A Clustering Algorithm for the DAP Placement Problem in Smart Grid

Guodong Wang[1(✉)], Yanxiao Zhao[1], Yulong Ying[2], Jun Huang[3], and Robb M. Winter[4]

[1] Department of Electrical and Computer Engineering, South Dakota School of Mines and Technology, Rapid City, SD 57701, USA
{guodong.wang,yanxiao.zhao}@sdsmt.edu
[2] School of Energy and Mechanical Engineering, Shanghai University of Electric Power, Shanghai, China
[3] Department of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[4] Department of Chemical and Biological Engineering, South Dakota School of Mines and Technology, Rapid City, SD 57701, USA

Abstract. In this paper, we investigate the DAP placement problem and propose solutions to reduce the distance between DAPs and smart meters. The DAP placement problem is formulated to two objectives, e.g., the average distance minimization and the maximum distance minimization. The concept of network partition is introduced in this paper and practical algorithms are developed to address the DAP placement problem. Extensive simulations are conducted based on a real suburban neighborhood topology. The simulation results verify that the proposed solutions are able to remarkably reduce the communication distance between DAPs and their associated smart meters.

Keywords: Smart meter · DAP placement · Network partition Transmission routes

1 Introduction

Smart grid is generally referred to as the next generation power grid which enables high-speed and two-way communications to increase efficiency, management and reliability of energy resource. Neighborhood area networks play a significant role for the communications in smart grid. A neighborhood area network is typically composed of smart meters and Data Aggregation Points (DAPs). Smart meters are responsible for recording energy consumption or billing information of smart houses. DAPs collect the information from different smart meters and forward it to wide area network gateways. Wireless communications are recommended for neighborhood area networks due to their advantages in deployment flexibility and economy efficiency. For a neighborhood area network, the location of DAPs greatly affects the performance of communications between DAPs and their associated smart meters. Take the communication distance as an

G. Sun and S. Liu (Eds.): ADHIP 2017, LNICST 219, pp. 349–359, 2018.
https://doi.org/10.1007/978-3-319-73317-3_41

example. The location of DAPs influences the communication distance between DAPs and smart meters, which further influences the energy consumption, transmission rate, and end-to-end latency in neighborhood area networks. Therefore, it is critical to investigate how to appropriately select locations of DAPs in a neighborhood area network.

The problem of selecting locations for DAPs is termed as the DAP placement problem, which aims to properly choose locations of DAPs and allocate appropriate smart meters to achieve an objective. In the literature, the DAP placement problem is under-explored, and only a few DAP placement strategies are proposed [1,2]. To fill the gap, we have proposed a method to shorten the maximum distance between DAP and smart meters [3]. In this paper, we continue our research and extend the objective of DAP placement problem in neighborhood area networks. Specifically, we develop this problem into two objectives: the average distance minimization and the maximum distance minimization. The first objective is to minimize the average distance between DAPs and smart meters. The second objective is to minimize the maximum distance between DAPs and smart meters. To achieve those goals, we introduce the concept of network partition technique in this paper and clustering based algorithms are developed to address those two problems. In particular, the DAP placement problem is formulated as a network partition problem and a Clustering-based DAP Placement Algorithm (CDPA) is proposed to tackle the DAP placement problem. An actual suburban neighborhood is adopted as a topology to evaluate the performance of our solution. Simulation results verify that the proposed solution is able to significantly reduce the distance between DAP and their associated smart meters.

The rest of the paper is organized as follows. In Sect. 2, the related work is introduced. Section 3 formulates the problem mathematically. In Sect. 4, the proposed new solution is described in detail. Section 5 presents the performance evaluation of the solution. Section 6 concludes this paper

2 Related Work

One of the objectives of the DAP placement is to reduce the deployment cost of DAPs. In [1], the total cost of DAP placement is formulated as the operating time of a network, the cost of installing a DAP and the price of energy consumption. The optimal location of DAP is calculated to minimize the total cost of deploying DAPs in a network. The solution is heavily dependent on the model of cost function, which limits its application in practical systems. In addition, authors of [1] assume one-hop communications from smart meters to DAPs, which may be inapplicable to smart meters that have limited transmission ability. Therefore, it is reasonable to extend the one-hop assumption to a more common situation.

Authors of [2] propose another approach to decrease the deployment cost of DAPs and enable all meters can establish a reliable communication to one or more DAPs. In this paper, the DAP problem is converted to a set covering problem, which is addressed by heuristic approaches. Specifically, a subset of reliable links is pre-constructed based on characteristics of neighborhoods, communication technologies, transmission rates of antennas and their height. The

DAP placement problem is divided into several independent subsets, which is helpful for reducing the execution time and memory for solving the problem.

