IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 23, 2019, accepted October 17, 2019, date of publication October 23, 2019, date of current version November 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949100

Optimal Multi-Base-Station Selection for Belt-Area Wireless Sensor Networks

ZHANGLING DUAN, XING WEI, JIANGHONG HAN, GUOQIANG YANG, AND YANG LU
School of Computer and Information, Hefei University of Technology, Hefei 230009, China

Corresponding author: Xing Wei (lijiajia@mail.hfut.edu.cn)

This work was supported in part by the National Key Research and Development Program under Grant 2018YFC0604404, in part by the Fundamental Research Funds for the Central Universities under Grant PA2019GDPK0079, and in part by the Anhui Provincial Key Research and Development Program under Grant 201904d08020040.

ABSTRACT Since underground environments, such as urban subways, tunnels, underground pipe corridors, and mine roadways, etc., are complex and changeable, the software functions of the sensor nodes in underground spaces must be updated regularly to satisfy the requirements of the monitoring center. Software updates cannot be traditionally conducted due to the large number of nodes, the wide distribution range and the poor underground environment; they can only be made using wireless reprogramming. Current reprogramming protocols are all designed for topologically unconstrained networks and are inefficient in routing constrained underground belt-area wireless sensor networks (BAWSNs). We propose a new reprogramming mechanism for reducing the energy and time consumption of the data downstream transmission process in BAWSNs. For network energy optimization, we identified the highest energy efficiency transmission radius. For network base-station location time optimization, an approximate $(1+\varepsilon)$ algorithm that is based on gradient cyclic descent is proposed, which is of complexity $O\left(kn^2\right)$. The simulation results demonstrate that, compared with classical algorithms, the BAWSNs approximation algorithm can locate the optimal base station accurately with low time consumption.

INDEX TERMS Belt-area wireless sensor networks, base station selection, p-center problem, optimization algorithm.

I. INTRODUCTION

With the accumulation of urban population, urban resource and environmental problems have become increasingly severe. Underground spaces, such as urban subways, tunnels, underground pipe corridors, and mine roadways, etc., are important for addressing the problems of urban resources and the environmental crisis, and are an important resource for sustainable development. In June 2016, the Ministry of Housing and Urban-Rural Development issued the ''Thirteenth Five-Year Plan for Urban Underground Space Development,'' with the objectives of completing the planning and approval of underground space development by 2020 in more than half of China's cities and establishing a more complete management system for underground space planning [1]. Underground spaces suffer from various problems, such as being of complex structure, being relatively closed, and hosting many types of accidents, including random accidents. Considering gas leakage under a mine, underground oil or gas pipeline leakage, tunnel roof deformation or other high-risk accidents, it is necessary to identify the fault point and to locate it in the shortest time possible; hence, it is essential to conduct all-weather, uninterrupted monitoring and warning of underground spaces. A wireless sensor network [2] is a type of multi-hop wireless self-organizing network that is composed of many sensor nodes and several base stations. Due to its flexibility in deployment and its rapid sensing of environmental parameters, a wireless sensor network can be used for real-time monitoring of underground space environments. Underground space scenes have a belt area distribution feature; namely, the belt area refers to a long-distance belt-shaped closed area. Such areas refer to longer, narrower areas that can reach tens of kilometers in length and often only tens of meters in width. Therefore, a wireless communication network composed of a WSN that is deployed in a special belt-like area of an underground space [3] is considered, as illustrated in Fig. 1.

However, the underground environment is complex and variable, and the software of each node must be updated and maintained regularly to satisfy the monitoring

The associate editor coordinating the review of this manuscript and approving it for publication was Hongwei Du.

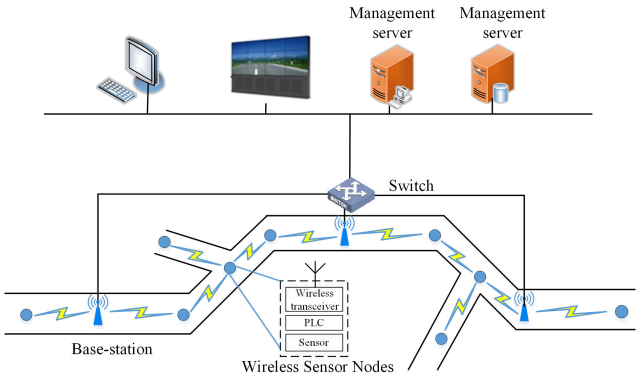

FIGURE 1. Belt-area wireless network reprogramming diagram.

requirements of the ground monitoring center. The belt-area wireless sensor networks (BAWSNs) cannot be traditionally artificially updated due to the large number of nodes, the wide distribution range and the poor underground environment. Software updates can only be made using wireless reprogramming. This gives rise to the need for ''over-the-air'' network reprogramming for updating sensor nodes in place. In some scenarios, all nodes of the network must be updated in the reprogramming process, such as system updating of sensors, and the transmission direction of reprogramming codes is from the base station to each sensor node. The most efficient approach is to use relay nodes to broadcast codes to the nodes that are within their transmission radii, some of which will be selected as new relay nodes. The process will continue until all nodes can be covered [4]. Many wireless reprogramming protocols share design challenges. Here, we consider the three most important challenges [5]:

(1) Completion time: The reprogramming completion time affects services that use WSNs. When we reprogram the network, we must suspend the services until the code update has been completed. Thus, we must minimize the reprogramming completion time.

(2) Energy efficiency: Sensor nodes are typically battery-powered and the sensor node battery provides the energy that is used in reprogramming. This battery also supplies energy for computing, communication, and sensing functions. Therefore, reprogramming must be energy-efficient.

(3) Reliability: Reprogramming requires the new code to be delivered throughout the entire network, and the delivered code must be executed correctly on the sensor nodes.

Therefore, it is necessary to design a new transmission mechanism for reducing the energy and time consumptions in the data downstream transmission process in a belt-area network scenario.

All available protocols, which include XNP [6], MOAP [8], Deluge [9], MNP [10], ACDP [11], and ACODI [12], for wireless reprogramming focus on transmission routing for energy efficiency. They utilize a very low-cost link estimation method that dynamically adjusts the size of the payload to increase the energy efficiency, thereby minimizing the energy consumption. These reprogramming routing protocols are all directed by a two-dimensional unconstrained network, which becomes inefficient in routing constrained belt-area wireless sensor networks. Reprogramming is a type of downlink data communication method. Under routing restrictions, the location selection of the base station, which is the initial point of reprogramming code transmission, has an important impact on the energy and time consumptions of the entire code update process. Therefore, selecting the optimal base-station position is of substantial importance for reducing the time and energy costs of the entire reprogramming process. Unfortunately, in the field of network energy optimization site selection, most studies [14]–[19] consider the application scenarios of two-dimensional unconstrained topology networks. Few studies have been conducted on the base-station location in the wireless network topology with the network topology limited belt-area wireless network and data downlink propagation.

In the BAWSNs information transmission process, due to the narrowness of the monitoring area and the limited transmission distance of the nodes, the data must be forwarded by intermediate nodes to the base station. The increased number of data forwarding hops leads to an increase in the transmission delay. In the two-dimensional unconstrained network topology, the shortest communication path can be selected via routing to reduce the network data forwarding delay. However, the spatial belt-shaped topology of the underground space causes the transmission direction of the node to completely rely on the extension direction of the belt-area; therefore, routing must follow the underground space topology and it is impossible to optimize the data forwarding hop count from the routing aspect. At this point, the deployment of the base station as a data aggregation center has a crucial impact on the data forwarding delay of the entire network. The objective of this paper is to select the optimal base-station location such that the image code is transmitted from the base station to the nodes with the most energy-efficient transmission radius and the time for updating the software of all nodes is the shortest. We propose a gradient-based $(1+\varepsilon)$-approximate optimal multi-base-station selection algorithm, namely, the optimal multi-base-station selection (OMSS) algorithm, in which the shortest data forwarding delay is the optimization target for the uniform linear network topology of BAWSNs. This study consists of the following steps:

(1) First, a time-efficient multi-base-station p-location model is presented. Next, the model is simplified. Under the BAWSNs special network structure, the location problem is transformed into the longest non-cyclic path problem. The NP-completeness of the time-efficient multi-base-station location model is indicated.

(2) Then, an exact algorithm for the problem index time and the gradient-based $(1+\varepsilon)$ strategy and algorithmic steps for an approximately optimal algorithm of the BAWSNs are presented. The complexities of the two algorithms are analyzed in detail.

(3) Finally, the time performances of the two algorithms are evaluated via simulation experiments. According to the results of the simulation experiments and direct testing, the OMSS algorithm can accurately locate the optimal position in a shorter time in the BAWSNs linear topology than the original p-center location algorithm.

The remainder of the paper is organized as follows: Section II discusses the shortcomings of the classical algorithms for the p-center problem. Section III established the p-center location model for multi-base-station selection. Section IV proves the NP-hardness of the belt-area network location problem and proposes an approximate algorithm that is based on gradient cyclic descent. In section V, the OSSM algorithm is simulated, and the differences between the OSSM algorithm and other classical algorithms are identified; additionally, the differences between OSMM, SAA and CIK algorithm are discussed, and the evaluation results are presented. Section VI presents the conclusions of the paper.

II. RELATED WORK

A. WIRELESS REPROGRAMMING

In wireless reprogramming, updating a traditional wireless sensor network is very common, e.g., updating parameters, upgrading software, and fixing security vulnerabilities. XNP [6] is the first data dissemination protocol for TinyOS [7]; however, it only provides the single-hop network solution, and all nodes must be in the communication range of the source node. In addition, data dissemination uses only the flooding mechanism and does not support incremental updating of the program image. MOAP [8] extends the code delivery to multi-hop networks. It disseminates data via a hop-by-hop approach with local broadcasting and sliding window recovery. However, one of the core assumptions of the MOAP design is that the time delay is the least important resource; thus, time is used to represent energy. Deluge [9] is the de facto network reprogramming protocol that is based on TinyOS, which employs ADV-REQ-DATA three-way message exchanges to ensure its eventual consistency, and it disseminates codes via an epidemic routing algorithm. The multi-hop network reprogramming (MNP) protocol [10] is an improved version of Deluge. The routing strategy is to use the node with the most connected nodes to update at the same communication radius as the forwarding node, thereby improving the energy utilization efficiency. The adaptive code dissemination protocol (ACDP) is an adaptive code distribution protocol [11]. The main strategy is to maximize the network life cycle by selecting the node with the most residual energy as the routing forwarding node. The adaptive code dissemination based on link quality (ACODI) protocol [12] is a link-quality-based adaptive code distribution. It provides a very low cost link estimation method that dynamically adjusts the size of the payload to increase the energy efficiency, thereby minimizing the energy consumption. In a recent publication on reprogramming, Teng and Liu [13] argue that recruiting many vehicles to disseminate the update code for roadside smart devices (RSDs) via vehicle-to-sensing device communications technology is an effective method. They proposed a cost-efficient greedy code mule selection scheme (CGCSS) for disseminating code to a huge number of RSDs in a smart city.

All available reprogramming protocols are designed for wireless sensor networks with unlimited two-dimensional topologies, which lose their original advantages in topologically constrained network topologies. This paper proposes an energy-efficient and time-efficient reprogramming method for topologically constrained BAWSNs.

