Wireless Networks
https://doi.org/10.1007/s11276-019-02152-x

Cost-efficient approximation algorithm for aggregation points planning in smart grid communications

Yue Li[1] · Tianyu Wang[1] · Shaowei Wang[1]

Abstract
Smart grid is in need of an efficient communication network to guarantee reliable two-way data transmission between the control center and smart meters (SMs). In this work, a software-defined networking (SDN) based smart grid communication (SGC) scheme is introduced to fulfill the information transmission requirement, where the control plane is separated from the data plane to support diverse services flexibly in the smart grid. In such an SDN-based SGC system, to guarantee effective data processing and forwarding between the SMs and the control center, aggregation points (APs) are introduced. These APs should be deployed in an optimal way so as to cut down the total capital expenditure of the SGC system. The total cost generally includes the transmission cost between APs and the control center as well as APs and SMs. The construction and maintenance cost of the APs is also included. An approximation algorithm is introduced in this paper. The algorithm can deal with the formulated intractable APs planning task and produce performance-guaranteed solutions with reasonable complexity. Experiments indicate that the proposed algorithm works well for geographical areas with different densities of SMs. Our proposal yields cost-efficient APs deployment scheme and sheds insight into the reduction of the capital expenditure of the SGC system.

Keywords Aggregation point · Approximation algorithm · Smart grid communications · Software defined network

1 Introduction

As smart grid is regarded as a revolutionary technology for the next generation power system with intelligent operations on power management, smart grid communication (SGC) has been calling much attention because it is the prerequisite to realize the smart grid [1]. An SGC provides real-time data exchange between suppliers and consumers and sustains data flows between smart components and sophisticated computing resources to satisfy the increasing demands of power generation, transmission and distribution for a metropolis even a whole nation. Consumers are provided with various services via the SGC, such as integrating renewable and alternative energy resources [2]. On the other hand, suppliers can carry out load management and real-time price broadcast to increase their profit via the SGC. Moreover, power disturbance and outage that may incur disastrous consequences can be largely avoided [3] with a reliable SGC. Overall, the SGC is an indispensable part to reform the conventional power system toward an intelligent one.

Since the SGC generally needs to support services with different performance requirements such as latency and traffic load from the viewpoint of both the consumers and the suppliers, an agile network infrastructure is required [4–6], where software defined networking (SDN) is a promising choice to construct such a flexible communication system. Different from traditional vertically integrated networks which introduce challenges such as inconsistent interfaces and frequent handovers, separating control plane from data plane makes the switches in an SDN only under the instruction of a programmable control center [7]. Thus the control plane is able to achieve seamless handover and detect real-time failure in this way,

✉ Shaowei Wang
wangsw@nju.edu.cn

Yue Li
njuliyue@smail.nju.edu.cn

Tianyu Wang
tianyu.alex.wang@nju.edu.cn

[1] Nanjing University, Nanjing, China

Published online: 11 October 2019

as well as to allocate radio resources intelligently and efficiently. For instance, a flexible SDN orchestration can substitute for current static resource allocation in small cell network to increase network throughput remarkably [8]; and an SDN-based phasor measurement unit can adapt different data rates in the smart grid to achieve lower latency as compared to conventional approaches.

For an SDN-based SGC system, the smart meters (SMs) which are connected to in-home appliances gather energy-consumption-related information in the data plane and send them to the control center for further processing and analysis. The control center responses to the demand from the consumers and broadcast price information to consumers, as well as perform load management among the whole grid to maintain system stability. Since there are a large number of SMs and these SMs are generally far away from the control center, deploying dedicated links between the SMs and the control plane to transfer data flow is costly and unacceptable for a commercial power provider. On the other hand, it is also infeasible to adopt existing powerline directly to perform these data transmission due to the attenuation and relay limit of powerline. So aggregation points (APs) that serve certain number of surrounding SMs are essential for the smart grid [9]. These APs can be taken as transition points that perform information transmission between the control center and the SMs. For example, demand response and power quality measurement generated by SMs can be sent through these APs. On the other hand, the control information generated by the control center is sent to the SMs through another way of the data flow also via the APs, such as load management information, which can greatly improve the demand flexibility of the smart grid. Moreover, the APs can also compress the raw data so as to improve the transmission efficiency. The compression ratio depends on the correlation of the data [10], which usually goes between 1/2 and 2/3. Obviously, it is of great significance to plan the APs optimally for a reliable and cost-efficient SDN-based SGC system. This is the motivation of our work.

Considering the efficiency and reliability requirements of the SGC, the APs should be designed carefully to minimize the deploying cost while providing QoS-guaranteed services for consumers and suppliers. The deployment of APs includes the selection of their locations as well as the assignment of their serving SMs. As power line communication (PLC) is usually available in the legacy grid and suitable for short-distance communications [11], it is reasonable to serve as connections between SMs and their associated APs, which can lower the total cost of the huge number of links. Fiber optic communication is recommended for connections between APs and the control plane due to its high data speeds and low signal losses, which can satisfy the needs of real-time massive data exchange and long-distance data transmission between them. In the following section, we will discuss the connection costs in detail. Our optimization task is to find out the minimal deployment cost for the SGC. The deployment cost consists of the opening cost of the APs, the connection cost between the APs and the SMs, and the connection cost between the APs and the control center.

We first discuss some related works briefly. A large SDN network consists of multiple controllers is investigated in [12], where multiple controllers are introduced to exploit the benefits of SDN and the controller placement problem is addressed by a so-called k-Critical algorithm. Lange et al. [13] also addresses a controller placement problem. It aims at choosing the proper controllers and minimizing the cost of assigning all the switches to them. Sallahi and St-Hilaire [14] provides a re-organizing resolution when a new network design is introduced or new switches are added in a controller planning problem. In [15], an asynchronous distributed algorithm is presented to address the data aggregation services placement problem in the smart grid. The scheme aims at finding the optimal aggregation points placement strategy with the minimal deployment cost. In [16], a data aggregation points placement problem is investigated. It leads to an integer program problem for advanced metering infrastructure and a K-means algorithm is introduced to minimize the total cost.