A smart grid is a large and complex system which consists of power generation, transmission and distribution as well as operations and management such as metering and billing [4]. A huge amount of data needs to be processed and exchanged in a smart grid. Availability of the smart grid requires time latency to be met for different operations and data transmissions [5]. For example, signal of protective actions needs to be generated and transmitted in the order of milliseconds. SCADA data needs to be transmitted within several seconds [6]. The results of [7,8] demonstrate that in a smart grid, the propagation delay of a data packet in the application of fast faults detection should be within the order of tens of milliseconds for a small size network, and 100 ms is acceptable in a medium size network.

In the context of networking, shortest distance path is one of the most common methods adopted for relaying messages in a wide variety of networks [9], since it provides an efficient way to decrease the energy and latency [10,11], which is significant to the overall performance of a network [12–14]. Therefore, in our paper, we focus on a delay-sensitive smart meter network and aim to minimize the communication distance between DAPs and their associated smart meters. In order to fully explore the transmission distance between DAPs and smart meters, we formulate and investigate the problem through two situations. The first situation is to minimize the average distance between DAPs and smart meters. The second one is to minimize the maximum distance between all the DAPs and their associated smart meters. Afterwards, the concept of network partition is introduced in this paper to address the DAP placement problem. Specifically, for a given neighborhood area network, the entire network is divided into subnetworks and DAPs are placed in the locations to minimize the average distance or the maximum distance between DAPs and smart meters. An actual suburban neighborhood is adopted in this paper to evaluate the performance of our solution. Simulation results verify that the proposed solution is able to significantly reduce the distance between DAP and their associated smart meters.

3 Problem Formulation

In this section, we briefly introduce the terminology and definitions that are used in this paper and formulate the DAP placement problem in a smart meter neighborhood area network.

Given a specific suburban neighborhood, let $V = \{v_i\}_{i=1,\cdots,|V|}$, where v_i is the i^{th} smart meter. The location of i^{th} smart meter is denoted by $v_i^l(x_i, y_i)$, where x_i and y_i are the longitude and latitude of smart meters, respectively. The task of DAP placement is to partition the network G = (V, E) into subnetworks and allocate DAPs to those subnetworks. Denote the set of DAPs by $DAP = \{dap_1, \cdots, dap_k\}$, where k is the number of DAPs. The set of smart meters allocated to dap_i is denoted by $S_i = \{s_0, s_1, \cdots s_{n_i}\}$, where $n_i = |S_i|$, representing the number of smart meters allocated to the i^{th} DAP. Denote the distance between any two nodes by d(u,v) $(u, v \in V)$.

Distance plays a significant role in wireless communications, since it greatly affects the energy consumption and routing optimization [15,16]. For example, in a wireless sensor network, sensors are expected to forward a packet to the neighbor that has the shortest distance to the destination [17,18]. As a special wireless sensor network, DAPs in a neighborhood area network should be placed appropriately to shorten the distance between DAPs and smart meters. Focusing on the distance minimization, there are two objectives regarding the distance between DAPs and smart meters. The first one is the average distance minimization, which stands for minimizing the average distance between smart meters and DAPs. The other one is the worst-case distance minimization, which stands for minimizing the maximum distance between smart meters and DAPs. Those two objectives are formulated as follows.

Average Distance Minimization. The average distance for a DAP placement P' is:

$$D_{avg}(P') = \frac{1}{|V|}\sum_{j=1}^{k}\sum_{s_i \in S_j} \mathbf{min}\ d(s_i, dap_j) \tag{1}$$

Worst-case Distance Minimization. The worst-case distance for a DAP placement P' is:

$$D_{wc}(P') = \mathbf{min}\{\mathbf{max}\{d(s_i, dap_j)\}\} \tag{2}$$

Denote subnetworks by $A = \{A_i\}_{i=1,\cdots,k}$, then Eqs. (1) and (2) are subjected to:

$$DAP_i, s_i \in A_i, \forall\ i = 1, 2, \cdots, k \tag{3}$$

The DAP placement problem resembles a facility location problem [19], which is NP-hard and requires heuristic approaches [20]. We investigate the DAP placement problem and seek methods to minimize the distance between DAPs and their associated smart meters. Specifically, two clustering algorithms are developed to address the objectives, which are formulated in Eqs. (1) and (2), respectively. We also evaluate and compare the performance associated with those two objectives and present implications to the DAP placement problem.