B. ENERGY OPTIMIZATION BASE-STATION SELECTION

In the field of network energy optimization site selection, reference [14] selects the center of gravity of the network as the optimal location of the base station in the wireless network with data traffic equalization, and the anycast base-station selection (ABS) algorithm [15] uses a heuristic method to solve the network lifetime maximization base-station location problem; however, it only obtains an approximate solution. The optimal base-station placement (OBSP) algorithm [16] divides the entire network area into sub-areas by controlling the radius with all the network nodes as centers. Then, it identifies the optimal sub-area where the base station is located after selection. In [17], the network is divided into sub-networks that have the same area, and the optimal location algorithm of the multi-base-station with the most preferred location is independently executed in each sub-area area. Reference [18] proposes a base-station location algorithm that uses the center of the minimum bounding circle of the network as the best location. In [19], a lexicographically stratified programming (LSP) model is proposed for a specified topology. Based on the complete hierarchical sequence and the grid search method, the optimal location of the base station is identified. However, this energy-efficient base-station location algorithm is designed for two-dimensional unconstrained networks of data uplink, whereas the above energy-efficient base-station location algorithm is designed for two-dimensional unconstrained networks of data uplink and is not suitable for BAWSNs reprogramming scenarios of data downlink.

C. TIME OPTIMIZATION BASE-STATION SELECTION

In the field of network base-station location time optimization, the data forwarding delay between nodes and base stations is linearly related to the distance between the nodes and the base stations. Moreover, the data forwarding delay of the network depends on the node that is farthest from the base station. This problem is a typical p-center problem in facility location strategies. In this field of research, Hakimi [20] proposed the p-center problem in the network and Kariv and Hakimi [21] proved that the p-center problem is an NP-hard problem. Megiddo and Supowit [22] also proved that the p-center problem that is based on the linear distance and the Euclidean distance is an NP-complete problem. Minieka [23] studied the p-center location problem in the case of $p > 1$ and Minieka *et al.* proposed a heuristic

algorithm for the p-center coverage problem. Based on this, Gonzalez [24] proposed an approximation algorithm, namely, the Gon algorithm, which has complexity $O(kn)$. The algorithm mainly adopts a greedy strategy; however, it sacrifices the location effect by randomly selecting the initial node to reduce the time complexity. Hochbaum and Shmoys [25] proposed an approximation algorithm that is based on parameter pruning according to the similarity between the p-center coverage problem and the dominating set problem. Chen and Chen [26] proposed a binary relaxation iterative (BRA) algorithm for network subset search that is based on the research of Minieka et al. A binary search is conducted by adding a fixed number of nodes to find a corresponding approximate optimal solution. However, if the network is large, the search space of the nodes becomes extremely complicated. To overcome this problem, reference [27] introduces an extensible relaxation-based iterative algorithm that does not depend on the calculation of the entire matrix, but rather on the calculation of a submatrix that is typically much smaller. This sub-matrix is only amplified when it is deemed necessary, thereby substantially reducing the complexity of the operation. Chen and Chen [28] combined the relaxation method with the p-center coverage problem of α neighbors, which is based on the research of Minieka *et al.* Drezner *et al.* [29] proposed a p-center random location model that is based on hybrid linear programming for solving the problem of emergency facility location. The most preferred address algorithm and the multi-facility optimization algorithm for a single facility have also been proposed.

Other scholars have used heuristic and metaheuristic algorithms to solve p-center location problems, such as the scatter search (SS) algorithm, which was proposed by Pacheco and Casado [30], and the genetic algorithm, which was proposed by Pullan [31]. Levin and Ben-Israel [32] proposed a heuristic algorithm for large-scale p-center problems. Drezner and Drezner [33] presented a heuristic-based greedy random adaptive algorithm for solving the sequential location problem of two facilities. Garcia-Diaz *et al.* [34] proposed a heuristic algorithm with algorithmic complexity for solving the p-center facility location problem. Garcia-Diaz *et al.* [35] used a structure-driven randomization (SDR) method to solve the p-center problem. In the heuristic algorithm for the p-center location problem, Resende, and Ribeiro proposed a multi-start metaheuristic algorithm that is based on the greedy randomized adaptive search procedure (GRASP) algorithm [36]. Callaghan *et al.* [37] and Mihelic and Robic [38] proposed a scoring (Scr) algorithm that is based on the relationship between the p-center problem and the dominating set problem. Albareda-Sambola *et al.* [39] extended the p-center problem according to the practical requirements and established a new hierarchical p-center problem (SpCP) model. A heuristic method that is based on sample average approximation (SAA) is proposed for this model.

The protocol that is proposed in this paper differs from most current reprogramming protocols for two-dimensional unconstrained networks. We mainly focus on the energy and time efficient reprogramming multi-base-station selection problem of BAWSNs with topological constraints. The novelty of the proposed approach is that it identifies the most energy-efficient transmission radius for the topologically constrained strip network, thereby avoiding the large energy consumption that is caused by the reprogramming protocols that are discussed above. However, the use of an energy-efficient transmission radius, which results in an increase in the number of data forwarding times, will require more network normal working hours, thereby causing channel congestion, which will increase the security risks of the underground space. Another innovation in this paper is that we reduce the time for network reprogramming by selecting the optimal base-station location. We are mainly studying the p-center problem for multi-base-station selection in BAWSNs-specific scenarios with topological constraints. According to our theoretical proof and experimental results, our method is more efficient than other methods in the strip network scenario.

III. SYSTEM MODEL

A. SYSTEM MODEL

The entire BAWSNs is represented by $G(V, E)$, wherein $V = \{v_1, v_2, \cdots, v_N\}$ represents a set of nodes and E represents the set of edges. K represents the number of base stations selected from the nodes throughout the OTA process, and the set of base stations is denoted by $B = \{b_1, b_2, \cdots, b_K\}$. G_i denotes a sub-graph composed of a node and a path traversed by the base station b_i throughout the OTA transmission process, and V_i denotes a node set covered by the base station b_i as a base station in the entire reprogramming process, and the final goal is to achieve $\bigcup_{i=1}^{K} V_i = V$.

B. ENERGY CONSUMPTION MODEL AND OPTIMAL TRANSMISSION RADIUS SELECTION

In terms of energy consumption, the minimum transmission energy required per bit of the sensor node is proportional to the square of the distance between the nodes. The power consumption model is shown in Equation 1:

$$\varepsilon_{i,j} = \varepsilon_{amp} d^2_{v_i, v_j} \tag{1}$$

where ε_{amp} represents the power gain (in J/bit/m2) and d_{v_i,v_j} represents the distance between nodes v_i and v_j.

In the wireless network, the node energy determines the service life of the whole network. Therefore, this paper studies energy consumption as the first optimization goal, and the energy consumption in the downlink transmission process is different from the uplink. Since the amount of data in the downlink is fixed by K (in bit), the energy consumption E of each data transmission is as shown in Equation 2:

$$\mathrm{E} = \mathrm{K}\varepsilon_{amp} d^2_{v_i, v_j} \tag{2}$$

According to Equation 2, since the data amount K and the power gain $\mathcal{E}_{\mathrm{amp}}$ are constant, the transmission distance C determines the transmission energy consumption of the real

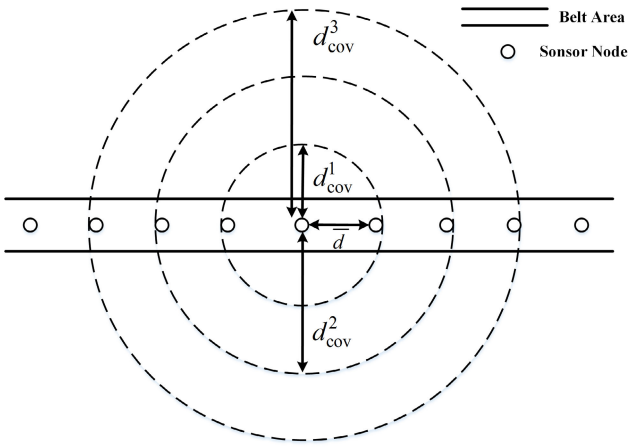

FIGURE 2. Belt-area node coverage schematic diagram.

network. The problem of optimizing the transmission energy consumption of the entire network is transformed to the problem of selecting the optimal forwarding radius. To study this problem, the following definitions are used:

Definition 1: Assuming α is the energy efficiency parameter in the wireless network data transmission process. If the node transmits with radius d_{cov}, the number of nodes that the transmission radius can cover is N_{coV}, then the energy consumed per bit of a node is α, which is expressed as:

$$\alpha = \frac{\varepsilon_{amp} d^2_{\mathrm{cov}}}{N_{\mathrm{cov}}} \tag{3}$$

The smaller α is, the less energy is consumed per bit of transmission, and the higher the energy utilization efficiency.

A very small number of nodes at the belt intersection are removed in the BAWSNs, and the remaining nodes are evenly distributed in the belt-area. Assuming that the distance between adjacent nodes is $\overline{d}$. The coverage of nodes by a node with a specified transmission radius is illustrated in Fig. 2.

As shown by the node coverage of the strip region in Fig. 2, the transmission radius d_{cov} and the number of nodes it covers N_{cov} are related as follows:

$$N_{\mathrm{cov}} = 2\left\lfloor \frac{d_{\mathrm{cov}}}{\overline{d}} \right\rfloor \tag{4}$$

The total amount of code updated during the network update process is K, α is the energy consumed per bit per node, then the total energy consumed by the entire network update process is:

$$\mathrm{E}_{\mathrm{total}} = \mathrm{K}N\alpha \tag{5}$$

The number of updated codes K and the number of network nodes N are both constant, and the minimum energy of the entire network update process can be obtained as long as the energy value α value is minimized.

The relationship between the transmission radius d_{cov} and the number of nodes N_{cov} it covers, given by the main body

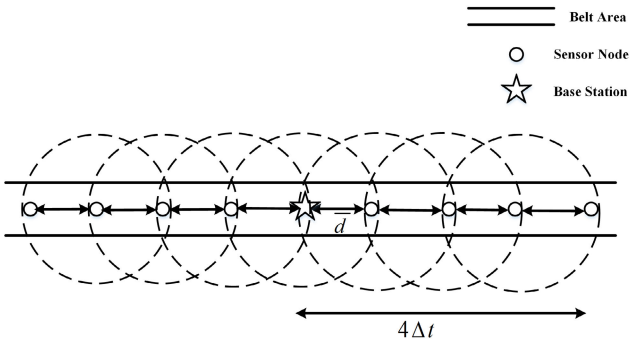

FIGURE 3. Data transmission delay diagram.

equation (4), can be transformed as follows:

$$\begin{aligned} \alpha &= \frac{\varepsilon_{amp} d^2_{\mathrm{cov}}}{N_{\mathrm{cov}}} \\ &= \frac{1}{2}\varepsilon_{amp} d^2_{\mathrm{cov}} \left\lfloor \frac{\overline{d}}{d_{\mathrm{cov}}} \right\rfloor \\ &\leq \frac{1}{2}\varepsilon_{amp} d^2_{\mathrm{cov}} \frac{\overline{d}}{d_{\mathrm{cov}}} \\ &\leq \frac{1}{2}\varepsilon_{anp} \overline{d} d_{\mathrm{cov}} \end{aligned} \tag{6}$$

It can be seen from the above equation that $\mathcal{E}_{amp}$ and $\overline{d}$ are constants, and the size of the energy efficiency parameter α depends on the selection of the transmission radius d_{cov}, and the minimum value of the radius is $\overline{d}$. Based on the above relationship, this conclusion can be get: if the belt monitoring area nodes are evenly distributed, the total energy that is consumed for transmitting the entire network update using the transmission radius $\overline{d}$ is minimal.