In this work, we introduce an SDN-based SGC system, which exploits the potentials of SDN to realize a reliable and cost-efficient two-way data transmission for the smart grid network. The network framework introduced in this paper is more practical as compared to the model investigated in [17], which uses optical fiber cable as links between APs and SMs. As mentioned above, it is expensive and unacceptable for a commercial power grid. Notice that wireless links between APs and SMs are also suggested in the literature [18], however, licensed radio spectrum is always scarce and unavailable, as discussed in [19]. Unfortunately, the formulated APs planning problem is also NP-hard. We introduce an efficient approximation algorithm to deal with the optimization task. The algorithm show great potentials for practical applications. Numerical results also show that our algorithm is performance-guaranteed and robust in different scenarios. Moreover, it yields lower deployment cost than the state-of-the-art one.

The remaining of the paper is as follows. System model is presented in Sect. 2, as well as problem formulation. The local search-based approximation algorithm is introduced in Sect. 3. Numerical results and discussions are given in Sect. 4. The conclusion and future researches are given in Sect. 5.

2 System model and problem formulation

As shown in Fig. 1, $\mathcal{R}$ is the area of interest served by an SDN-based SGC system. Let $\mathcal{M} = \{1, 2, \ldots, M\}$ be the set of SMs contained in this area. M_c is the set of control centers in $\mathcal{R}$. Without loss of generality, we consider $M_c = 1$ in this paper. A number of APs need to be installed in $\mathcal{R}$ to serve these SMs. To be more specific, an SM should be associated with only one AP, while an AP serves a cluster of its surrounding SMs. Denote $\mathcal{N} = \{1, 2, \ldots, N\}$ as the candidate set of APs that can be chosen to install. c_{nm} is the PLC connection cost between an SM and its connected AP. Fiber optic is employed to provide links between the APs and the control center. The cost of fiber optic is f_n. Denote r as the radius of $\mathcal{R}$ and ρ_{SM} as the density of SMs in the area [16], the number of SMs can be calculated as $M = \rho_{SM} r^2$.

2.1 Channel model

As discussed above, PLC can transmit data between APs and SMs using the existing power lines, which is of low cost. The channel characteristics of PLC involve transfer function H_f and additive noise N_f. The transfer function H_f can be obtained as follows:

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

g is a weighting factor which is always equal or less than one, that is, $g \leq 1$. It can represent the production of transmission and refection factor long the power line.

$A(f, dis_{nm})$ is the attenuation of powerline cable and can be calculated as follows:

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

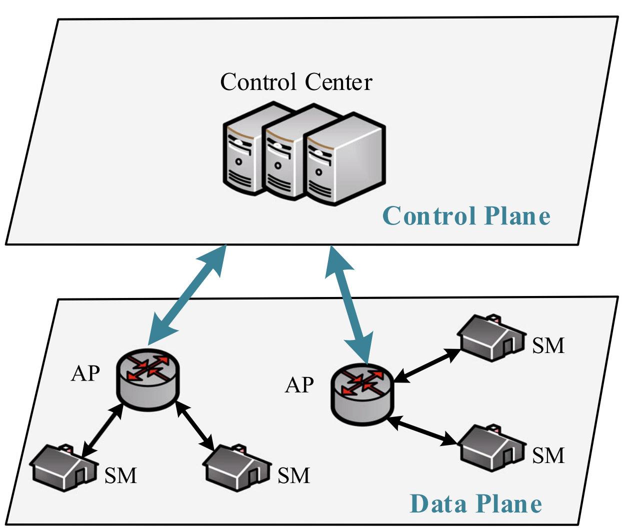

Fig. 1 Architecture of SDN-based SGC system

where f is the frequency, dis_{nm} is the distance between the nth AP and the mth SM. a_0 and a_1 are cable parameters. Detailed discussions of them can be found in [20].

$B(f, dis_{nm})$ is the delay portion. The delay factor τ is described as

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

where v_0 represents light speed and ε_r is a dielectric constant that depends on insulating material. We can write the delay portion $B(f, dis_{nm})$ into

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

Then we can get the transfer function H_f:

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

N_f is the PLC background noise that can be calculated as

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

where we can find that the power spectral density of which decreases with frequency f. The value of K depends on the receiver or transmitter's location. In our system model, for a given frequency, once an AP is selected, the value of K is fixed, as well as the PLC channel noise.

2.2 Mathematical model

Our APs planning optimization task focuses on minimizing the sum cost of deploying APs while meeting the traffic requirements of all SMs. Assuming that c_n^{AP} is the installation and maintenance cost of each AP, and c_{nm} refers to the cost to link SM m and its associated AP n. The cost to connect this AP and the control center is denoted by f_n. c_{nm} and f_n are both related to the transmission distances dis_{nm} and dis_n, which can be calculated as: $c_{nm} = \phi(dis_{nm}), f_n = \psi(dis_n)$. Here ϕ and ψ are linear functions. dis_{nm} is the distance between AP n and SM m. dis_n is the distance between the control center and the AP n. In this paper, the link cost is considered as a linear function to distance, which is reasonable and can simplify analysis.

Let x_n be index variable that indicates AP n is open or not, and y_{nm} shows whether SM m is connected with AP n or not.

$$x_n = \begin{cases} 1 & \text{AP } n \text{ is open;} \\ 0 & \text{otherwise.} \end{cases}$$

$$y_{nm} = \begin{cases} 1 & \text{Assign SM } m \text{ to AP } n; \\ 0 & \text{otherwise.} \end{cases}$$

Denote d_m as the traffic requirement of SM $m \in \mathcal{M}$ and w_n is the max capacity of each candidate AP $n \in \mathcal{N}$. In our system, all the traffic demands should be satisfied while selected APs should not exceed their capacity limitation.

Taking account of all these constraints, We now summarize our APs planning problem mathematically in (1), where C_i refers to constraint i.

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

where

$Q = Q_a \cdot Q_b.$

Q_a is the PLC attenuation and Q_b is the PLC delay limit. The object function consists of the APs installation and maintenance cost, the cost of the links between APs and SMs, and the cost of the fiber optic link between APs and control center. C_1 ensures that the traffic demands covered by a selected AP will not exceed its capacity limitation. C_2 explains the attenuation and relay requirement between the SMs and the APs. C_3 means that an AP should be open if it serves any of the SMs. C_4 ensures that it should be served by only one AP for any SM.