4 Clustering-Based DAP Placement Algorithms

In this section, we elaborate how to develop clustering algorithms to partition a neighborhood area network into subnetworks and place DAPs accordingly. Since the DAP placement problem resembles facility location problems, solutions can be borrowed from contexts of clustering algorithms. However, the standard clustering algorithms, e.g., K-means, cannot be directed adopted to address network partition and the DAP placement problem. To facilitate understanding, we first introduce the concept of clustering algorithms and discuss their shortcomings in partitioning a network topology. Afterwards, a clustering-based DAP placement algorithm will be developed to conduct the network partition and the DAP placement. Note that there are two critical parameters in clustering algorithms:

center and centroid. For clarification, the initial nodes which are selected to perform clustering algorithms are termed as center. The actual node, which is eventually found by clustering algorithms, is termed as centroid.

A typical clustering algorithm, e.g., K-means, includes four main steps:

- Initialize k clusters and allocate one center for each cluster using random sampling;
- Allocate nodes into one of the clusters based on Euclidean distance;
- Recalculate centroid for each cluster;
- Repeat step 2 and 3 until there is no change in each cluster.

The standard clustering algorithms cannot be directly applied to network partitions due to the following reasons. First, the Euclidean distance cannot be used in calculating the distance between two nodes, because physical links may not exist in the path of the Euclidean distance. Second, due to the limited transmission range of smart meters, relay smart meters are needed to forward messages from one smart meter to the DAP. Third, the centroid, where DAP is placed, should be chosen from existing smart meters ($DAP_i \in V$) to guarantee there are routes between the centroid and its associate smart meters.

In this section, we propose a Clustering-based DAP Placement Algorithm (CDPA) to partition a neighborhood area network into subnetworks and place DAPs to appropriate locations. The associated algorithm is presented in Algorithm 1.

The input of CDPA consists of the number (n) of smart meters of a given network, the coordinates of smart meters ($sl_i(x_i, y_i)_{\{i=1,\cdots,n\}}$), the transmission range (r_c) of each smart meter, and the number of DAPs (k). The output of CDPA includes the identification of DAPs (DAP), the set of smart meter clusters (S) and the routing set for each subnetwork (R). Following is the detailed explanation for each step.

The first step of CDPA is to calculate the shortest path of each pair of smart meters among their possible routes. Note that, the distance adopted in this algorithm represents the shortest path distance. In step 2, k smart meters are randomly selected from the network as centers of each cluster to initialize the clustering. In Step 3, smart meters are allocated into different clusters based on their shortest path distances to the k centers. Specifically, smart meters are allocated to the cluster which is closest to the k centers. For each round, since new smart meters are allocated to those k clusters, the cluster center should be recalculated for each cluster. This process is conducted in step 4, which consists of two methods to address different objectives. In step 4 (1), the new cluster center is obtained aiming at minimizing the average distance between cluster members and the center. In contrast, the worst-case situation is achieved in step 4 (2), which aims to minimize the maximum distance between cluster members and the center. The processes of step 3 and 4 are repeated until there is no change in each cluster, as presented in step 5. The output of the algorithm is obtained in Step 6.

Algorithm 1. Clustering-based DAP Placement Algorithm

Input:
(1) n (the number of smart meters)
(2) $SL = \{sl_i(x_i, y_i)\}_{i=1,\cdots,n}$ (the coordinate of smart meters)
(3) r_c (transmission Range)
(4) k (the number of DAPs)
Output:
(1) $DAP = \{dap_i\}_{i=1,\cdots,k}$ (the instance of DAP)
(2) $S = \{s_i\}_{i=1,\cdots,k}$ (the set of smart meters)
(3) $R = \{r_i\}_{i=1,\cdots,k}$ (the routing paths of each subnetwork)
Step 1: Calculate the shortest path distance between any two nodes. (In the rest of this algorithm, distances represent shortest path distances)
Step 2: Initialize k centers ($C = \{c_i\}_{i=1,\cdots,k}$) by randomly selecting k smart meters from S.
Step 3: Distribute the smart meter s_i ($s_i \in S$) to one of the k clusters using the relation,

$$v \in sc_i,\ if\ \ d(v, c_i) < d(v, c_j), \forall j \in \{1, 2, \cdots, k\}$$

where $d(u, v)$ represents the shortest path between smart meter u and v.
Step 4: Update centroids $C' = \{c_i\}_{i=1,\cdots k}$ such that:
(1) the **average distance is minimized** as formulated in Eq. 1, which is termed as ($CDPA_{avg}$).