The entire BAWSNs is represented by $G(V, E)$, wherein $V = \{v_1, v_2, \cdots, v_N\}$ represents a set of nodes and E represents the set of edges. We select K nodes from the set of nodes, V to be built and denote them as $B = \{b_1, b_2, \cdots, b_K\}$. Assuming that the maximum reliable transmission radius of the node and the distance between adjacent nodes are both $\overline{d}$, d_{v_i,b_j} represents the shortest non-cyclic path length of node v_i and base station b_j.

The data forwarding delay of BAWSNs is directly related to the node forwarding radius and the number of forwarding nodes. If the maximum reliable transmission radius is $\overline{d}$, the time that is taken for data to be forwarded by each intermediate node during network data transmission is Δt. The complete process is illustrated in Fig. 3.

According to the transmission process that is illustrated in Fig. 3, the data transmission time $T(v_i, b)$ of node v_i and base station b_j can be expressed as follows:

$$T(v_i, b) = \frac{d_{v_i}, b}{\overline{d}} \cdot \Delta t \tag{7}$$

The network data forwarding delay that is obtained via equation (1) is linear in the distance between the node and the base station. Therefore, the time optimization problem

is converted into a distance optimization problem. In each round of network data communication, the data forwarding time depends on the farthest node from the base-station set, namely, $B = \{b_1, b_2, \cdots, b_K\}$. Therefore, $d(v, B)$ represents the shortest distance of each base station in the base-station set, namely, B, to a specified node, namely, v, which is expressed as follows:

$$d(v, B) = \min_{v \in V} \left\{ d_{v,b_1}, d_{v,b_2}, \cdots, d_{v,b_K} \right\} \tag{8}$$

Then, according to equation (8), in selecting the base-station set B_i, the largest distance, which is denoted as $d^i_{\max}$, from base-station set B_i among all nodes in the network can be expressed as follows:

$$d^i_{\max} = \max_{v \in V} d\,(v, B_i) \tag{9}$$

The optimization objective in this paper is to select the base station set, namely, B^*, that minimizes the farthest distance, namely, $d^i_{\max}$, from base station set B_i among all the nodes in the network as follows:

$$\begin{aligned} &\min d^i_{\max} \\ &\text{s.t.} \cup c_i v_i = B_i \\ &\quad c_i = \{0, 1\} \\ &\quad \sum c_i = K \\ &\quad 1 \le i \le N \end{aligned} \tag{10}$$

In the above equation, c_i denotes the location factor and subscript i represents the ID of the corresponding selected node, namely, v_i. If node v_i has been selected as the base station location, $c_i = 1$; if it has not been selected, $c_i = 0$. And B_i denotes the set of base stations selected from the set V. Equation (10) shows how to select the appropriate base stations set B_i from the set V such that the distance $d^i_{\max}$ of any node v from its nearest base station is the smallest. The problem that is defined above is a combinatorial optimization problem and the above model must be simplified.

C. SIMPLIFIED MODEL

In the BAWSNs data transmission, for a single base station, the convergence time of each round of data transmission to the base station is related to the longest non-cyclic path between the base station and the node in the entire network. Therefore, the longest non-cyclic path length of the node is defined as follows:

Definition 2: L represents the longest non-cyclic path length between nodes in the entire graph G and can be expressed as follows:

$$L = \arg\max \left\{ d_{v_i,v_j} | v_i, v_j \in V \right\} \tag{11}$$

The following lemma follows from the relationship between Definition 1 and the transmission time, which is expressed in equation (7).

Lemma 1: Under a single base station, the shortest time, namely, T^*, of the entire network data forwarding delay and the longest non-cyclic path length, namely, $L_{\max}$ between the nodes in the entire network G are related as follows:

$$T^* = \left\lceil \frac{L}{2\overline{d}} \right\rceil \cdot \Delta t \tag{12}$$

(The proof is presented in Appendix)

According to Lemma 1, in the problem of efficiently determining a single base station location time, the optimal base station is located halfway down the longest non-cyclic path in network graph G. The problem of location selection is simplified to the longest non-circulating loop problem in the network graph G, for which current research is relatively mature.

D. MULTI-BASE-STATION TIME-EFFICIENT LOCATION MODEL

The multi-base-station time-efficient location model adopts the strategy of divide and conquer. The graph, namely, G, divides the edge set, namely, E, into a subset of connected edges: $\{E_1, E_2, \cdots, E_N\}$. According to the subset, a plurality of the derived subgraph sets, namely, $\{G_1, G_2, \cdots, G_N\}$, are segmented and the middle node set of each derived subgraph is $\{V_1, V_2, \cdots, V_N\}$. To divide the graph G into K subgraphs, K base stations are selected, namely, K subgraphs are selected from the plurality of connected derived subgraph sets, namely, $\{G_1, G_2, \cdots, G_N\}$, such that the K subgraph node sets satisfy $\cup V_i = V$. According to Lemma 1, the subgraph data transmission time, namely, T_i, is determined by the longest non-cyclic path length, namely, L_i, in its subgraph. The optimal selection method for the shortest non-cyclic path L_i in each derived subgraph G_i and the shortest time for solving the network update under the multi-base-station distribution can be rewritten as follows:

$$\begin{aligned} &\min L^i_{\max} \\ &\text{s.t. } c_i = \{0, 1\} \\ &\quad \cup c_i V_i = V \\ &\quad \sum_1^N c_i = K \\ &\quad 1 \le i \le N \end{aligned} \tag{13}$$

The above problem is difficult to solve due to too few constraints and too large a search space for polynomial time. According to the characteristics of the graph G, the constraints of the optimization target of the above problem can be further tightened. First, from Definition 1 and conclusion, the following lemma can be obtained:

Lemma 2: If the longest non-cyclic path in an undirected simple graph G is L, if G is divided into K derived subgraphs, namely, G_i, and if L_i is the length of the longest non-cyclic path in subgraph G_i, then $\sum_{i=1}^{K} L_i \ge L$. (The proof is presented in Appendix)

Lemma 3: If the graph G is divided into subgraphs of K longest non-cyclic path lengths, the longest subgraph length, namely, L_i, for each subgraph G_i is: $L_i \ge \left\lfloor \frac{L}{K} \right\rfloor$. (The proof is presented in Appendix.)

From Lemma 3, we conclude that the lower limit of the optimal solution of the problem is $\forall i \neq j, L_i = L_j = \lfloor \frac{L}{K} \rfloor$. Then, the problem can be transformed into a subgraph set, namely, $\{G_1^*, G_2^*, \cdots, G_K^*\}$, that divides the graph $G(V, A)$ into K longest non-cyclic paths that are each of length $(1 + \Delta) \lfloor \frac{L}{K} \rfloor$, which corresponds to the longest non-circulating path length when the optimal subgraph is divided. To solve this problem, we specify a set of subgraph nodes, namely, $S = \{S_1, S_2, \cdots, S_N\}$, that are centered on each node in graph G and have a path length of $(1 + \Delta) \lfloor \frac{L}{2K} \rfloor$ to the node, where S_i can be formally defined as follows:

$$S_i = \{v_i\} \cup \left\{ v_j | d_{v_i, v_j} \leq (1 + \Delta) \left\lfloor \frac{L}{2K} \right\rfloor \right\} \tag{14}$$

Then, the problem can be expressed as follows:

$$\begin{aligned} \min \quad & \max L_i \\ & c_i = \{0, 1\} \\ & V = \bigcup_{S_i \subseteq V} c_i S_i \\ s.t. \quad & \sum_1^N c_i = K \\ & S_i = \{v_i\} \cup \left\{ v_j \left| d_{v_i v_j} \leq (1 + \Delta) \right| \frac{L}{2K} \rfloor \right\} \\ & 1 \leq i \leq N \end{aligned} \tag{15}$$

In the above equation, S_i is a set of nodes that were selected within the optimal radius $\frac{L_i^*}{2}$ centered on v_i; hence, v_i is the optimal base station position that corresponds to node set S_i. Then, for the optimal node set S_i^*, which satisfies the optimality condition that is specified above, the set $\{v_1^*, v_2^*, \cdots, v_K^*\}$ of the central nodes v_i^* that correspond to each optimal node set S_i^* is the final optimal base-station set: $\{b_1^*, b_2^*, \cdots, b_K^*\}$.

E. NP-COMPLETENESS PROOF

To prove the NP-completeness property of problem (13), we must first prove that the problem belongs to the NP class of problems. If it can be reduced to an existing NP-complete problem, it is also NP-complete because it can be proven to satisfy the NP-completeness conditions without loss of generality.

To prove that the problem is NP-complete, we prove that it reduces to the subset sum problem, which defined as follows: For a finite set $X \subset N$ and a known target $t \in N$, does there exist a subset $X' \subseteq X$ such that t is equal to the sum of its elements? This problem is expressed as follows:

SUBSET-SUM $= \langle X, t \rangle$:there is a subset $X' \subseteq X$ that satisfies $t = \sum_{x \in X} x$.

Then, problem (15) is defined as the following NP-complete problem: K-GRAPHCUT $= \langle S, V \rangle$:there is a subset $S' \subseteq S$ of elements K that satisfies $V = \bigcup_{v_i \in S'} S_i$.

Proof: First, we prove that problem (11) belongs to the class of NP problems. For an instance of the problem, namely, $\langle S, V \rangle$, let subset S' be the optimal subset. With a verification algorithm, it is possible to check whether $V = \bigcup_{v_i \in S'} S_i$ in polynomial time. As shown in Fig. 4, (b) and (c) illustrate the optimal cuts into two subgraphs of (a). The correctness can be verified within polynomial time complexity; hence, K-GRAPHCUT is an NP problem.

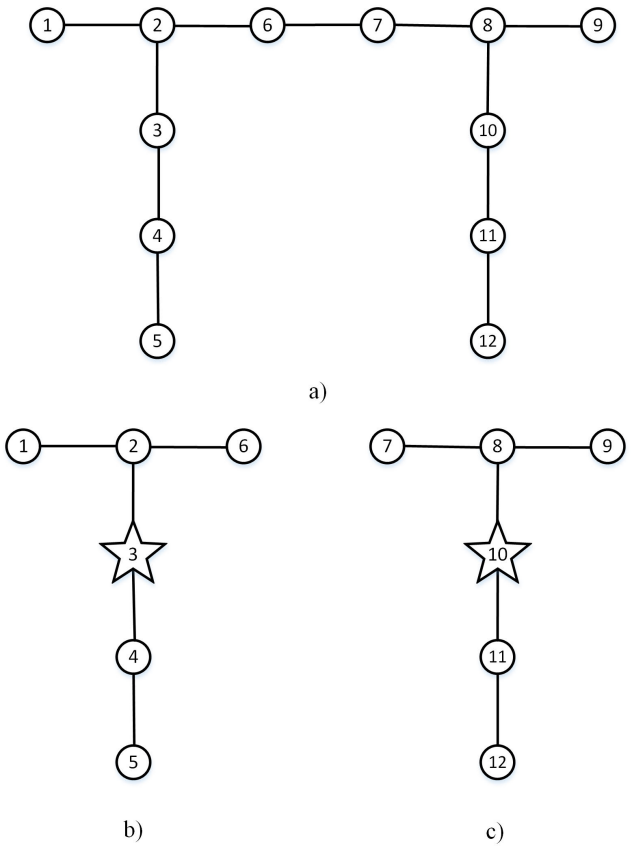

FIGURE 4. Schematic diagram of subgraph segmentation and division.