3 Local search based approximation algorithm

The formulated mathematical model for APs planning leads to an integer programming problem. Generally, it is hard to work out the global optimal solution. More specifically, it formulates an unsplittable demands capacitated facility location problem [21]. The integer constraints in (1) bring us exponential complexity if calculated by brute search. These kind of problems can be solved by intuitive methods such as tabu search and genetic algorithm. The limitation lies in the absence of provable gap between the optimal solution and the produced ones. In this paper, we introduce a local search (LS) algorithm that can provide an approximation ratio of $(8 + \varepsilon)$ to deal with the problem. Numerical shows that it works quite well for all considered scenarios.

The intuitiveness of our proposed LS-based algorithm is as follows: First, let $\mathcal{S}$ ($\mathcal{S} \subseteq \mathcal{N}$) be an initial solution to Eq. (1); then repeatedly updating $\mathcal{S}$ based on local search procedure to reach (near) optimal solution. Given an initial solution $\mathcal{S}$, the integer programming problem described in (1) can be transformed as follows:

$$\begin{aligned} \min_{y_{nm}} \quad & \sum_{n \in \mathcal{S}} c_n + \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{S}} y_{nm} c_{nm} \\ & C_1 : \sum_{m \in \mathcal{M}} d_m y_{nm} \leq w_n, \quad \forall n \in \mathcal{S}, \\ & C_2 : | H_f | \geq Q \times y_{nm}, \quad \forall n \in \mathcal{S}, \quad \forall m \in \mathcal{M}, \\ & C_3 : \sum_{n \in \mathcal{S}} y_{nm} \geq 1, \quad \forall m \in \mathcal{M}, \\ & C_4 : y_{nm} \in \{0, 1\}, \quad \forall n \in \mathcal{S}, \quad \forall m \in \mathcal{M} \end{aligned} \tag{2}$$

We can have the relaxation form of (2) as follows:

$$\begin{aligned} \min_{y_{nm}} \quad & \sum_{n \in \mathcal{S}} c_n + \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{S}} y_{nm} c_{nm} \\ s.t. \quad & C1 \sim C3, \\ & C_4 : 0 \leq y_{nm} \leq 1, \quad \forall m \in \mathcal{M}, \quad n \in \mathcal{S}. \end{aligned} \tag{3}$$

Eq. (3) defines a linear programming problem. Some developed algorithms or toolboxes, such as CVX [22], can be employed to solve it efficiently. Then the feasible integer solution can be obtained via rounding procedure as follows:

$$y_{nm} = \begin{cases} 1 & n = \arg\max_{m^* \in \mathcal{M}} y_{nm^*}, \\ 0 & \text{otherwise}, \end{cases} \quad \forall n \in \mathcal{N}, \tag{4}$$

where y_{nm^*} is the solution to (3). Here we calculate the deploying cost for the current set $\mathcal{S}$, which is denoted by $c(\mathcal{S})$. The objective is to minimize the $c(\mathcal{S})$ with a certain set $\mathcal{S} \subseteq \mathcal{N}$.

The following steps are taken to update $\mathcal{S}$: Given an initial solution $\mathcal{S}$ to Eq. (2), an *improvement operation* can yield a cost reduction of $c(\mathcal{S})/p(n)$ at least. Denote $p(n)$ as a fixed polynomial in n that is large enough; keep on updating $\mathcal{S}$ through *improvement operation* until the cost can no longer be decreased. The choice of $p(n)$ ensures that we will arrive at a solution with fixed approximation factor after $p(n)log(C(S_0)/C(S^*))$ *operations*. Since each *operation* takes polynomial time. We can say that the whole algorithm runs in polynomial time. The *operation* is described as follows:

Initialization: Take $\mathcal{S} = \mathcal{N}$ as the initial solution, which can make sure that Eq. (3) is solvable. Then we use the following *operation* to update $\mathcal{S}$:

Adding: Open an AP $n \in \mathcal{N} \backslash \mathcal{S}$. Assign the demand and calculate the cost reduction that can be calculated as $c(\mathcal{S}) - c(\mathcal{S} \cup \{n\})$. If the cost reduction is more than $c(\mathcal{S})/p(n)$, set $\mathcal{S} \leftarrow \mathcal{S} + \{n\}$;

Dropping: Close an AP $n \in \mathcal{N}$. Reassign the demand and calculate the cost reduction which is defined as $c(\mathcal{S}) - c(\mathcal{S} \backslash \{n\})$. If the cost reduction is more than $c(\mathcal{S})/p(n)$, set $\mathcal{S} \leftarrow \mathcal{S} - \{n\}$;

Swapping: Swap an AP $n \in \mathcal{N}$ with an AP $n \in \mathcal{N} \backslash \mathcal{S}$. Reassign the demand and calculate the cost reduction which is defined as $c(\mathcal{S}) - c(\mathcal{S} \backslash \{n\} \cup \{n'\})$. If the cost reduction is more than $c(\mathcal{S})/p(n)$, set $\mathcal{S} \leftarrow \mathcal{S} - \{n\} + \{n'\}$, where n' is a close AP in $\mathcal{N} \backslash \mathcal{S}$;

The solving procedure can be find in Table 1.

Corollary 1 *If the expansion of facility capacity does not lead to an increase of the facility cost, the LS-based algorithm has an $(8+\varepsilon)$-approximate ratio $(\varepsilon > 0)$ and can yield a solution in polynomial time.*

Proof The detailed proof is deferred to "Appendix". □

4 Numerical results

As analyzed above, we propose a worst-case performance-guaranteed algorithm to address the APs planning problem. We will then evaluate the performance of the proposed algorithm via extensive numerical experiments for a practical SDN-based SGC system. We consider geographical areas with different density of SMs and discuss the factors that influence the capital expenditure of the SGC system.