$$c_i' = v_m,\ if\ \ \frac{1}{n'} \sum d(v_m, v) = minimum,$$
$$\frac{1}{n'} = size(cluster_i)$$
$$\forall\ m\ , v\ \ \in cluster_i,\ \ i = \{1, 2, \cdots, k\}$$

or **(2)** the **maximum distance is minimized** as formulated in Eq. 2, which is termed as ($CDPA_{wc}$).

$$c_i' = v_m,\ if\ \ Max\{d(v_m, v)\} = minimum,$$
$$\forall\ m\ , v\ \ \in cluster_i,\ \ i = \{1, 2, \cdots, k\}$$

Step 5: Repeat steps 3 and 4 until there is no change in each cluster.
Step 6: Save C' to DAP, $\{cluster_i\}_{i=1,\cdots k}$ to S and routing pathes to R.

5 Performance Evaluation

In this section, we demonstrate the network partition as well as DAP placement results achieved by CDPA, and evaluate its performance in terms of distance minimization. A real suburban neighborhood is selected from Rapid City, SD, USA, as the neighborhood area network.

5.1 Demonstration of DAP Placement Achieved by CDPA and the Associated Routes

In Sect. 4, CDPA is proposed to conduct network partition and DAP placement. Actually, CDPA consists of two algorithms, $CDPA_{avg}$ and $CDPA_{ws}$, which are adopted to achieve two different objectives as formulated in Sect. 3. Specifically,

$CDPA_{avg}$ targets at minimizing the average distance between DAPs and smart meters. $CDPA_{ws}$ aims at minimizing the maximum distance between DAPs and their associated smart meters. In this subsection, we demonstrate and compare the DAP placement results conducted by $CDPA_{avg}$ and $CDPA_{ws}$, respectively.

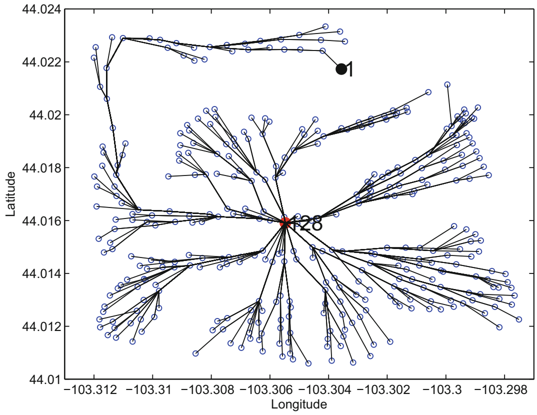

Fig. 1. Demonstration of DAP placement and associated routes achieved by $CDPA_{avg}$

Figure 1 depicts the DAP placement and associated routes achieved by $CDPA_{avg}$. In particular, when only one DAP is deployed in the network, the position of DAP is selected as node 128, which has the smallest sum distance to all other smart meters. The maximum distance among the network is from the DAP (node 128) to node 1, where the distance is 1870.4 m. The average distance between DAP and other smart meters is 549.42 m. The optimal routes from all

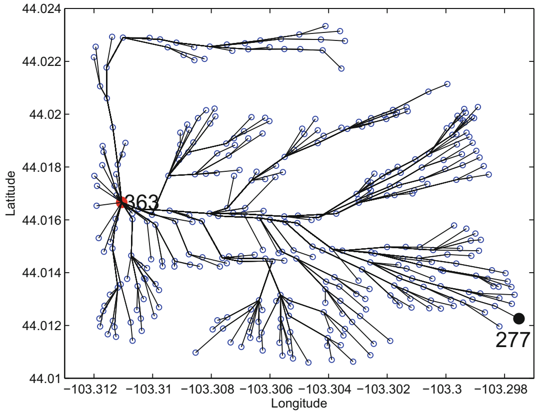

Fig. 2. Demonstration of DAP placement and associated routes achieved by $CDPA_{ws}$

of the smart meters to the DAP are also presented, which are the shortest pathes among all of the possible routes.