Next, it is reduced to the subgraph problem according to the corresponding constraint. Then, we only need to prove that SUNBSET-SUM $\leq_p$ K-GRAPHCUT: Given a set of problem variable sets $X = \{x_1, x_1, \cdots, x_n\}$, there exists a subset $X' \subseteq X$ such that the sum of its elements is t. The statistical algorithm constructs an instance, namely, $\langle S, V \rangle$, of K-GRAPHCUT such that there exists $S' \subseteq S$ that satisfies $V = \bigcup_{S_i \in S'} S_i$.

The following SUNBSET-SUM is reduced to problem (15) via the following two steps, which can be completed in polynomial time. First, the following two constraints are satisfied, as illustrated in Fig. 5:

<1> The subset selection parameter, namely, c_i, in problem (13) corresponds one-to-one with the elements in set $\{x_1, x_1, \cdots, x_n\}$;

<2> The value of c_i satisfies the following conditions.

If K-GRAPHCUT has an optimal subgraph partitioning set $\{V_1^*, V_2^*, \cdots, V_K^*\}$ with K partitions, according to the constraint in problem (11), there must exist sets $\{c_1^*, c_2^*, \cdots, c_K^*\}$ and $\{V_1^*, V_2^*, \cdots, V_K^*\}$ that are in one-to-one correspondence and satisfy $\mathbf{c}_i^* = 1$, and according to the constraints in conditions <1> and <2>, set $\{c_1^*, c_2^*, \cdots, c_K^*\}$ must correspond to $\{x_1^*, x_2^*, \cdots, x_K^*\}$ and $x_i^* = 1$ must be satisfied. Therefore,

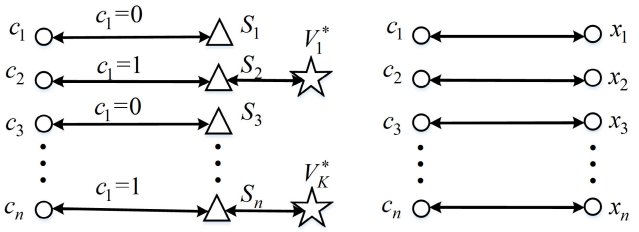

FIGURE 5. Node and base station association graph for the subset problem.

the optimal partition subset, namely, $\{V_1^*, V_2^*, \cdots, V_K^*\}$, of problem (15) satisfies constraints <1> and <2>. Thus, $t = K$, that is, for the set $\{x_1, x_1, \cdots, x_n\}$, there is a subset $X' \subseteq X$ that satisfies the constraint, and $K = \sum_{x \in X'} x$ is satisfied. To obtain the solution, namely, $\{V_1^*, V_2^*, \cdots, V_K^*\}$, of the K-GRAPHCUT in problem (15), it is necessary to solve the subset problem, namely, to satisfy $K = \sum_{x \in X'} x$, where the number of elements in subset X' is K, which is expressed as $|X'| = K$, which follows from the NP-completeness of the subset problem. K-GRAPHCUT in problem (15) is also NP-complete. Thus, the proposition is proved.

IV. DESIGN AND ANALYSIS OF ALGORITHMS

A. ACCURATE ALGORITHM WITH EXPONENTIAL TIME COMPLEXITY

Suppose that for each node v_i in G, the set of nodes within radius $\lfloor \frac{L}{2K} \rfloor$ from the base station is calculated by v_i. Next, a full combination of K base stations is selected from n nodes and a combination of $\cup S_i = V$ is selected among all the combined base-station combinations.

In this paper, an accurate algorithm with exponential time complexity is proposed. First, all the subgraph sets, namely, $S = \{S_1, S_2, \cdots, S_N\}$, in the graph G that are centered on each node and have path length $\lfloor \frac{L}{2K} \rfloor$ to the node are specified. Then, K base stations are selected from n nodes. All combinations of sets

$$R_n^K = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1K} \\ r_{21} & r_{22} & \cdots & r_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\varphi 1} & r_{\varphi 2} & \cdots & r_{\varphi K} \end{bmatrix}$$

are formed, where $\varphi = C_n^K$ and r is the node ID number and the merged set, which is denoted as O_i^K, is calculated via the following expression:

$$O_i^K = O_i^{K-1} \cup S_{r_{i,K}} \tag{16}$$

The solution method for $S_{r_{i,K}}$ is based on the distance matrix, namely,

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & d_{3n} \\ d_{n1} & d_{n2} & \cdots & d_{mn} \end{bmatrix},$$

between the nodes in the network and the longest non-circular path length, namely, L, of the network. According to equation (14), $S_i = \{v_i\} \cup \left\{v_j | d_{v_i,v_j} \leq \left[\frac{L}{2K}\right], v_i, v_j \in V\right\}$: centring on the node v_i, row i is traversed to identify all nodes v_j such that $d_{v_i,v_j} \leq \lfloor \frac{L}{2K} \rfloor$ and the set of all nodes v_j that satisfy this condition is the set S_i.

Algorithm 1 : The Optimization Algorithm of the Model

Input: $A = (a_{i,j})_{n \times n}$: Initial adjacency matrix;
$D = (d_{i,j})_{n \times n}$: Distance matrix;
Output: S^*:Final base station location set.
Initialization: $O_i^0 \leftarrow \varnothing$;
$n = C_N^K$: The iteration of the algorithm;
R_N^K: Select all combinations of base stations from the set of nodes;

```
EXACT-BASE (R_m^K);
for i = 1 to n do
    for j = 1 to K do
        if O_i^K == V then
            return r_i1, r_i2, ..., r_iK
        end if
    end for
end for
```

For the above algorithm, the time complexity of the outermost loop is $O\left(C_n^K\right)$, namely, $O\left(\frac{n!}{K!(n-K)!}\right)$; the time complexity of the second-layer loop is $O(K)$; and the complexity of the $S_{r_{i,j}}$ algorithm is $O(n)$. Therefore, the time complexity of the entire algorithm is $O\left(\frac{K \cdot n \cdot n!}{K!(n-K)!}\right)$, which can be reduced to $O\left(\frac{n \cdot n!}{(K-1)!(n-K)!}\right)$. When the value of K is small, the time complexity is polynomial. When K is large, it is exponential. In the problem that is studied in this paper, the underground area is larger. To reduce the data transmission time, many base stations are required as the initial transmitting nodes of the data transmission code and the value of K must be large; hence, the complexity is exponential for the problem that is studied in this paper.

B. GRADIENT LOOP-BASED $(1 + \varepsilon)$ ALGORITHM

For the above algorithm, which is of exponential time complexity, it is necessary to traverse the entire base-station combination and its coverage space. The search space is large, and the convergence speed is too low for the optimal solution to be obtained in a short time. For converging to the optimal solution as quickly as possible, this paper presents a $(1 + \varepsilon)$ approximation algorithm that is of polynomial time complexity and is based on gradient descent. The algorithm mainly improves the convergence speed from two aspects: the gradient descent algorithm is used to reduce the search space for the base-station combination and the base-station coverage radius is adjusted to increase the convergence speed of the algorithm.

The core strategy of gradient descent is to find the optimal solution along the direction of gradient descent. It is formally expressed as follows:

$$a_{k+1} = a_k + \rho_k \bar{s}^{(k)} \tag{17}$$

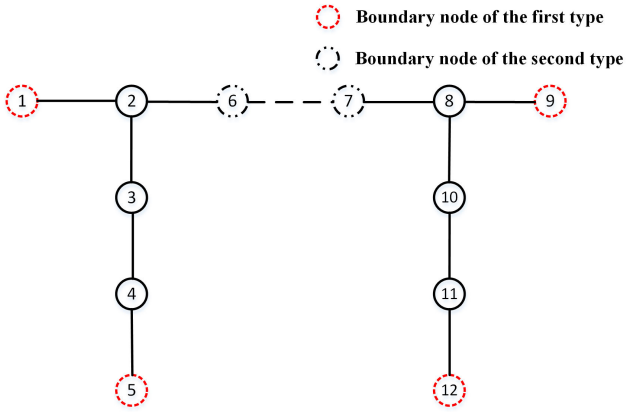

FIGURE 6. Graph of the boundary nodes.

where $\overline{s}^{(k)}$ represents the negative direction of the gradient and ρ_k represents the search step size. For each round of the search process for the problem that is studied in this paper, once the initial base station, namely, b_0, has been determined, it is necessary to select an appropriate gradient direction, namely, $\overline{s}^{(k)}$, and search with step length ρ_k among the remaining nodes. Finally, the algorithm identifies the optimal base station as quickly as possible.

1) SELECTION OF THE STEP SIZE ρ_K FOR THE GRADIENT ALGORITHM

Suppose the nodes at the ends of the longest non-cyclic path of the subgraph are v_p and v_q. According to Lemma 3, $d_{b_i,v_p} = d_{b_i,v_q} \geq \lfloor \frac{L}{2K} \rfloor$. Then, for the problem that is studied in this paper, G_i is divided based on base station b_i and the subgraph that is obtained by dividing the radius by d_{q_i,v_p}. According to the distance between the base station b_i and the boundary nodes in subgraph G_i, the boundary nodes are divided into the following two cases:

(1) The first is the set of boundary nodes of graph G that are contained in graph G_i. However, not all the adjacent nodes of these nodes will become connection points between the two subgraphs in graph G_i. Therefore, the distance between the base station and such boundary nodes is independent of the distance between the two subgraph base-stations. This is beyond the scope of our research;

(2) The second set of boundary nodes is divided into subgraphs and represented by $Z = \{z_1, z_2, \cdots, z_k\}$. Among the adjacent nodes, there are nodes that belong to another subgraph, as shown in illustrated 6:

According to the above analysis, for subgraph G_i, the distance between base station b_i and any node z_j in the set of boundary points of the second category satisfies $d_{b_i,z_j} \geq \lfloor \frac{L}{2K} \rfloor$ and the following inference can be drawn:

Inference 1: If b_i and b_j are two adjacent optimal base stations in graph G, then the distance, namely, d_{b_i,b_j}, between b_i and b_j satisfies $d_{b_i,b_j} \geq \lfloor \frac{L}{K} \rfloor$.

From inference 1, it follows that the lower limit of the value of search step ρ_k is $\lfloor \frac{L}{K} \rfloor$. Via this approach, the combination case in which the distance between adjacent nodes is less than $\lfloor \frac{L}{K} \rfloor$ is excluded from the total combination space of nodes, which substantially reduces the size of the search space.

2) $(1+\varepsilon)$ APPROXIMATELY OPTIMAL SELECTION OF THE OPTIMAL SEARCH STEP SIZE ρ_K^*

The above inference only specifies the lower limit of the search step; it does not specify the optimal search step size. It remains necessary to filter the optimal search step size to obtain the base-station distribution under the optimal search step size, which is costly. For the selection of the optimal step size, the $(1+\varepsilon)$ approximately optimal selection method is proposed.