Simulation parameters are given as follows. SMs and candidate APs are distributed randomly in an ares of interest $\mathcal{R}$, which is 1×1 km^2. The channel of PLC between SMs and their serving APs occupies a frequency of 15 M. We set $Q_a = 15$ and $Q_b = 20$. The weighting factor $g = 1$. $a_0 = 6.5 \cdot 10^{-3}$ and $a_1 = 2.46 \cdot 10^{-9}$, which have been proposed in [20]. The capacity w_n of AP n and the traffic demand d_m of SM m are uniformly distributed and can be expressed as: $w_n = w + 100\xi_n, d_m = d + 10\xi_m$, where w and d are set as $w = 500$ and $d = 20$. ξ_n and ξ_m are uniformly distributed within (0,1). The connection cost is $c_{nm} = 0.001 dis_{nm}$, where dis_{nm} is the distance between SMs and associated APs. To simplify analysis, we set the installation cost c_n^{AP} and the link cost f_n as constants.

Figures 2, 3 and 4 show the APs planning results in three scenarios with different densities of SMs: rural, suburban and urban areas. The dotted lines indicate the links between the SMs and its associated APs. Table 2 gives the total deployment cost and the number of selected APs produced by the algorithm. We compared it with an LP-based algorithm proposed in [23]. Obviously, in suburban and urban areas with more clients, the LS-based algorithm tends to select less APs and leads to lower capital cost. Overall, we can find that an SM tends to associate with the almost nearest AP which has been opened in all scenarios. It is reasonable since shorter distance leads to lower connection cost between SMs and APs. From Figs. 2, 3 and 4, we find that an AP averagely serves much fewer SMs in the rural areas than in the suburban and urban areas. It can be explained as follows: The SMs are sparsely distributed in rural areas so the connection costs between APs and SMs govern the total cost. In other words, to serve a given number of SMs, more APs are required in rural areas compared to urban and suburban areas to reduce the total connection cost between SMs and APs. When the density of SMs becomes higher, e.g., for the cases in the urban or suburban areas, each AP should serve as many as possible its surrounding SMs so that the total cost can be minimized by opening APs as few as possible, where the capacity limitation of APs holds dominant position at this time. Moreover, our proposed algorithm is robust and can yield balanced load among the opened APs as can be seen in Figs. 2, 3 and 4.

Figures 5 and 6 give the number of selected APs as the functions of the cost of APs and the capacity of APs, where there are 100 SMs with 50 candidate APs and $c_n = 18$. From Fig. 5 we can find out that the total required APs can

Table 1 The solving procedure

Procedure
Initialization: Set initial solution $\mathcal{S}$ $(\mathcal{S} = \mathcal{N})$
Calculate $c(\mathcal{S})$:
1: Get relaxation form of (2) and calculate y_{nm^*}
2: round y_{nm^*}, calculate y_{nm} and $c(\mathcal{S})$
Update $\mathcal{S}$:
3: **Repeat** Renew set $\mathcal{S}$
4: **if** problem (3) can be solved
5: **if** *operation* is *good*
6: update $\mathcal{S}$
7: **end if**
8: **else**
9: renew $\mathcal{S}$, back to step 4
10: **end if**
11: **Until** none of *operation* is *good*

