alpha_m y_{nm} &= \sum_m c_{nm} y_{nm} + \sum_m \beta_{nm} y_{nm} + \sum_m \gamma_n y_{nm} \\ &= \sum_m c_{nm} y_{nm} + \sum_m \beta_{nm} x_n + w_n \gamma_n x_n \\ &= \sum_m c_{nm} y_{nm} + c_n + \delta_n . \end{aligned} \tag{8}$$

$\sum_m \beta_{nm} + w_n \gamma_n = c_n + \delta_n$. So Lemma 4.5 can be proved by sum all n with $x_n = 1$. ■

Finally, we can prove our main results that is shown in the following Lemma 4.6 using these bounds.

Lemma 4.6. *The total cost of the solution returned by Algorithm 2 is at most* $5 \cdot OPT$.

Proof: The proof is referred to Appendix A. ■

IV. Numerical Results and Discussions

As can be observed from Lemma 4.6, our proposed algorithm is approximation algorithm that can provide the worst performance guarantee for the APs planning problem in the SDN-based smart gird. In this section, we verify the efficiency of the algorithm in three representative scenarios: urban, suburban and rural. We compare the performance of the proposed algorithm with other heuristic methods that can be employed to address NP-hard problems with reasonable complexity: GA and TS. We will show that our proposal outperforms others for all considered scenarios. The parameters of the GA and the TS employed in this work is summarized in Table V.

Consider a certain area served by an SDN-based smart grid communication system, where all the SMs and the candidate sites to install the APs are uniformly distributed within an area of 1 by 1 kilometer for urban, suburban and rural environments. The capacity w_n of AP n is distributed uniformly within $(600, 900)$. The traffic demand d_m of SM m is distributed uniformly within $(20, 30)$. The number of SMs, the control center and the candidate sites to install APs are summarized as Table VI for different scenarios. The communication technology in data plane is PLC and the frequency of PLC channel is 15M. The weighting factor g is 1. The attenuation limit Q_a and delay limit Q_b of the links between APs and SMs are 15 and 20, respectively. $a_0 = 6.5 \cdot 10^{-3}$ and $a_1 = 2.46 \cdot 10^{-9}$, which have been proposed in [29]. Without loss of generality, we assume the installation cost of each AP is the same.

Numerical results yielded by our proposed algorithm are given in Fig.2-4, where the black circles represent the candidate sites for installing APs and the red squares are the locations of SMs. Fig.2, Fig.3 and Fig.4 correspond to urban,

TABLE V
PARAMETERS OF GA AND TS

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

TABLE VI
PARAMETERS OF THE SYSTEM IN DIFFERENT SCENARIOS

Scenario	ρ_{SM} $(per km^2)$	N	M_c	SMs per AP
Urban	1000	500	1	34
Suburban	500	500	1	22
Rural	10	500	1	2

suburban and rural areas, respectively. Table VI shows the average number of SMs served by an AP in different scenarios. In Fig.2-4, the sub-figure on the left shows the distribution of SMs and the candidate sites for deploying APs. And the right one shows the selected APs by our proposed algorithm, as well as the links between the SMs and the AP serving them.

Since the SMs are sparse in the rural area, we can observe from Fig.2 that there are on average much fewer SMs served by an AP as compared to the suburban and urban scenarios shown in Fig.3-4, where on average 22 and 34 SMs are connected to one AP. Moreover, from Fig.2-4 we can see that our proposed algorithm can produce load balanced planning results for all considered cased. In other words, our proposed algorithm can fully exploit the capacity potential of each AP, leading to the minimal number of required APs for serving given number of SMs. So we can conclude conservatively that our proposal is cost-efficient.

To verify the promising performance of our proposed algorithm, we also compare it with other representative heuristic methods that can tackle our formulated problem. Fig.5 shows the average costs as a function of the number of SMs, which are produced by our proposed algorithm, the GA and the TS discussed above. The cost of AP n is set to $c_n^{AP} = 9.3$ and the connection cost between AP n and the control center is set to $P = 5$. The connection cost between AP n and SM m for PLC is $c_{nm} = 0.001 dis_{nm}$, where dis_{nm} is the distance between the SM and the serving AP. As can be seen from the Fig.5, our proposed algorithm outperforms the GA and the TS remarkably. The cost of our proposal is about 20% lower than that of the TS. Obviously, when the number of SMs increases, the total deployment cost is also increases. Recall that our proposed algorithm has the worst case performance guarantee which is difficult even impossible for other heuristic methods to achieve such a performance. It is promising for applications in the SDN-based smart grid communications where scalability and reliability are essential prerequisites.