According to Inference 1, the longest non-circulating path distance between adjacent optimal base stations b_i^* and b_j^* is the optimal step length ρ_k^*, namely, $\rho_k^* = d_{b_i^*,b_j^*}$. In the problem that is studied in this paper, the graph G is divided into K subgraphs that have the longest non-cyclic path, are of equal length, and minimize the longest non-cyclic path of each subgraph. To ensure that the longest path of the subgraph is as short as possible, repeated nodes should be avoided between the subgraphs. It is assumed that z_i is a subgraph that is centered on base station b_i^* to obtain a subgraph boundary node of the first type and z_j is a subgraph that is centered on base station b_j^* for obtain a subgraph boundary node of the second type. According to the division of the subgraph, for the subgraph that is obtained by dividing G with the base station as the center and the boundary node of the second type, which is denoted as d_{b_i,z_i}, as the radius, the following inferences are drawn:

Inference 2: Suppose z_i divides the subgraph with the base station b_i^* as the center to obtain a subgraph boundary node of the second type and suppose z_j is a subgraph that is centered on the base station b_j^* and is used to obtain a subgraph boundary node of the second type. Then, $d_{b_i^*,z_i}$ and $d_{b_j^*,z_j}$ are the radii of the subgraphs of base stations b_i^* and b_j^*, respectively, which satisfy the following:

$$d_{b_i^*,z_i} = d_{b_j^*,z_j} = \frac{d_{b_i^*,b_j^*}}{2} \tag{18}$$

For the problem that is studied in this paper, the shorter the distance d_{b_i,b_j} between adjacent base stations, the shorter the time consumption for data transmission; however, the shorter the search step, the slower the search convergence. According to Inference 1, the theoretically optimal step size is $\lfloor \frac{L}{K} \rfloor$; however, in searching with step size $\lfloor \frac{L}{K} \rfloor$ in an underground wireless network, the optimal base-station distribution is not realized and the convergence speed is too slow.

To increase the convergence speed of the algorithm, the search step is reduced to $(1+\varepsilon)\lfloor \frac{L}{K} \rfloor$ and $H(\lfloor \frac{L}{K} \rfloor)$ is used to denote the convergence time when the search step size is $\lfloor \frac{L}{K} \rfloor$. $H\left((1+\varepsilon)\lfloor \frac{L}{K} \rfloor\right)$ denotes the convergence time under the reduced search step size. The following lemma compares $H\left((1+\varepsilon)\lfloor \frac{L}{K} \rfloor\right)$ with the original convergence time:

Lemma 4: The following inequality holds:

$$H\left(\left\lfloor \frac{L}{K} \right\rfloor\right) \leq (1+\varepsilon) H\left((1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor\right) \tag{19}$$

Lemma 5: If the network topology of the graph G has been determined, the data transmission time $T\left((1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor, S\right)$, under the base-station distribution that is obtained by the reducing search step size to $(1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor$ and the data transmission time $T\ (d^*, S^*)$ under the optimal base-station distribution are related as follows:

$$\frac{T\left((1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor, S\right)}{T\ (d^*, S^*)} \leq (1+\varepsilon) \tag{20}$$

According to Lemma 4 and Lemma 5, gradient search with the reduced step size of $(1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor$ can yield higher search efficiency; however, the accuracy of the algorithm is reduced. To more accurately approximate the optimal step size between adjacent base-stations, once the topology of the graph G network has been determined, there must exist a set of optimal base station sets, namely, S^*. Subgraph partitioning with the optimal radius, namely, r^*, centered on each base station in S^* minimizes the longest non-cyclic path in each subgraph. According to Lemma 2, the optimal radius satisfies $r^* \geq \left\lfloor \frac{L}{2K} \right\rfloor$. According to Inference 2, the relationship between the division radius and the distance between the adjacent base stations is $d^* = 2\ r^*$ and $d^* \geq \left\lfloor \frac{L}{K} \right\rfloor$. It is assumed that a set of optimal base stations is obtained using the reduced step size of $(1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor$ and all nodes can be covered by the division radius of $(1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor$; however, the shortest division radius is not guaranteed. Combined with the above relationships, it follows from $(1+\varepsilon)\left\lfloor \frac{L}{2K} \right\rfloor \geq r^*$ that:

$$\left\lfloor \frac{L}{K} \right\rfloor \leq d^* \leq (1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor \tag{21}$$

According to the clamping criterion, the value of ε is continuously adjusted to approximate the optimal distance d^* of the adjacent base station and the iterative process of its parameters is expressed as follows:

$$\varepsilon^{(k)} = \varepsilon^{(k-1)} - \Delta \tag{22}$$

The following conditions are satisfied:

$$\varepsilon^* = \arg(1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor = d^* \tag{23}$$

where Δ is the $\mathcal{E}$ adjustment parameter. The distance between adjacent nodes is assumed to be 1, which is also the minimum unit of the division radius; hence, the adjustment condition of parameter Δ is as follows:

$$\begin{aligned} (1+\varepsilon-\Delta)\left\lfloor \frac{L}{K} \right\rfloor &= (1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor - 1 \\ \Delta \cdot \left\lfloor \frac{L}{K} \right\rfloor &= 1 \\ \Delta &= \left\lceil \frac{K}{L} \right\rceil \end{aligned} \tag{24}$$

3) SELECTION OF GRADIENT DESCENT DIRECTION $\overline{s}^{(k)}$

After the search step has been determined, the selection of the gradient direction directly affects the speed at which the algorithm converges. Therefore, the greedy strategy is adopted for selecting the gradient descent direction, which is denoted as $\overline{s}^{(k)}$, and the number of nodes is constant for the graph G. After the initial base station has been selected, the candidate base station that is adjacent to the initial base station is searched using step size ρ_k as the search radius. To avoid duplicate coverage of nodes in neighboring base stations, the neighboring base stations cover the network nodes with a radius of $\frac{\rho_k}{2}$. The base stations that can cover additional nodes are the optimal direction for gradient descent, namely, $\overline{s}^{(k)}$ solves the following optimization problem:

$$\begin{aligned} &\max\left(N^i_{cover}\right) \\ &s.t. \quad s_i = \{v_j | d_{b_i,v_j} \leq \frac{\rho_k}{2}\} \\ &\qquad N^i_{cover} = |s_i| \end{aligned} \tag{25}$$

where N_{cover} is the number of nodes that the next base station covers with a radius of $\frac{\rho_k}{2}$.

4) INITIAL BASE STATION SELECTION

After the search step size, namely, ρ_k, and the gradient descent direction, namely, $\overline{s}^{(k)}$, have been determined, the selected position of the initial base station, namely, b_0, determines the positions of the remaining $K-1$ base stations, which is essential for location selection. According to Inference 2, the distance between base station b_i and boundary node of the second type z_i of the optimally partitioned subgraph and the search step size, namely, ρ_k, are related as $d_{b_i,z_i} = \frac{\rho_k}{2}$. It is shown that for each subgraph, the optimal base-station position is of distance $\frac{\rho_k}{2}$ from the boundary node. Combining with the greedy gradient descent direction, it is desirable to have as many base stations as possible to cover additional nodes with the division $\frac{\rho_k}{2}$ when selecting the base station. This involves finding the boundary nodes of the first type in a subgraph. The boundary nodes of the first type are boundary nodes of graph G. When they are divided into subgraphs, the base station cannot divide further than by $\frac{\rho_k}{2}$. These nodes cause insufficient division; therefore, when selecting the initial base station, a node that is separated from the boundary nodes of the first type by $\frac{\rho_k}{2}$ is selected, as illustrated below.

C. ALGORITHM DESCRIPTION

The solution steps and pseudocode of the $(1+\varepsilon)$ approximately optimal algorithm that is based on gradient descent are summarized as follows:

1) First, the network topology and distance matrix of graph G are specified.

2) Identify the boundary nodes of the first type in graph G.

3) In the first round of siting, the step-size lower limit $\left\lfloor \frac{L}{K} \right\rfloor$ and the step length ρ_0 are determined according to the longest path L and the number of base-station address K in the figure.

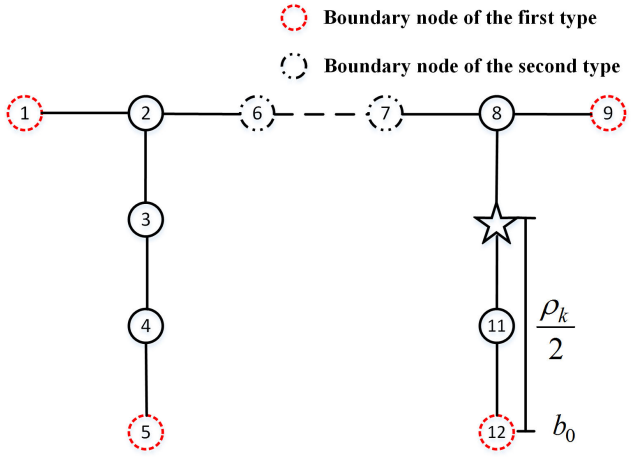

FIGURE 7. Initial base station selection graph.

4) According to Inference 1, the initial base station b_0 is determined in conjunction with the initial step size ρ_0 and the boundary nodes of the first type.

5) Select the gradient descent direction $\overline{S}^{(1)}$ according to equation (21).

6) According to the gradient descent algorithm, for which the iteration step is expressed as $b_{k+1} = b_k + \rho_k \overline{s}^{(k)}$, the best base station set $S^{(1)}$ is identified.

7) The base station set $S^{(1)}$ is judged. If the current set of base stations is centered on $S^{(1)}$ elements and the set of nodes covered by $\lfloor \frac{L}{2K} \rfloor$, which is the coverage radius is V, the current base station set is output as the best base-station set. Otherwise, proceed to the next step.

8) The update search step length $\rho_k = (1 + k\Delta)\rho_0$ is iteratively modified by applying equation (18) to modify the parameter k. Then, proceed to 3) to perform the above steps to select the base station $S^{(k)}$ under the optimal search step. Finally, apply 7) to obtain a set of optimized base stations that satisfy the conditions.

The pseudocode for the entire algorithm is as follows:

Algorithm 2 : Gradient Descent Location Algorithm

// Δ : Adjustment parameters
// ρ_k : Search step size
// k : Step growth
// $BS\_selected[K]$: Base station selection set
// $Full\_cover[n]$: Node coverage set
// $Node\_cover[k]$: Number of nodes with radius $\frac{\rho_k}{2}$ covered by v_j
// $CoverNum_{max}$: Maximum number of coverage
// b_0 : First type boundary node
Input: $A = (a_{i,j})_{n\times n}$: Initial adjacency matrix;
$D = (d_{i,j})_{n\times n}$: Distance matrix;
Output: S^*:The final set of base-station locations.
Initialization: $\Delta = \lceil \frac{K}{L} \rceil$, $\rho_k = (1 + k\Delta)\lfloor \frac{L}{K} \rfloor$, $k = 0$, $BS\_selected[K] = zeros[K]$, $N[n] = 0$,$N_{\max} = 0$, $Full\_cover[n] = zeros[n]$

```
while Full_cover[n] ≠ ones[n] do
  for i = 1 to K − 1 do
    for j = 1 to n do
      for u = 1 to n do
        if d_{S[i],v_j} = ρ_k and d_{v_j,v_u} ≤ ρ_k/2 then
          Node_cover[j] = Node_cover[j] + 1
        end if
        if Node_cover[j] ≥ CoverNum_max then
          CoverNum_max = Node_cover[j]
          BS_selected[i + 1] = j
        end if
        for W = 1 to n do do
          if d_{S[i],v_w} ≤ ρ_k/2 then
            Full_cover[w] = 1
          end if
        end for
      end for
    end for
  end for
  k = k + 1
  ρ_k = (1 + kΔ)⌊L/K⌋
end while
return BS_selected[K]
```

The complexity of the above algorithm is mainly attributed to the first filtering of the three-layer loop of the optimal base-station location set when the search step size is $(1 + \varepsilon)\lfloor \frac{L}{K} \rfloor$. The innermost loop is used to determine the direction of gradient descent and the selection is based on the number of covered nodes; the algorithmic complexity is $O(n)$. The second-layer loop mainly identifies the candidate base-station node, for which the distance, according to the previous iteration, is $(1 + \varepsilon)\lfloor \frac{L}{K} \rfloor$; the algorithmic complexity is $O(n)$. The outermost loop mainly controls the number of base stations; and its algorithmic complexity is $O(K)$.