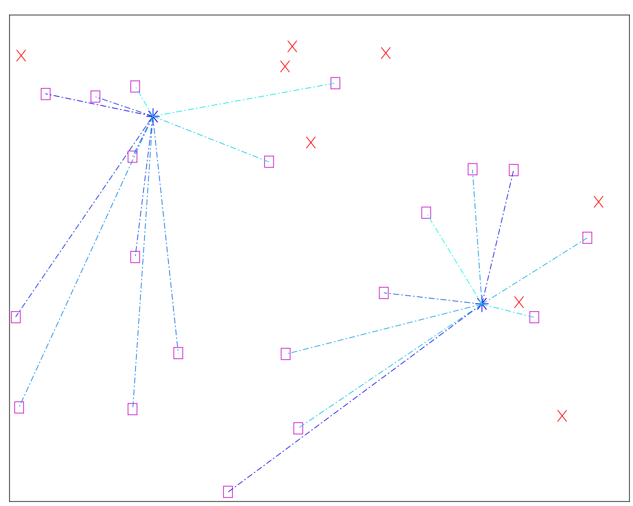

Fig. 2 APs placement of rural area: AP = 10, SM = 20

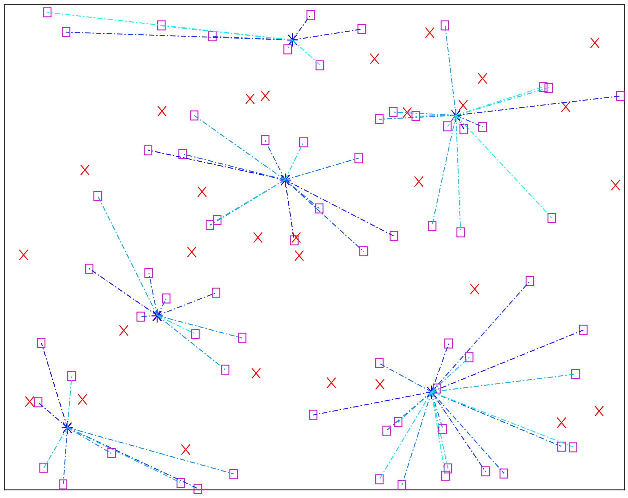

Fig. 3 APs placement of suburban area: AP = 35, SM = 70

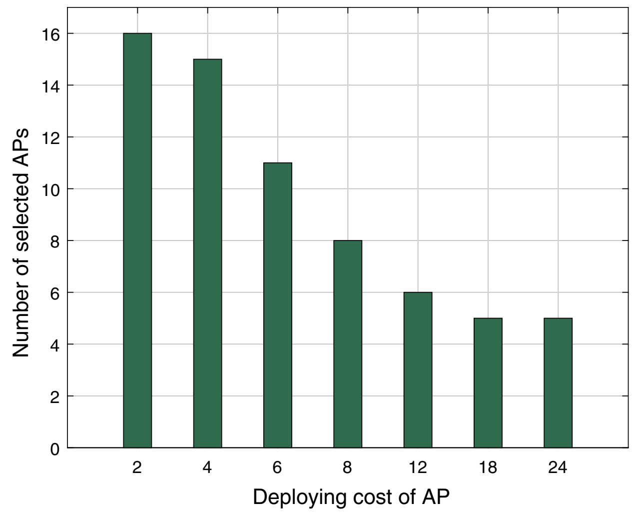

Fig. 5 Total selected APs with different cost of deploying AP

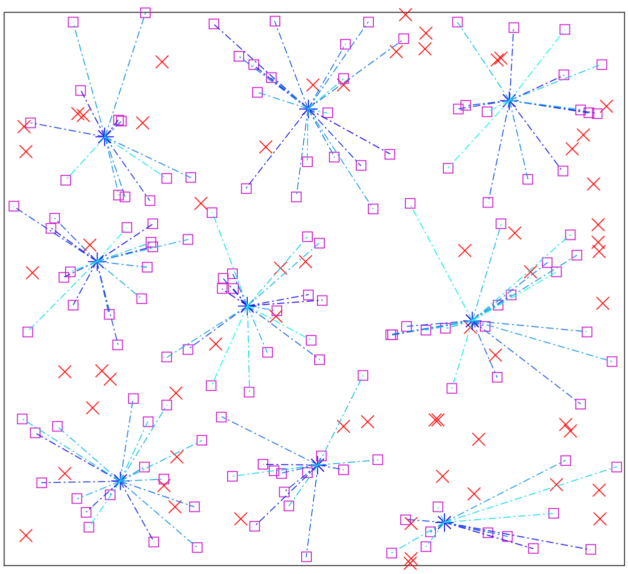

Fig. 4 APs placement of urban area: AP = 70, SM = 140

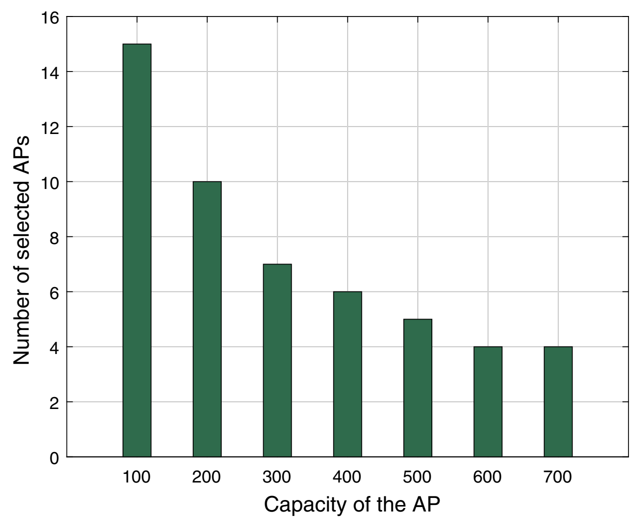

Fig. 6 Total selected APs with different capacity of AP

be cut down if the installation cost of APs increases. The reason is that the increasing cost of APs becomes the dominant part of the total expenditure of the SGC system. Figure 6 shows the total selected APs decreases as increasing of the capacity of APs. It is worthwhile to note that the number of selected APs does not keep decreasing when the capacity of APs is large enough as shown in Fig. 6, where the number of selected APs is the same for the cases that $w_n = 600$ and $w_n = 700$. Recall that fewer APs always result in the increase of the link cost between SM and APs for a given distribution of the SMs and APs. We can conclude that it is not necessary to equip AP too much capacity in practice.

Since the LS-based algorithm is an iterative one, we also investigate the convergence of the algorithm. Figure 7

Table 2 Comparison of two algorithms in different scenarios

Scenario	Total cost		Number of selected APs	
	LS-based	LP-based	LS-based	LP-based
Rural: AP = 10, SM = 20	7.9529	7.9529	2	2
Suburban: AP = 35, SM = 70	16.6887	17.5730	6	7
Urban: AP = 70, SM = 140	26.1748	26.3255	9	10

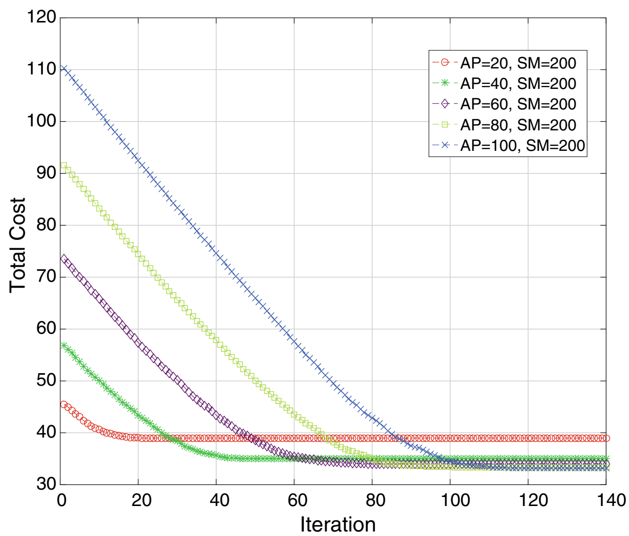

Fig. 7 Total cost during each iteration

shows the number of iterations for convergence in different scenarios. We can see more iterations are required if the number of candidate sites for installing APs becomes large. It is interesting to note that the total cost keeps decreasing when the number of candidate sites is increasing. It can be explained that the increasing of candidate sites for installing APs leads to more promising choices for the SMs to associate with the APs with relatively less link cost.

5 Conclusion

In this work, we developed an SDN-based smart grid communications system where aggregation points (APs) are introduced for information exchange between the smart meters and the control center. These APs should be carefully planned to minimize the total expenditure of our SGC system. We try to meet the traffic requirements of all the SMs in the area of interest with the minimal total expenditure. The formulated optimization task falls into an NP-hard facility location problem. we introduced an LS-based $(8+\epsilon)$-approximation algorithm to handle it efficiently. It is performance-guaranteed and numerical results show that they are promising to be applied to an SDN-based SGC system. Our proposed scheme provides a promising solution to the smart grid communication. In future work, we should investigate more scenarios with different constraints in practical smart grid network.

Acknowledgements The authors would acknowledge the financial and data support from State Grid Corporation of China (SGCC). This work was partially supported by the National Natural Science Foundation of China (61671233, 61801208, 61931023), the Jiangsu Science Foundation (BK20170650), the Postdoctoral Science Foundation of China (BX201700118, 2017M621712), the Jiangsu Postdoctoral Science Foundation (1701118B), and the open research fund of National Mobile Communications Research Laboratory (2019D02).