The DAP placement result of the $CDPA_{wc}$ is demonstrated in Fig. 2. Compared with $CDPA_{avg}$, the $CDPA_{ws}$ aims to shorten the maximum distance between DAPs and smart meters, so the DAP is preferred to be placed in the location where the maximum distance between DAP and smart meters is minimized. Specifically, the position of DAP obtained by $CDPA_{wc}$ is at node 363, which has the minimum distance to other smart meters. The maximum distance between DAP and other nodes in the network is 1380.4 m, which is from node 363 to node 277. We can find that the maximum distance is significantly decreased by $CDPA_{ws}$ in comparison with $CDPA_{avg}$. However, the average distance achieved by $CDPA_{ws}$ is 760.90 m, which is over 200 m larger than the result achieved by $CDPA_{avg}$.

5.2 Comparison of Distance Between DAPs and Smart Meters

In this subsection, we compare the performance of those two algorithms, $CDPA_{avg}$ and $CDPA_{wc}$, in terms of maximum distance minimization and average distance minimization, respectively. To reflect the real performance of those two algorithms, each of them is executed for 100 times. Cumulative Distribution Functions (CDFs) of those results are plotted and depicted in Figs. 3 and 4, respectively.

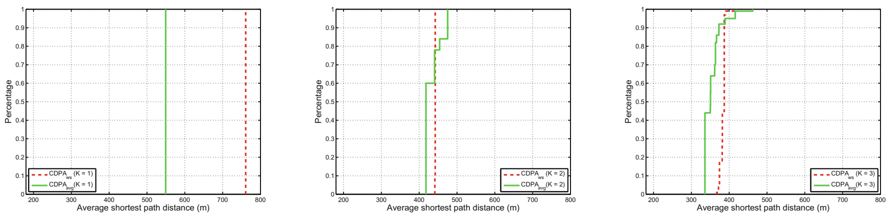

Fig. 3. Average distance between DAPs and smart meters ($CDPA_{avg}$)

As depicted in Fig. 3, $CDPA_{avg}$ outperforms $CDPA_{wc}$ in the average distance minimization, since the majority of the average distance achieved by $CDPA_{avg}$ is smaller than the one achieved by $CDPA_{ws}$, although the difference decreases with the increase of k . Those results further verify that the proposed algorithms $CDPA_{wc}$ and $CDPA_{avg}$ are capable of decreasing the maximum distance and average distance between DAPs and smart meters, respectively. In addition, since the difference of the average distance achieved by $CDPA_{wc}$ and $CDPA_{avg}$ decreases with the increase of k, $CDPA_{wc}$ is recommended to conduct DAP placement in a neighborhood area network due to its better performance in the maximum distance minimization.

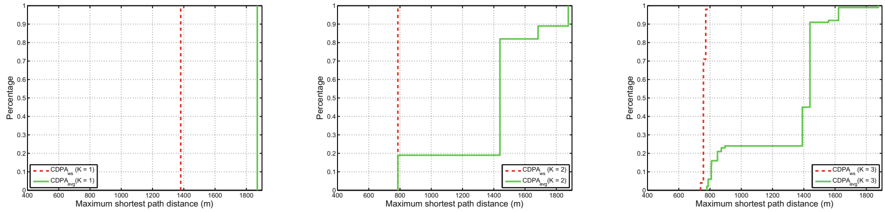

Fig. 4. Maximum distance between DAPs and smart meters ($CDPA_{ws}$) (Color figure online)

In contrast, Fig. 4 depicts the cumulative distribution function of the maximum distance achieved by both $CDPA_{avg}$ and $CDPA_{wc}$, which reveals how the distance between DAPs and smart meters is shortened for each execution. As depicted in Fig. 4, the dashed red curve represents the maximum distance achieved by $CDPA_{wc}$, while the solid green curve is the result achieved by $CDPA_{avg}$. It is clearly observed that $CDPA_{wc}$ outperforms $CDPA_{avg}$ in the maximum distance minimization, since the maximum distance achieved by $CDPA_{wc}$ is smaller than the one achieved by $CDPA_{avg}$.

6 Conclusion

In this paper, we focus on the DAP placement in a smart meter neighborhood area network and aim at minimizing the distance between DAPs and their associated smart meters. To achieve this goal, we formulate this problem and propose two objectives regarding the distance minimization. They are the average distance minimization and the maximum distance minimization. The concept of network partition is introduced in this paper and a clustering-based DAP placement algorithm is developed to tackle the DAP placement problems. Based on this approach, an entire network is divided into subnetworks and one DAP is deployed at an optimal position of each subnetwork. An actual suburban neighborhood is adopted in this paper to evaluate the performance of the proposed solution. Simulation results verify that the proposed solution is able to significantly reduce the distance between DAP and their associated smart meters. At this phase, we only focus on shortening the distance between DAPs and their associated smart meters. Besides distance minimization, there are many other challenges, e.g., energy saving, reliability and resistance to be tackled to pave a way for actual DAP placements in Smart Grid, which will be investigated in the future.