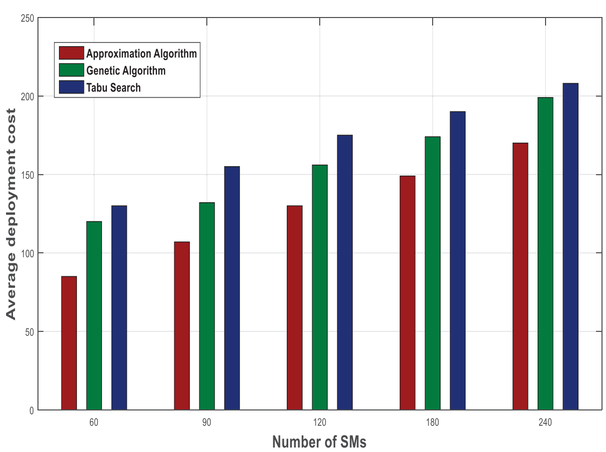


Fig. 5. Average deployment cost as the number of SMs M changes. $N = 500$, $c_n^{AP} = 9.3, \forall n \in \mathcal{N}$, $P = 10$, $c_{nm} = 0.001 dis_{nm}, \forall n \in \mathcal{N}, m \in \mathcal{M}$.

V. CONCLUSION

In this paper, we studied the SDN-based smart grid communications, where the data plane and the control control are decoupled to provide flexible two-way flow of data communications for the smart grid. The APs planning in data plane plays an important role for such an SDN-based smart grid communication network. We formulated a general APs placement problem and introduce a 5-approximation algorithm to address it. Our proposed algorithm can yield solutions with the worst case performance guarantee. Moreover, it outperforms other representative heuristic methods for all considered scenarios as verified by numerical results. In future work, hybrid networking model should be investigated for the communication links between APs and SMs, e.g., both wired and wireless technology can be selected for these links in practical environment, depending on the density of SMs, traffic requirements of SMs, as well as the installation and connection cost between APs and SMs.

APPENDIX A
PROOF OF THEOREM 4.6

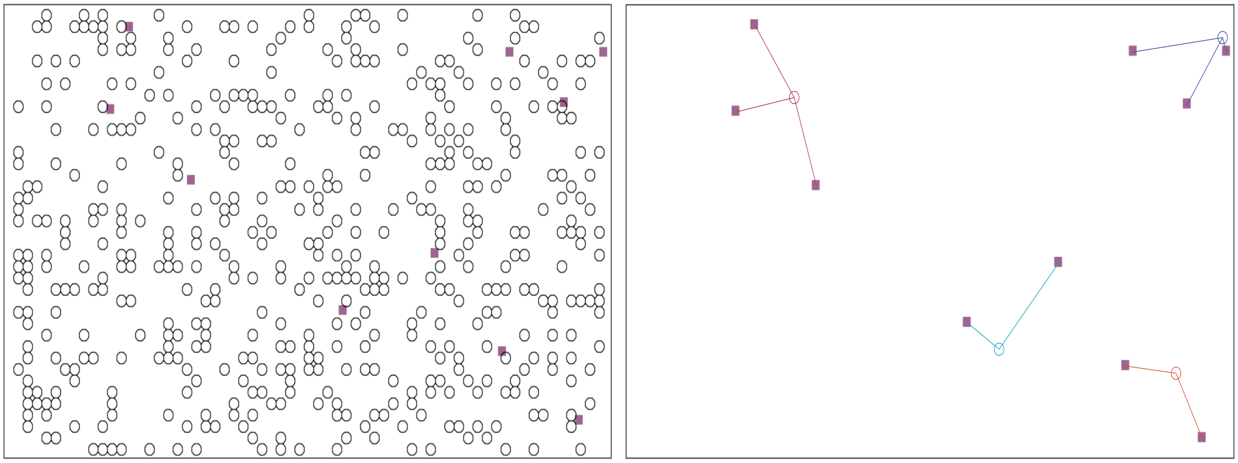
To bound the total cost, it suffices to give a fractional assignment $\widehat{z}_{nm}$ such that $\widehat{x}, \widehat{y}$ are feasible solutions to (2) and have the cost at most $5 \cdot OPT$. Set $\widehat{y}_{nm} = y_{nm}$ for every AP n with $x_n = 1 = \widehat{x}_n$. Then

$$\sum_m \sum_{R_2} c_n \widehat{x}_n + \sum_m \sum_{R_2} c_{nm} \widehat{y}_{nm} = \sum_m \sum_{R_2} \alpha_m y_{nm} - \sum_n \delta_n. \quad (9)$$