Appendix: Proof of the approximation ratio of LS-based algorithm

We first analyze the capacitated facility location problem with splittable demands (CFLS) and prove its approximation ratio.

Lemma 1.1 (Adding an AP) *If the current solution* $\mathcal{S}$ *satisfy* $c_s(\mathcal{S}) - c(\mathcal{S}^*) \geq \frac{nc(\mathcal{S})}{p(n)}$, *then there exists a* $n \in \mathcal{N}$ *that can be added to* $\mathcal{S}$ *to improve the current solution.*

Proof Noticing that $c_s(\mathcal{S}) > c_s(\mathcal{S}^*)$, we obtain $\mathcal{S}^* - \mathcal{S} \neq \emptyset$. Denote $n_1, \cdots, n_l$ as facilities in $\mathcal{S}^* - \mathcal{S}$. Now we need to prove that n_i is a facility among them that satisfies Lemma 1.1.

Denote σ and σ^* as the assignment for $\mathcal{S}$ and $\mathcal{S}^*$, respectively. We now need to analyze the difference between them. $G(\sigma, \sigma^*)$ is a difference graph for σ and σ^*. We can obtain a set of cycles $\mathscr{C}$ and paths $\mathscr{F}$ after decomposing the difference graph G. Clearly each one of the set $\mathscr{F} \cup \mathscr{C}$ contains a non-positive cost. Also we have $cost(\mathscr{C}) + cost(\mathscr{F}) = c_s(\mathcal{S}) - c_s(\mathcal{S}^*)$. Since $c_s(\mathcal{S}) - c(\mathcal{S}^*) \geq \frac{nc(\mathcal{S})}{p(n)}$, we have

$$\sum_{i=1}^{l} cost(\mathscr{F}_i) \geq c_s(\mathcal{S}) - c_s(\mathcal{S}^*). \tag{5}$$

For each n_i, we can have an assignment $\mathcal{S} + n_i$. The service cost will be

$$c_s(\mathcal{S} + n_i) = c_s(\mathcal{S}) - cost(\mathscr{F}_i). \tag{6}$$

From Eq. (5), we can obtain

$$\sum_{i=1}^{l} (cost(\mathscr{F}_i) - f_{n_i}) \geq c_s(\mathcal{S}) - c_s(\mathcal{S}^*) - c_f(\mathcal{S}^*) \tag{7}$$

after subtracting $c_f(\mathcal{S}^*)$.

The averaging result of Eq. (7) will be

$$cost(\mathscr{F}_i) - f_{n_i} \geq \left(c_s(\mathcal{S}) - c_s(\mathcal{S}^*) - c_f(\mathcal{S}^*)\right)/l. \tag{8}$$

Combining these bounds, we have

$$\begin{aligned} c(\mathcal{S}) - c(\mathcal{S} + u_i) &= c_s(\mathcal{S}) - c_s(\mathcal{S} + u_i) - f_{u_i} \\ &= cost(\mathscr{F}_i) - f_{u_i} \\ &\geq \frac{c_s(\mathcal{S}) - c_s(\mathcal{S}^*) - c_f(\mathcal{S}^*)}{l} \\ &\geq \frac{c_s(\mathcal{S}) - c_s(\mathcal{S}^*)}{n} \\ &\geq \frac{c(\mathcal{S})}{p(n)}. \end{aligned} \tag{9}$$

Consider that $c_s(\mathcal{S}) - c(\mathcal{S}^*) \geq \frac{nc(\mathcal{S})}{p(n)}$, the Proof of Lemma 1.1 is completed. □

Lemma 1.2 (Dropping/swapping APs) *$\mathcal{S}$ is a subset of $\mathcal{N}$. An upper bound holds if the cost of facility $\mathcal{S}$ cannot be improved by at least $c(\mathcal{S})/p(n)$ through dropping or swapping:*

$$c_f(\mathcal{S})\left(1 - \frac{n^2}{p(n)}\right) < 5c(\mathcal{S}^*) + 2c_s(\mathcal{S}) + \frac{c(\mathcal{S})}{n}. \tag{10}$$

Proof $\mathcal{S}^*$ is the optimal solution while $\mathcal{S}$ is a solution that satisfies Lemma 1.2. The following candidate operations for $\mathcal{S}$ are analyzed to prove the hypothesis of Lemma 1.2.

When an AP i satisfies $f_i \leq c(\mathcal{S})/n$, we call it *cheap*, and *expensive* otherwise. If an AP i in $\mathcal{S} - \mathcal{S}^*$ satisfies $D_i'' \leq M/2$, we call it *light*. Facilities that are not light are called *heavy*. The closest facility to be chosen is called *primary*, and *secondary* otherwise. Now $\mathcal{S}$ can be partitioned into several classes. $\mathcal{S}_C$ refers to facilities in $\mathcal{S} - \mathcal{S}^*$ that are *cheap*. $\mathcal{S}_{EL}$ refers to costly facilities in $\mathcal{S} - \mathcal{S}^* - \mathcal{S}_C$, as well as and $\mathcal{S}_{EH}$. $\mathcal{S}_{ELP}$ and $\mathcal{S}_{ELS}$ denote those *expensive light* APs which can be called *primary* or *secondary*.

Then we have the following candidate operations for APs in different classes:

- AP $i \in \mathcal{S} \cap \mathcal{S}^*$ and $\mathcal{S}_C$: do nothing.
- AP $i \in \mathcal{S}_{EH}$ and $\mathcal{S}_{ELP}$: use i^* to replace i.
- AP $i \in \mathcal{S}_{ELS}$: drop i and rearrange the load of i.

A *good* operation means that the total cost can be cut down by $c(\mathcal{S})/p(n)$ or more with this operation. β is a refined allocation strategy with the cost of $q_{\hat{\beta}}$, which satisfies $q_{\hat{\beta}} \leq c_s(\mathcal{S}) + c_s(\mathcal{S}^*)$.