Therefore, the complexity of filtering the three-layer loop of the optimal base station location set is $O\left(Kn^2\right)$ if the search step size is $(1+\varepsilon)\lfloor \frac{L}{K} \rfloor$. In addition, to determine whether the selected base station set can completely cover the graph G, it is determined in the second layer loop that the K base-stations that can be selected are centered on the node that is covered by the partition radius $(1 + \varepsilon)\lfloor \frac{L}{K} \rfloor$ and the algorithmic complexity is $O(Kn)$. Therefore, the complexity of the optimal base station location set algorithm is $O\left(Kn^2 + Kn\right)$ if the search step size is $(1 + \varepsilon)\lfloor \frac{L}{K} \rfloor$.

To identify the optimal base-station location, a recursive algorithm is applied. Each time a set of base station, namely, $S^{(i)}$, is obtained with a step size of $(1 + \varepsilon)\lfloor \frac{L}{K} \rfloor$, a recursion must be performed in the address selection algorithm. If the number of recursions is M, the complexity of the whole algorithm is $\sum_{i=1}^{M}\left(Kn^2 + Kn\right)$. The adjustment of each parameter by $\Delta = \lceil \frac{K}{L} \rceil$ changes the search step as $\rho_{k+1} = \rho_k - 1$ and the search step satisfies $\rho_{k+M} = \rho_k - M \geq 0$, namely, $\rho_k \geq M$. Due to the properties of the graph G, the maximum value of

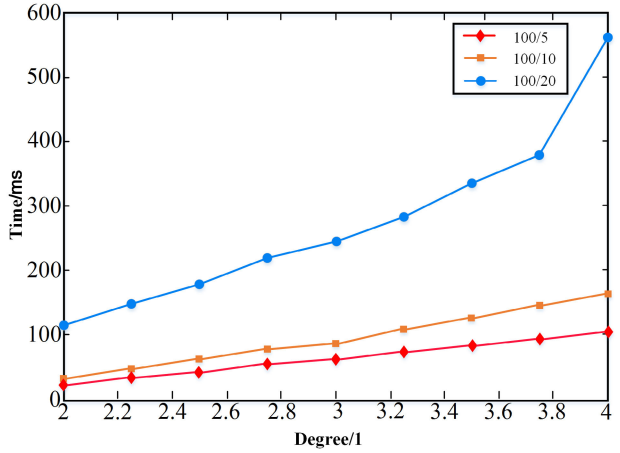

FIGURE 8. Time performance of the OMSS algorithm as a function of the network complexity.

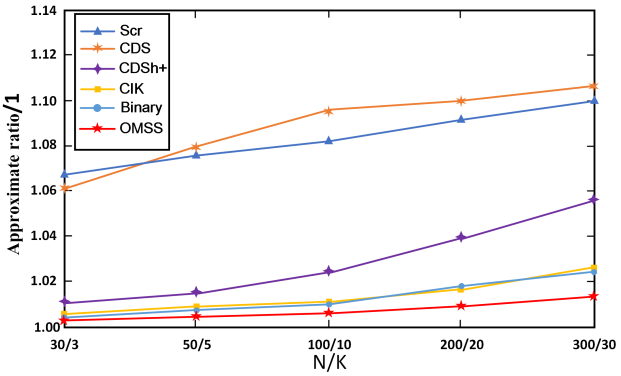

FIGURE 9. Comparison of algorithm location accuracies under various numbers of nodes and numbers of base stations.

the search step is the longest non-circulating path length l of G, namely, $\rho_k \leq l$, and $M \leq \rho_k \leq l$ is satisfied; thus, $\sum_{i=1}^{M} \left(Kn^2 + Kn\right) \leq l\left(Kn^2 + Kn\right)$.

Therefore, the complexity of the complete algorithm is $O\left(lKn^2 + lKn\right)$, where lK is a constant term and $O(lKn)$ can be neglected if n is large. In summary, the complexity of the complete algorithm is $O\left(kn^2\right)$.

V. SIMULATION EXPERIMENTS AND ANALYSIS

In this paper, MATLAB 2018b which developed by MathWorks was used for algorithm simulation verification. A CPU was used with an Intel core i7-6700HQ processor with a main frequency of 2.6 GHz and a maximum frequency of 3.5 GHz. The GT-ITM Topology Generator was used to generate the network topology. According to the number of nodes, the number of selected base stations, and the network complexity, a comprehensive simulation verification of the performance of the OMSS algorithm is conducted. The complexity of the network is calculated by dividing the sum of the degrees of all nodes in the network by the number of vertices, as expressed in equation (26).

$$\theta = \frac{\sum_{i=1}^{n} b_i}{n}, \theta \geqslant 1 \tag{26}$$

where b_i represents the degree of node n_i and θ represents the network complexity. We select 100 nodes and compare the influence of the network complexity on the performance of the algorithm under the same number of nodes and various numbers of base stations by adjusting θ and the number of base stations. The results are plotted in Fig. 8.

According to Fig. 8, the calculation time of the algorithm is proportional to the network complexity: the higher the network complexity, the longer the calculation time. The number of base station selections also has a substantial impact on the performance of the algorithm. If the number of network nodes is the same as the network complexity, the more base-station selections, the longer the algorithm calculation time.

To further evaluate the effects on the time performance of the OMSS algorithm of the number of nodes, the number of selected base stations and the network complexity, the time for selecting the optimal base station with the OMSS algorithm is measured under various values of the variables. The results are listed in Table 1.

According to the data in Table 2, when other conditions are unchanged and the number of nodes is increased, the growth rate of the algorithm calculation time is lower than the growth rate of the number of nodes and the algorithm has higher stability relative to the number of nodes.

To evaluate the accuracy of the base-station location selection algorithm, we compare it with various p-center location algorithms: the classical algorithm for the p-center problem, namely, the Scr algorithm; the CDS algorithm; and the CDSh + algorithm, which is based on CDS and for which the source code is available from (https://github.com/jesgadiaz/vertex-k-center). The source code for the Binary algorithm is obtained from the URL (https://www.tau.ac.il/∼chenr/or_research.html) of R. Chen. To objectively evaluate the performances of the algorithms for a network with a specified topology, the minimum-maximum distance, namely, d_i^*, in each algorithm for selecting the optimal base station must be compared with the d_{EXACT}^*-row under the optimal base station that is selected by the accurate exponential-time algorithm; the ratio is called the approximate optimal ratio, which is denoted by μ and formally expressed as follows:

$$\mu_i = \frac{d_i^*}{d_{EXACT}^*} \tag{27}$$

In this paper, the numbers of nodes and base stations are selected to evaluate the location performance of the algorithm more comprehensively. The OMSS algorithm and the five classical algorithms are run on the same network topology to compare their location performances. Moreover, to ensure the credibility of the results under the same topology, each algorithm was run 10 times and the average value was used to evaluate the performance. The results are plotted in Fig. 9.

TABLE 1. Time performance of the OMSS algorithm under various numbers of nodes, network complexities, and numbers of base stations.

K	N	θ_1	Average time (ms)	Shortest time (ms)	θ_2	Average time (ms)	Shortest time (ms)	θ_3	Average time (ms)	Shortest time (ms)
3	10	2	3.21	2.05	2.5	3.83	2.11	3	4.13	2.38
3	15	2	3.81	2.34	2.5	4.09	2.43	3	6.09	3.57
3	20	2	4.84	2.88	2.5	5.92	2.92	3	6.85	4.05
3	25	2	4.99	3.02	2.5	7.25	3.45	3	8.37	5.17
3	30	2	5.13	3.22	2.5	9.87	4.79	3	11.48	6.28
5	30	2	8.61	4.37	2.5	18.04	10.31	3	27.78	15.49
5	35	2	9.94	4.99	2.5	20.81	12.56	3	34.74	18.89
5	40	2	11.75	5.96	2.5	21.65	15.77	3	41.54	23.55
5	45	2	13.15	6.45	2.5	24.47	18.91	3	46.61	28.63
5	50	2	16.01	8.31	2.5	27.58	20.14	3	51.79	35.11
10	50	2	38.02	23.12	2.5	50.06	32.47	3	68.13	40.84
10	60	2	41.03	31.67	2.5	60.23	36.98	3	71.25	47.91
10	70	2	47.34	35.33	2.5	66.25	43.59	3	76.53	52.88
10	80	2	51.57	39.58	2.5	78.37	52.81	3	84.63	63.76
10	90	2	54.65	42.91	2.5	79.85	59.46	3	94.67	70.32
10	100	2	56.13	46.61	2.5	81.34	63.04	3	110.38	84,23
20	100	2	120.36	98.59	2.5	173.36	143.43	3	267.45	189.98
20	150	2	157.39	110.47	2.5	220.91	167.71	3	321.82	243.76
20	200	2	164.74	116.95	2.5	253.27	182.29	3	394.39	277.64
20	250	2	166.34	121.77	2.5	312.31	230.73	3	452.89	343.17
20	300	2	167.91	128.92	2.5	345.41	271.29	3	525.67	434.95

TABLE 2. Performance comparison of OMSS with other classical algorithm.

K	N	Binary		CIK		SAA		OMSS	
		Average time (s)	Shortest time (s)	Average time (s)	Shortest time (s)	Average time (s)	Shortest time (s)	Average time (s)	Shortest time (s)
500	5	0.18	0.1	0.83	0.26	0.14	0.06	**0.09**	**0.03**
700	5	0.26	0.09	1.04	0.64	0.19	**0.04**	**0.11**	0.05
1000	5	0.31	0.13	2.17	1.58	0.23	0.12	**0.13**	**0.06**
2000	5	3.12	1.71	3.79	1.89	6.05	4.78	**2.43**	**1.04**
5000	5	19.35	10.71	7.56	3.67	9.17	0.63	**3.34**	**1.58**
8000	5	23.05	13.86	10.3	5.21	10.29	6.98	**4.25**	**2.76**
500	10	0.28	0.11	0.95	0.62	0.29	0.13	**0.13**	0.06
700	10	0.56	0.34	1.54	0.79	0.47	0.21	**0.32**	**0.15**
1000	10	0.65	0.45	2.24	1.36	0.59	0.41	**0.45**	**0.21**
2000	10	4.31	2.11	3.92	2.13	4.02	3.08	**2.33**	**1.32**
5000	10	11.33	5.47	7.78	**2.52**	8.19	5.01	**5.29**	2.64
8000	10	23.11	12.78	10.59	6.31	20.17	11.38	**5.74**	**3.37**
500	20	2.02	1.03	2.71	1.82	1.79	1.13	**1.52**	**0.61**
700	20	17.63	8.22	11.73	4.31	13.86	9.37	**6.34**	**3.93**
1000	20	30.21	14.86	18.55	**6.58**	23.15	10.39	**9.75**	6.89
2000	20	123.21	103.91	65.17	38.19	100.11	73.14	**17.18**	**12.62**
5000	20	388.39	286.29	94.86	54.07	153.24	104.31	**50.27**	**32.78**
8000	20	1231.56	982.82	361.93	131.85	683.37	287.13	**83.49**	**75.63**

To further evaluate the location accuracy of the algorithm as a function of the network complexity, 10 base stations were selected from 500 nodes under various network complexities and these results were compared with the results of the classical algorithms; the comparison results are plotted in Fig. 10.