$(\widehat{x}, \widehat{y})$ satisfies $\sum_m \widehat{y}_{nm} = \sum_{R_1} y_{nm} + \sum_{R_2} z_{nm}$. Since the clusters N_k are disjoint, by Lemma 4.5 we have,

$$\begin{aligned} &\sum_{R_1} c_n \widehat{x}_n + \sum_m \sum_{R_1} c_{nm} \widehat{y}_{nm} \\ &\leq \sum_{k \in C} O_k^* + 2 \sum_m \sum_{R_1} c_{nm} y_{nm} + \sum_m \sum_{R_1} c_{nm} y_{nm} \\ &\leq 3 \sum_n c_n x_n + \sum_m \sum_{R_1} c_{nm} y_{nm} + 2 \sum_m \sum_{R_1} c_{nk(n)} y_{nm}. \end{aligned} \quad (10)$$

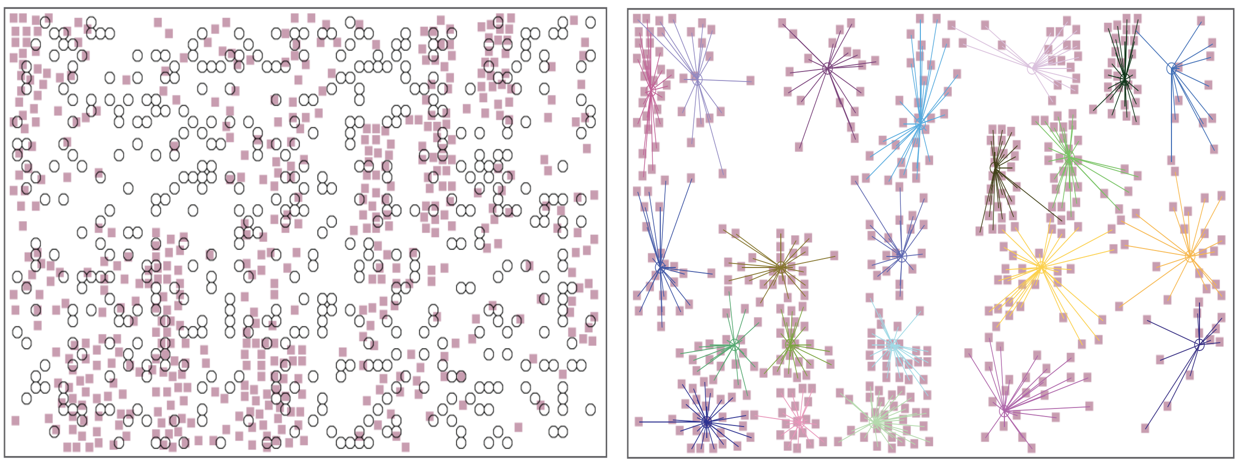
For any SM m and AP $n \in F_m$, if $n \in A_m$, then we have $c_{nk(n)} \leq \alpha_m$; otherwise, $c_{nk} \leq c_{nm} + \alpha_m$ for $m \in C$, and