If we get *bad* operations for all facilities in $\mathcal{S}_{EH}$, then

$$\begin{aligned} &c_f(\mathcal{S}_{EH})\left(1 - \frac{n^2}{p(n)}\right) \\ &\quad \leq 4c_f(\mathcal{S}^* - \mathcal{S}) + \sum_{i \in \mathcal{S}_{EH}} q_{\hat{\beta}}(i). \end{aligned} \tag{11}$$

If we get *bad* operations for all facilities in $\mathcal{S}_{ELP}$, then

$$\begin{aligned} &c_f(\mathcal{S}_{ELP})\left(1 - \frac{n^2}{p(n)}\right) \\ &\quad \leq c_f(\mathcal{S}^* - \mathcal{S}) + \sum_{i \in \mathcal{S}_{ELP}} q_{\hat{\beta}}(i). \end{aligned} \tag{12}$$

If we get *bad* operations for all facilities in $\mathcal{S}_{ELS}$, then

$$c_f(\mathcal{S}_{ELS})\left(1 - \frac{n^2}{p(n)}\right) \leq \sum_{i \in \mathcal{S}_{ELS}} 2q_{\hat{\beta}}(i). \tag{13}$$

We have the inequality $c_f(\mathcal{S}_C) \leq c(\mathcal{S})/n$ and $q_{\hat{\beta}} \leq c_s(\mathcal{S}) + c_s(\mathcal{S}^*)$. Noticed that

$$\sum_{i \in \mathcal{S}_{EH}} q_{\hat{\beta}}(i) + \sum_{i \in \mathcal{S}_{ELP}} q_{\hat{\beta}}(i) + \sum_{i \in \mathcal{S}_{ELS}} q_{\hat{\beta}}(i) = \sum_{i \in \mathcal{S} - \mathcal{S}^*} q_{\hat{\beta}}(i).$$

Combining all the conclusions above, we obtain

$$\begin{aligned} &c_f(\mathcal{S})\left(1 - \frac{n^2}{p(n)}\right) \\ &\quad = (c_f(\mathcal{S} \cap \mathcal{S}^*) + c_f(\mathcal{S}_C) + c_f(\mathcal{S}_{EH}) \\ &\qquad + c_f(\mathcal{S}_{ELP}) + c_f(\mathcal{S}_{ELS})) \times \left(1 - \frac{n^2}{p(n)}\right) \\ &\quad \leq c_f(\mathcal{S} \cap \mathcal{S}^*) + \frac{c(\mathcal{S})}{n} + 5c_f(\mathcal{S}^* - \mathcal{S}) \\ &\qquad + \sum_{i \in \mathcal{S} - \mathcal{S}^*} 2q_{\hat{\beta}}(i) \\ &\quad \leq \frac{c(\mathcal{S})}{n} + 2q_{\hat{\beta}} + 5c_f(\mathcal{S}^*) \\ &\quad \leq \frac{c(\mathcal{S})}{n} + 2c_s(\mathcal{S}) + 2c_s(\mathcal{S}^*) + 5c_f(\mathcal{S}^*) \\ &\quad \leq \frac{c(\mathcal{S})}{n} + 2c_s(\mathcal{S}) + 5c_s(\mathcal{S}^*). \end{aligned} \tag{14}$$

This completes the Proof of Lemma 1.2. □

Theorem 1 can be proved with Lemmas 1.1 and 1.2.

Theorem 1 (CFLS) *For any constant $\epsilon > 0$, the LS-based algorithm yields an $(8 + \epsilon)$-approximate solution in polynomial time.*

Proof According to Lemma 1.1, we have

$$c_s(\mathcal{S}) < c(\mathcal{S}^*) + \frac{nc(\mathcal{S})}{p(n)}. \tag{15}$$

According to Lemma 1.2, we have

$$\begin{aligned} c_f(\mathcal{S})\left(1 - \frac{n^2}{p(n)}\right) &\leq 5c(\mathcal{S}^*) + 2c_s(\mathcal{S}) + \frac{c(\mathcal{S})}{n} \\ &< 7c(\mathcal{S}^*) + nc(\mathcal{S})\left(\frac{2}{p(n)} + \frac{1}{n^2}\right). \end{aligned} \tag{16}$$

If we add the upper bound to the cost of AP, we can have

$$c(\mathcal{S})\left(1 - \frac{n^2}{p(n)}\right) < 8c(\mathcal{S}^*) + nc(\mathcal{S})\left(\frac{3}{p(n)} + \frac{1}{n^2}\right). \tag{17}$$

By rearranging, we obtain

$$c(\mathcal{S})\left(1 - \frac{n^2}{p(n)} - \frac{3n}{p(n)} - \frac{1}{n}\right) < 8c(\mathcal{S}^*). \tag{18}$$

This completes the Proof of Theorem 1. □

In general, when obtaining the assignment of unsplittable case from the assignment of splittable case, the capacities of all APs in unsplittable case increase by a factor of at most two. If the expansion of capacity leads to a increase of facility cost, the cost of AP in an unsplittable scenario is double times than that in a splittable case at most. This conclusion has been confirmed in [24]. But in the system model of this paper, the installation cost of an AP is much bigger than the price of the general purpose processor, we assume that the facility cost of an AP will not change with the increase of SMs it serves, which means the facility cost of unsplittable case in this paper is equal to the splittable case. Corollary 1 below summarize our results for CFLU in this paper.

Corollary 1 *(CFLU) If the expansion of facility capacity does not lead to an increase of the facility cost, the LS-based algorithm has an $(8+\varepsilon)$-approximate ratio ($\varepsilon > 0$) and can yields a solution in polynomial time.*

Proof For the special case in this paper, the cost of installing an AP is much bigger than the price of the general purpose processor in it, so we have

$$c_f^u(\mathcal{S}) = c_f^s(\mathcal{S}). \tag{19}$$

Known that

$$c_s^u(\mathcal{S}) \leq c_s^s(\mathcal{S}). \tag{20}$$

With simple mathematical operations, we have

$$c_f^u(\mathcal{S}) + c_s^u(\mathcal{S}) \leq c_f^s(\mathcal{S}) + c_s^s(\mathcal{S}) < (8+\epsilon)c(S^*). \tag{21}$$