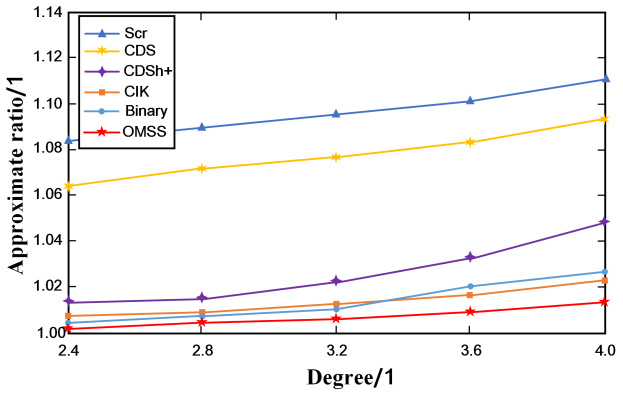

FIGURE 10. Comparison of the algorithms' siting accuracies under various complexities.

According to Fig. 9 and Fig. 10, the results of the OMSS algorithm are closer to the optimal site selection results than those of the other algorithms under the same network topology with various numbers of nodes and base stations and its location accuracy is more stable under various network complexities. To further evaluate the stability of the OMSS algorithm's running time, this paper selects various numbers of network nodes and base stations for running the OMSS algorithm, the SAA algorithm, CIK algorithm and the Binary algorithm. The results are shown in table2, Fig. 11 and Fig. 12.

According to Fig. 9 and Fig. 10, the results of the OMSS algorithm are closer to the optimal site selection results than those of the other algorithms under the same network topology with various numbers of nodes and base stations, and its location accuracy is more stable under various network complexities. To further evaluate the stability of the OMSS algorithm's running time, this paper selects various numbers of network nodes and base stations for running the OMSS algorithm, the SAA algorithm, CIK algorithm and the Binary algorithm. The results are presented in Table 2, Fig. 11 and Fig. 12.

According to Fig. 8, the running time of the OMSS algorithm is much shorter than those of the CDSh+ and EXACT algorithms and the results of the above simulation experiment demonstrate that the OMSS algorithm exhibits excellent performance in terms of both location accuracy and computational stability and can satisfy the time- efficiency requirements for address selection in large-scale BAWSNs.

Table 2 lists the times taken when the number of network nodes is increased by 8000 from 500 and when the number of locations is increased from 5 to 20, for selecting the best base-station location. The data from Table 2 demonstrate that when the number of locations is fixed, as the number of nodes increases, the location time of each algorithm increases overall. When the number of nodes does not change, the address time of each algorithm also increases within a range as the number of locations increases. In addition, according to Table 2, when the number of nodes and the number of locations have been determined, the OMSS algorithm outperforms the other two algorithms in terms of both the average location time and the most preferred address time.

Fig. 11 plots the time consumptions for calculating 2 and 5 optimal node positions for each algorithm when the number of locations is 2 and the number of nodes is from 2000 to 10,000. In Fig. 12, the time consumptions when the number of locations is 2, the number of nodes is from 10,000 to 100,000, and the algorithm selects 2 and 5 optimal node positions are plotted. Comparing with Fig. 11, when the number of locations is fixed and the number of nodes is increased from 2000 to 10,000, the overall site selection time of the algorithm increases, but the trend is not strong. Comparing Fig. 11 and Fig. 12, when the number of nodes is unchanged and the number of selected nodes is increased from 2 to 5, the address time increases of the SAA algorithm and the CIK algorithm are larger than that of the OMSS algorithm. Moreover, from the trends in Fig. 11 and Fig. 12, as the number of nodes increases and the number of selected sites increases, the time lags of the other algorithms increase substantially. However, the time

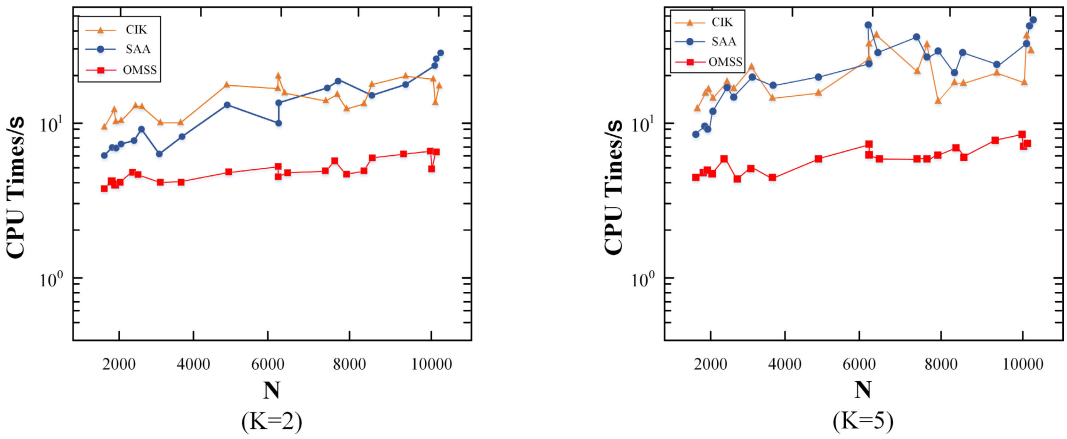

FIGURE 11. Computational times for instances with $N \in [2000, 6000]$ and $K \in \{2, 5\}$.

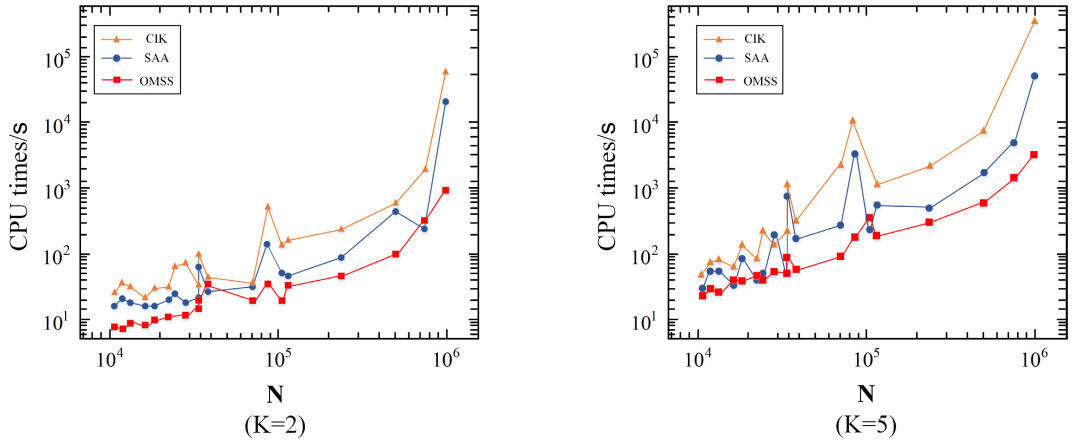

FIGURE 12. Computational times for instances with $N \in \left[10^4, 10^6\right]$ and $K \in \{2, 5\}$.

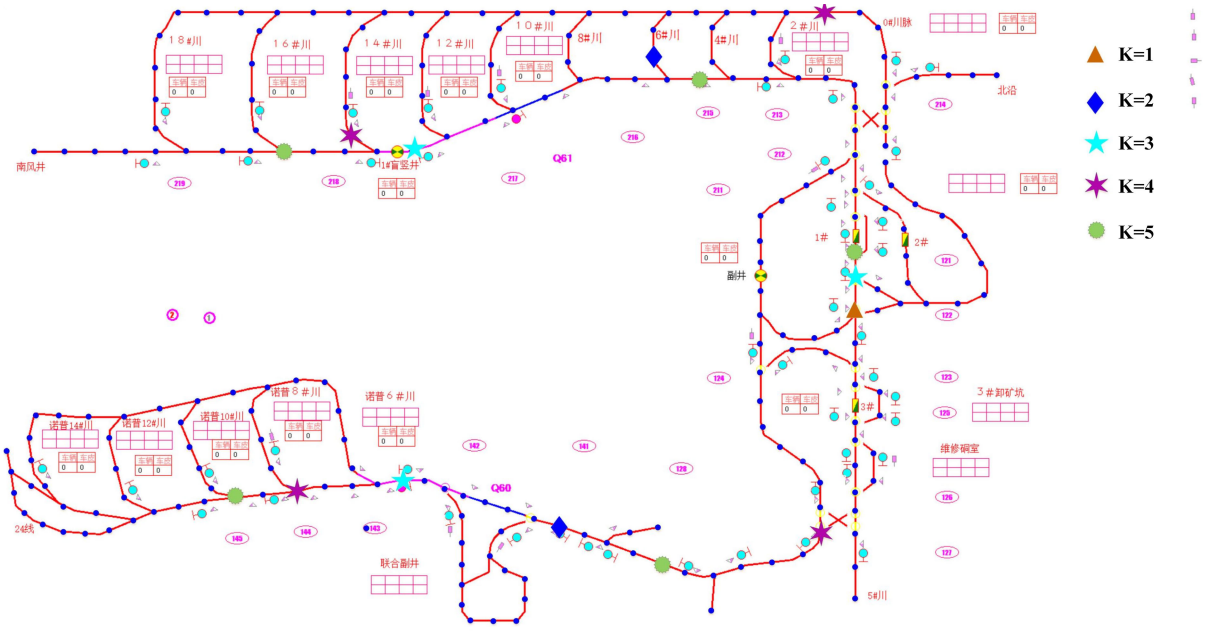

FIGURE 13. The effect graph of OMSS site selection under the real mine structure.

consumption of the OMSS algorithm does not increase substantially and the fluctuations are relatively stable. Based on these simulation results, the OMSS algorithm realizes excellent performance in terms of both location accuracy and computational stability, and can satisfy the address selection time efficiency requirements of large-scale band-area wireless networks.

Finally, the topology of a real strip-shaped wireless sensor network in an underground roadway with an area of 3 km × 5 km is selected for practical evaluation. The node spacing is 100 meters and the number of nodes is 225. The network topology and node distribution are presented in Fig. 10. The OMSS algorithm is run on this topology under various numbers of base stations and its running times are plotted in Fig. 13.

According to the results that are presented in Fig. 13, under the fixed belt network topology that is illustrated in Fig. 10, as the number of locations increases, the location time of the OMSS algorithm increases and the two are linearly related for the examined numbers of locations.

VI. CONCLUSION

In this paper, the problem of base-station location in a BAWSNs is studied, and the minimum-maximum base-station discrete p-center location model is established. The NP-hardness of the belt network location problem is proved. A $(1 + \varepsilon)$ approximation algorithm-OMSS algorithm that is based on gradient loop descent and has an algorithmic complexity of $O\left(kn^2\right)$ is proposed.

The results of our theoretical analysis and simulation verification demonstrate that OMSS can accurately locate the optimal base station with a low time consumption. Moreover, this algorithm outperforms the classical algorithms for the p-center location problem in terms of time consumption and stability.

In the future, we plan to study the problem of continuous p-center location for wireless networks in belt regions.

APPENDIX

Lemma 6: Under a single base station, the entire network forwarding delay time, which is denoted as T^*, is related to the longest non-cyclic path length, namely, $d_{\max}$, between nodes in the entire network G as follows:

$$T^* = \left\lceil \frac{d_{\max}}{2\overline{d}} \right\rceil \Delta t \tag{28}$$

Proof: In the propagation scenario that is shown in Fig. 3, the update time, namely, T_{total}, of the entire network is determined by the base station b and the node v_i that is farthest from it:

$$T_{\text{total}} = \frac{d_{b,v_i}}{\overline{d}} \Delta t \tag{29}$$

where d_{b,v_i} represents the distance between the base station s and the node v_i that is farthest from it.