(a) Rural area

(b) Selected APs and the connected SMs in rural area

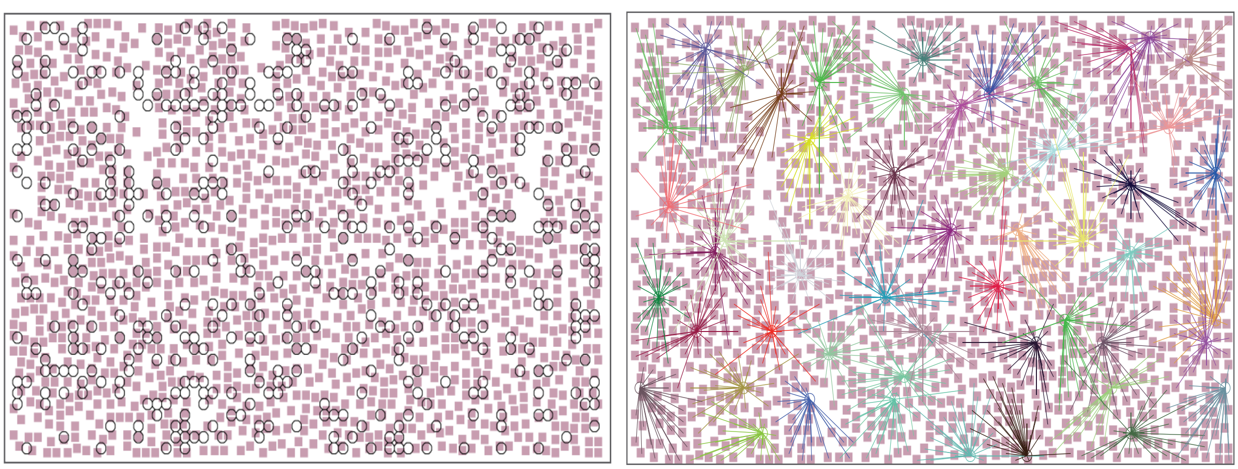
Fig. 2. APs placement for rural area



(a) Suburban area

(b) Selected APs and the connected SMs in suburban area

Fig. 3. APs placement for suburban area



(a) Urban area

(b) Selected APs and the connected SMs in urban area

Fig. 4. APs placement for urban area

$c_{nk} \leq c_{nm} + c_{n^*(m)m} + \alpha_m$ for $m \notin C$. Plugging this in the above expression,

$$\begin{aligned} &\sum_{R_1} c_n \widehat{x}_n + \sum_m \sum_{R_1} c_{nm} \widehat{y}_{nm} \\ &\quad \leq 3 \sum_n c_n x_n + \sum_m \sum_{R_1} c_{nm} y_{nm} + 2 \sum_m \sum_{R_1} \alpha_m y_{nm} \\ &\quad + 2 \sum_m \sum_{n: x_n < 1} c_{nm} y_{nm} + \sum_{m \notin C} 2 c_{n^*(m)m} \sum_{R_1} y_{nm}. \end{aligned} \tag{11}$$

For $m \notin C$, denote $n \in A_m$ as R_3,

$$c_{n^*(m)m} = \min_{R_3} c_{nm} \leq \frac{\sum_{R_3} c_{nm} y_{nm}}{\sum_{R_3} y_{nm}} < 2 \sum_{R_3} c_{nm} y_{nm}. \tag{12}$$

This implies that

$$\begin{aligned} &\sum_{R_1} c_n \widehat{x}_n + \sum_m \sum_{R_1} c_{nm} \widehat{y}_{nm} \\ &\quad \leq 3 \sum_n c_n x_n + \sum_m \sum_{R_2} c_{nm} y_{nm} + 2 \sum_m \sum_{R_2} \alpha_m y_{nm} \\ &\quad + 2 \sum_m \sum_{R_1} c_{nm} y_{nm} + 2 \sum_{m \notin C} \sum_{R_3} c_{nm} y_{nm} \\ &\quad \leq 2 \sum_m \sum_{R_1} \alpha_m y_{nm} + 3 (\sum_n c_n x_n + \sum_m \sum_n c_{nm} y_{nm}). \end{aligned} \tag{13}$$

Combine (12) and (13), we obtain that

$$\begin{aligned} C_{total} \quad &\leq (\sum_m \sum_{R_2} \alpha_m y_{nm} - \sum_n \delta_n) + 2 \sum_m \sum_{R_2} \alpha_m y_{nm} \\ &+ 3 (\sum_n c_n x_n + \sum_m \sum_n c_{nm} y_{nm}) \\ &\leq 2 (\sum_m \sum_{R_2} \alpha_m y_{nm} - \sum_n \delta_n \\ &+ \sum_m \sum_{R_1} \alpha_m y_{nm}) + 3 \cdot OPT_1 \\ &= 5 \cdot OPT_1. \end{aligned} \tag{14}$$

Acknowledgement

This work was partially supported by JiangsuSF (BK20151389), the Fundamental Research Funds for the Central Universities (021014380013) and the open research fund of National Mobile Communications Research Laboratory, Southeast University (2016D08).

References

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

[2] Z. Fan, P. Kulkarni, S. Gormus, and Efthymiou, "Smart grid communications: Overview of research challenges, solutions, and standardization activities," *IEEE Commun. Surv. Tut.*, vol. 15, no. 1, pp. 21–38, Jan. 2013.

[3] M. H. U. Ahmed, M. G. R. Alam, R. Kamal, C. S. Hong, and S. Lee, "Smart grid cooperative communication with smart relay," *IEEE Comm. and Netw.*, vol. 14, no. 6, pp. 640–652, Dec. 2012.

[4] A. Gelberger, N. Yemini, and R. Giladi, "Performance analysis of software-defined networking (SDN)," in *Proc. IEEE MASCOTS'13*, Aug. 2013, pp. 389–393.

[5] T.-C. Yen and C.-S. Su, "An SDN-based cloud computing architecture and its mathematical model," in *Proc. ISEEE'14*, vol. 3, Apr. 2014, pp. 1728–1731.