This is the Proof of Corollary 1. □

References

1. Nghia Le, T., Chin, W. L., & Chen, H. H. (2017). Standardization and security for smart grid communications based on cognitive radio technologies: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *19*, 423–445.
2. Li, Z., & Liang, Q. (2016). Capacity optimization in heterogeneous home area networks with application to smart grid. *IEEE Transactions on Vehicular Technology*, *65*, 699–706.
3. Fan, Z., Kulkarni, P., Gormus, S., Efthymiou, C., Kalogridis, G., Sooriyabandara, M., et al. (2013). Smart grid communications: Overview of research challenges, solutions, and standardization activities. *IEEE Communications Surveys & Tutorials*, *15*, 21–38.
4. Vallejo, A., Zaballos, A., Selga, J. M., & Dalmau, J. (2012). Next-generation QoS control architectures for distribution smart grid communication networks. *IEEE Communications Magazine*, *50*, 128–134.
5. Aijaz, A., & Aghvami, A. H. (2015). PRMA-based cognitive machine-to-machine communications in smart grid networks. *IEEE Transactions on Vehicular Technology*, *64*, 3608–3623.
6. Monteiro, V., Carmo, J. P., Pinto, J. G., & Afonso, J. L. (2016). A flexible infrastructure for dynamic power control of electric vehicle battery chargers. *IEEE Transactions on Vehicular Technology*, *65*, 4535–4547.
7. Wang, K., Wang, Y., Zeng, D., & Guo, S. (2017). An SDN-based architecture for next-generation wireless networks. *IEEE Wireless on Communication*, *24*, 25–31.
8. Thyagaturu, A. S., Dashti, Y., & Reisslein, M. (2016). SDN-based smart gateways (Sm-GWs) for multi-operator small cell network management. *IEEE Transactions on Network and Service Management*, *13*, 740–753.
9. Bartoli, A., Hernandez-Serrano, J., Soriano, M., Dohler, M., Kountouris, A., & Barthel, D. (2011). Secure lossless aggregation over fading and shadowing channels for smart grid M2M networks. *IEEE Transactions on Smart Grid*, *2*, 844–864.
10. Ringwelski, M., Renner, C., Reinhardt, A., Weigel, A., & Turau, V. (2012). The Hitchhiker's guide to choosing the compression algorithm for your smart meter data. In *Proceedings of IEEE ENERGYCON'12* (pp. 935–940).
11. Ahmed, M. O., & Lampe, L. (2013). Power line communications for low-voltage power grid tomography. *IEEE Transactions on Communications*, *12*, 5163–5175.
12. Jimenez, Y., Cervello-Pastor, C., & Garcia, A. J. (2014). On the controller placement for designing a distributed SDN control layer. In *Proceedings of IFIP Networking'14*.
13. Lange, S., Gebert, S., Zinner, T., Tran-Gia, P., Hock, D., Jarschel, M., et al. (2015). Optimal model for the controller placement problem in software defined networks. *IEEE Transactions on Network and Service Management*, *12*, 4–17.
14. Sallahi, A., & St-Hilaire, M. (2017). Expansion model for the controller placement problem in software defined networks. *IEEE Communications Letters*, *21*, 274–277.
15. Lu, Z., & Wen, Y. (2014). Distributed algorithm for tree-structured data aggregation service placement in smart grid. *IEEE Systems Journal*, *8*, 553–561.
16. Aalamifar, F., Shirazi, G. N., Noori, M., & Lampe, L. (2014). Cost-efficient data aggregation point placement for advanced metering infrastructure. In *Proceedings of IEEE SmartGridComm'14* (pp. 344–349).
17. Huang, X., & Wang, S. (2015). Aggregation points planning in smart grid communication system. *IEEE Communications Letters*, *19*, 1315–1318.
18. Huang, X., Wang, S., & Wang, C. (2015). Aggregation points planning for smart grid communications: Wired and wireless cases. In *Proceedings of IEEE GLOBECOM'15*.
19. Huang, X., & Ansari, N. (2017). Resource exchange in smart grid connected cooperative cognitive radio networks. *IEEE Transactions on Vehicular Technology*, *66*, 6291–6298.
20. Zimmermann, M., & Dostert, K. (2002). A multipath model for the powerline channel. *IEEE Transactions on Communications*, *50*, 553–559.
21. Levi, R., Shmoys, D. B., & Swamy, C. (2012). LP-based approximation algorithms for capacitated facility location. *Mathematical Programming*, *131*, 365–379.
22. Grant, M., Boyd, S., & Ye, Y., (2009). CVX users' guide. http://cvxr.com/cvx/doc. Accessed 21 Jan 2018.
23. Wang, S., & Huang, X. (2016). Aggregation points planning for software-defined network based smart grid communications. In *Proceedings of IEEE INFOCOM'16*.
24. Shmoys, D. B., Tardos, É. & Aardal, K., (1997). Approximation algorithms for facility location problems (extended abstract). In *Proceedings of ACM STOC'97* (pp. 265–274).

Yue Li received the B.S. degree from Nanjing University, Nanjing, China, in 2018. She is currently pursuing the M.S. degree with the Department of Electrical Engineering, Stanford University, CA, USA. Her research interest includes wireless networks and machine learning.

Tianyu Wang received the B.S. degrees in physics and a double major in computer software from Peking University, Beijing, China, in 2011, and the Ph.D. degree from School of Electronics Engineering and Computer Science, Peking University, Beijing, China, in 2016. He is currently an associated researcher in the School of Electronic Science and Engineering at Nanjing University, China. He has published more than 30 IEEE journal and conference papers, and received the Best Paper Award from the IEEE International Conference on Communications (ICC) 2015, IEEE Global Communication Conference (GLOBECOM) 2014, and ICST China Communication Conference (ChinaCom) 2012. His current research interest focuses on machine learning based wireless communication technology.

Shaowei Wang received his B.S., M.S. and Ph.D. degrees from Wuhan University, China. He is currently a full professor of the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests mainly include telecommunication systems, operations research and machine learning. He is on the editorial board of IEEE Communications Magazine, IEEE Transactions on Wireless Communications, and Springer Journal of Wireless Networks. He serves/served on the technical or executive committee of reputable conferences including IEEE INFOCOM, IEEE ICC, IEEE GLOBECOM, IEEE WCNC.