References

1. Aalamifar, F., Shirazi, G.N., Noori, M., Lampe, L.: Cost-efficient data aggregation point placement for advanced metering infrastructure. In: 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 344–349 (2014)
2. Rolim, G., Passos, D., Moraes, I., Albuquerque, C.: Modelling the data aggregator positioning problem in smart grids. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), pp. 632–639 (2015)
3. Wang, G., Zhao, Y., Huang, J., Winter, R.: On the data aggregation point placement in smart meter networks. In: 2017 26th International Conference on Computer Communication and Networks, pp. 1–6 (2017)
4. Yan, Y., Qian, Y., Sharif, H., Tipper, D.: A survey on smart grid communication infrastructures: motivations, requirements and challenges. Commun. Surv. Tutor. IEEE **15**(1), 5–20 (2013)
5. Amin, M.: Challenges in reliability, security, efficiency, and resilience of energy infrastructure: toward smart self-healing electric power grid. In: 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, pp. 1–5 (2008)
6. Bennett, C., Wicker, S.B.: Decreased time delay and security enhancement recommendations for AMI smart meter networks. In: Innovative Smart Grid Technologies (ISGT), pp. 1–6 (2010)
7. Sood, V.K., Fischer, D., Eklund, J., Brown, T.: Developing a communication infrastructure for the smart grid. In: 2009 IEEE Electrical Power & Energy Conference (EPEC), pp. 1–7 (2009)
8. Aggarwa, A., Kunta, S., Verma, P.K.: A proposed communications infrastructure for the smart grid. In: Innovative Smart Grid Technologies (ISGT), pp. 1–5, 19–21 January 2010
9. Wang, G., Zhao, Y., Huang, J., Duan, Q., Li, J.: A K-means-based network partition algorithm for controller placement in software defined network. In: International Conference on Communications (2016)
10. Krishnamachari, L., Estrin, D., Wicker, S.: The impact of data aggregation in wireless sensor networks. In: 22nd International Conference on Distributed Computing Systems Workshops, pp. 575–578 (2002)
11. Yilmaz, O., Demirci, S., Kaymak, Y., Ergun, S., Yildirim, A.: Shortest hop multipath algorithm for wireless sensor networks. Comput. Math. Appl. **63**(1), 48–59 (2012)
12. Wang, G., Wu, Y., Dou, K., Ren, Y., Li, J.: AppTCP: the design and evaluation of application-based TCP for e-VLBI in fast long distance networks. Future Gener. Comput. Syst. **39**, 67–74 (2014)
13. Wang, G., Ren, Y., Dou, K., Li, J.: IDTCP: an effective approach to mitigating the TCP incast problem in data center networks. Inf. Syst. Front. **16**, 35–44 (2014)
14. Wang, G., Ren, Y., Li, J.: An effective approach to alleviating the challenges of transmission control protocol. IET Commun. **8**(6), 860–869 (2014)
15. Ganesan, D., Govindan, R., Shenker, S., Estrin, D.: Highly-resilient, energy-efficient multipath routing in wireless sensor networks. ACM SIGMOBILE Mob. Comput. Commun. Rev. **5**(4), 11–25 (2001)

16. Muruganathan, S.D., Ma, D.C., Bhasin, R.I., Fapojuwo, A.O.: A centralized energy-efficient routing protocol for wireless sensor networks. IEEE Commun. Mag. **43**(3), S8–13 (2005)
17. Goyal, D., Tripathy, M.R.: Routing protocols in wireless sensor networks: a survey. In: 2012 Second International Conference on Advanced Computing & Communication Technologies (ACCT), pp. 474–480 (2012)
18. Pantazis, N.A., Nikolidakis, S.A., Vergados, D.D.: Energy-efficient routing protocols in wireless sensor networks: a survey. IEEE Commun. Surv. Tutor. **15**(2), 551–591 (2013)
19. Drezner, Z., Hamacher, H.W.: Facility Location. Springer, New York (1995)
20. Farahani, R.Z., Hekmatfar, M., Fahimnia, B., Kazemzadeh, N.: Hierarchical facility location problem: models, classifications, techniques, and applications. Comput. Ind. Eng. **68**, 104–117 (2014)