[6] P. Xiao, W. Qu, H. Qi, Z. Li, and Y. Xu, "The SDN controller placement problem for WAN," in *Proc. IEEE ICC'14*, Oct. 2014, pp. 220–224.

[7] M. Obadia, M. Bouet, J.-L. Rougier, and L. Iannone, "A greedy approach for minimizing SDN control overhead," in *Proc. IEEE NetSoft'15*, Apr. 2015, pp. 1–5.

[8] Y. Jimenez, C. Cervello-Pastor, and A. Garcia, "On the controller placement for designing a distributed SDN control layer," in *Proc. IFIP Networking Conference'14*, June 2014, pp. 1–9.

[9] H. Farhad, H. Lee, and A. Nakao, "Data plane programmability in SDN," in *Proc. IEEE ICNP'14*, Oct. 2014, pp. 583–588.

[10] A. Cahn, J. Hoyos, M. Hulse, and E. Keller, "Software-defined energy communication networks: From substation automation to future smart grids," in *Proc. IEEE SmartGridComm'13*, Oct. 2013, pp. 558–563.

[11] A. Goodney, S. Kumar, A. Ravi, and Y. H. Cho, "Efficient pmu networking with software defined networks," in *Proc. IEEE SmartGridComm'13*, Oct. 2013, pp. 378–383.

[12] J. Zhang, B.-C. Seet, T.-T. Lie, and C. H. Foh, "Opportunities for software-defined networking in smart grid," in *Proc. ICICS'2013*, Dec. 2013, pp. 1–5.

[13] N. Dorsch, F. Kurtz, H. Georg, C. Hagerling, and C. Wietfeld, "Software-defined networking for smart grid communications: Applications, challenges and advantages," in *Proc. IEEE SmartGridComm'14*, Nov. 2014, pp. 422–427.

[14] M. Ringwelski, C. Renner, A. Reinhardt, A. Weigel, and V. Turau, "The hitchhiker's guide to choosing the compression algorithm for your smart meter data," in *Proc. IEEE ENERGYCON'12*, Sep. 2012, pp. 935–940.

[15] R. Yu, Y. Zhang, and Y. Chen, "Hybrid spectrum access in cognitive neighborhood area networks in the smart grid," in *Proc. IEEE WCNC'12*, Apr. 2012, pp. 1478–1483.

[16] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 53–59, Apr. 2011.

[17] Z. Fadlullah, M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 60–65, Apr. 2011.

[18] X. Huang and S. Wang, "Aggregation points planning in smart grid communication system," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1315–1318, Aug. 2015.

[19] Z. Lu and Y. Wen, "Distributed algorithm for tree-structured data aggregation service placement in smart grid," *IEEE Syst. J.*, vol. 8, no. 2, pp. 553–561, June 2014.

[20] S. Lange, S. Gebert, T. Zinner, P. Tran-Gia, D. Hock, M. Jarschel, and M. Hoffmann, "Optimal model for the controller placement problem in software defined networks," *IEEE Trans. Net. Serv. Manag.*, vol. 12, no. 1, pp. 4–17, Mar. 2015.

[21] W.-L. Wang and Q. Wu, "Relay placement problem in smart grid deployment," in *Proc. LNCS ISoLA'10*, Oct. 2010, pp. 411–424.

[22] F. Aalamifar, G. Shirazi, M. Noori, and L. Lampe, "Cost-efficient data aggregation point placement for advanced metering infrastructure," in *Proc. IEEE SmartGridComm'14*, Nov. 2014, pp. 344–349.

[23] K. Jain and V. V. Vazirani, "Primal-dual approximation algorithms for metric facility location and k-median problems," in *Proc. IEEE FOCS'99*, pp. 2–13.

[24] 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.

[25] 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.

[26] 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.

[27] I. Stievano, F. Canavero, W. Garcia Valverde, L. Guerrieri, and P. Bisaglia, "Multipath modeling of automotive power line communication channels," *IEEE Trans. Indus. Inform.*, vol. 10, no. 2, pp. 1381–1391, May 2014.

[28] G. Bumiller, L. Lampe, and H. Hrasnica, "Power line communication networks for large-scale control and automation systems," *IEEE Commun. Mag.*, vol. 48, no. 4, pp. 106–113, Apr. 2010.

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

[30] O. Hooijen, "A channel model for the residential power circuit used as a digital communications medium," *IEEE Trans. Electrom. Comp.*, vol. 40, no. 4, pp. 331–336, Nov. 1998.