In the above equation, both $\overline{d}$ and Δt are constants. Minimizing the update time T_{total} is equivalent to minimizing d_{b,v_i}. The network update must ensure that all nodes are updated. Therefore, d_{b,v_i} must satisfy the following constraint:

$$d_{b,v_i} \geq d_{b,v_j} \left(v_j \in V\right) \tag{30}$$

Due to the parallelism of time, the distance d_{b,v_i} should be transmitted in parallel to the two sides or in multiple directions with the base station b as the center in the update time T_{total}, where any one of the $\overline{b}$, $\overline{v}_i$ directions is assumed to be $\overline{b}$, $\overline{v}_i$, and the following constraint is imposed:

$$d_{b,v_i'} \geq \max d_{b,v_j'} \left(v_j' \in V\right) \tag{31}$$

Combining equation (28) and equation (29) yields:

$$d_{b,v_i} + d_{b,v_i} \geq \max d_{b,v_j} + \max d_{b,v_j} \tag{32}$$

Due to the parallelism of the transmission time, $d_{b,v_i} = d_{b,v_i}$ satisfies the following:

$$2d_{b,v_i} \geq \max d_{b,v_j} + \max d_{b,v_j} \tag{33}$$

Let d_{b,v_j} denote the length of path $R\left(b, v_j\right)$ and d_{b,v_j} denote the length of path $R\left(b, v_j'\right)$. Then, paths $R\left(b, v_j\right)$ and $R\left(b, v_j'\right)$ are connected end to end with b as the common node and the path length is as follows:

$$d_{v_j,v_j} \leq \max d_{b,v_j} + \max d_{b,v_j'} \left(v_j, v_j' \in V\right) \tag{34}$$

We obtain $d_{\max} = d_{v_j,v_j'}$ from the definition of $d_{\max}$. Then, equation (33) can be rewritten as follows:

$$\begin{aligned} 2d_{b,v_i} &\geq \max d_{b,v_j} + \max d_{b,v_j'} \\ 2d_{b,v_i} &\geq d_{v_j,v_j} \\ 2d_{b,v_i} &\geq d_{\max} \\ d_{b,v_i} &\geq \frac{d_{\max}}{2} \end{aligned} \tag{35}$$

According to equation (35), the minimum value of d_{b,v_i} is $\frac{d_{\max}}{2}$. Then:

$$\min T_{\text{total}} = \frac{\min d_{b,v_i}}{\overline{d}} t = \frac{d_{\max}}{2\overline{d}} t \tag{36}$$

The original proposition is proved.

Lemma 7: If the longest non-loop path in an undirected simple graph G is L, G is divided into K derived subgraphs, which are denoted as G_i, and l_i is the length of the longest non-loop path in subgraph G_i, then:

$$\sum_{i=1}^{K} l_i \geq L \tag{37}$$

Proof: The longest non-cyclic path, which is denoted as $w(L)$, must be separated into segments in the subgraph G_i and its length is L_i, where $L_i \leq l_i$. Then, $\sum_{i=1}^{K} L_i \leq \sum_{i=1}^{K} l_i$ because $\sum_{i=1}^{M} L_i = L$ satisfies $\sum_{i=1}^{M} l_i \geq L$ and the proposition is proved.

Lemma 8: If the graph G is divided into subgraphs of K longest non-cyclic path lengths, the longest subgraph length l_i for each subgraph G_i satisfies $l_i \geq \lfloor \frac{L}{K} \rfloor$.

Proof: We divide the graph G into subgraphs with the longest non-recurring path lengths. If $\forall i \neq j, l_i = l_j$, then $\sum_{i=1}^{K} l_i = K_i$. According to Lemma 6, $\sum_{i=1}^{K} l_i \geq L$ and, hence, $Kl_i \geq L$. Thus, $l_i \geq \lfloor \frac{L}{K} \rfloor$ and the proposition is proved.

Lemma 9: The following inequality is satisfied:

$$H\left(\frac{L}{K}\right)) \leq (1+\varepsilon) H\left((1+\varepsilon)\left\lfloor \frac{L}{K} \right\rfloor\right) \tag{38}$$

Proof: Find the convex hull of the node set V and enclose the entire graph G by a convex polygon of area S. If the search step size of the adjacent base station is $\lfloor \frac{L}{K} \rfloor$, then the distance between the adjacent base stations, namely, b_i and b_j, is $\lfloor \frac{L}{K} \rfloor$. From Inference 2, the corresponding division radius is $\lfloor \frac{L}{2K} \rfloor$, which is equivalent to base stations b_i and b_j being the centers and $\lfloor \frac{L}{2K} \rfloor$ being the radius coverage. The coverage area of each subgraph is $\pi\left(\lfloor \frac{L}{2K} \rfloor\right)^2$; thus, the convergence speed $H\left(\lfloor \frac{L}{2K} \rfloor\right)$ of the algorithm is as follows:

$$H(\left\lfloor \frac{L}{2K} \right\rfloor) = \frac{S}{\pi(L/2K\rfloor)^2} \tag{39}$$

Similarly, the convergence speed $H\left((1+\varepsilon)\left\lfloor\frac{L}{K}\right\rfloor\right)$ is as follows:

$$\begin{aligned} H\left((1+\varepsilon)\left\lfloor\frac{L}{K}\right\rfloor\right) &= \frac{S}{\pi[(1+\varepsilon)\lfloor L/2K\rfloor]^2} \\ &= \frac{S}{\pi(1+\varepsilon)^2\lfloor L/2K\rfloor^2} \\ &= \frac{H(L/K\rfloor)}{(1+\varepsilon)^2} \\ &\geq \frac{H(L/K\rfloor}{(1+\varepsilon)} \end{aligned} \tag{40}$$

Then, we obtain $H\left(\left\lfloor\frac{L}{K}\right\rfloor\right) \leq (1+\varepsilon)H\left((1+\varepsilon)\left\lfloor\frac{L}{K}\right\rfloor\right)$. The lemma is proved.

Lemma 10: If the network topology of the graph G has been determined, the data forwarding time $T\left((1+\varepsilon)\left\lfloor\frac{L}{K}\right\rfloor, S\right)$ under the base station distribution is obtained by reducing the search step size to $(1+\varepsilon)\left\lfloor\frac{L}{K}\right\rfloor$, which has the following relationship with the data forwarding time, namely, $T(d^*, S^*)$, under the optimal base station distribution:

$$\frac{T\left((1+\varepsilon)\left\lfloor\frac{L}{K}\right\rfloor, S\right)}{T(d^*, S^*)} \leq (1+\varepsilon) \tag{41}$$

Proof: The base station distribution that is obtained by searching with step size $(1+\varepsilon)\left\lfloor\frac{L}{K}\right\rfloor$ yields the longest division radius of the base station, namely, $(1+\varepsilon)\left\lfloor\frac{L}{2K}\right\rfloor$, and the longest division radius of the base station under the optimal base station distribution is $\frac{d^*}{2}$. According to the process of data transmission, the data transmission time of each subgraph is equal to the coverage radius of each subgraph base station:

$$T\left((1+\varepsilon)\left\lfloor\frac{L}{K}\right\rfloor, S\right) = (1+\varepsilon)\left\lfloor\frac{L}{2K}\right\rfloor \tag{42}$$

$$T\left(d^*, S^*\right) = \frac{d^*}{2} \tag{43}$$

The following relationship follows from Lemma 7:

$$\left\lfloor\frac{L}{2K}\right\rfloor \leq \frac{d^*}{2} \leq (1+\varepsilon)\left\lfloor\frac{L}{2K}\right\rfloor \tag{44}$$

Finally, we obtain the following:

$$T\left((1+\varepsilon)\left\lfloor\frac{L}{K}\right\rfloor, S\right) \leq (1+\varepsilon)T\left(d^*, S^*\right) \tag{45}$$

The proposition is proved.

REFERENCES

[1] H. Yuan, Y. He, and Y. Wu, "A comparative study on urban underground space planning system between China and Japan," *Sustain. Cities Soc.*, vol. 48, Jul. 2019, Art. no. 101541.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114. Aug. 2002.

[3] B. Wang, H. Xu, W. Liu, and H. Liang, "A novel node placement for long belt coverage in wireless networks," *IEEE Trans. Comput.*, vol. 62, no. 12, pp. 2341–2353, Dec. 2011.

[4] H. Cai, Y. Zhang, H. Yan, F. Shen, K. Zhou, and C. Zhang, "A delay-aware wireless sensor network routing protocol for industrial applications," *Mobile Netw. Appl.*, vol. 21, no. 5, pp. 879–889, Oct. 2016.

[5] A. Hashizume, H. Mineno, and T. Mizuno, "Multi-base station placement for reprogramming wireless sensor networks," in *Proc. 2nd Int. Conf. Symp. Appl. Internet (SAINT)*, Jul. 2009, pp. 284–285.

[6] J. Jeong, S. Kim, and A. Broad, "Network reprogramming," Berkeley, CA, USA, Tech. Rep. 8, Aug. 2003. Accessed: Jan. 14, 2013. [Online]. Available: http://www.tinyos.net/tinyos-1.x/doc/

[7] *TinyOS Community Forum*. Accessed: Jan. 14, 2013. [Online]. Available: http://www.tinyos.net

[8] T. Stathopoulos, J. Heidemann, and D. Estrin, "A remote code update mechanism for wireless sensor networks," UCLA, Los Angeles, CA, USA, Tech. Rep. 30, 2003.

[9] J. W. Hui and D. Culler, "The dynamic behavior of a data dissemination protocol for network programming at scale," in *Proc. 2nd Int. Conf. Embedded Networked Sensor Syst. (SenSys)*, Nov. 2004, pp. 81–94.

[10] S. Kim and D.-S. Eom, "Dynamic transmission power control for wireless sensor network reprogramming," in *Proc. 17th Asia–Pacific. Conf. Commun. (APCC)*, Oct. 2011, pp. 145–150.

[11] C. Dong and F. Yue, "An efficient network reprogramming protocol for wireless sensor networks," *Comput. Commun.*, vol. 55, no. 10, pp. 41–50, Jan. 2015.

[12] D. Kim, H. Nam, and D. Kim, "Adaptive code dissemination based on link quality in wireless sensor networks," *IEEE Internet Things J.*, vol. 4, no. 3, pp. 685–695, Jun. 2017.

[13] H. Teng, W. Liu, T. Wang, A. Liu, X. Liu, and S. Zhang, "A cost-efficient greedy code dissemination scheme through vehicle to sensing devices (V2SD) communication in smart city," *IEEE Access*, vol. 7, pp. 16675–16694, 2019.

[14] J. Wang, P. Urriza, Y. Han, and D. Cabric, "Weighted centroid localization algorithm: Theoretical analysis and distributed implementation," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3403–3413, May 2011.

[15] Y. T. Hou, Y. Shi, and H. D. Sherali, "Optimal base station selection for anycast routing in wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 55, no. 3, pp. 813–821, May 2016.

[16] Y. Shi and Y. T. Hou, "Optimal base station placement in wireless sensor networks," *ACM Trans. Sensor Netw.*, vol. 5, no. 4, pp. 2855–2869, Jun. 2006.

[17] I. Slama, B. Jouaber, and D. Zeghlache, "Energy efficient scheme for large scale wireless sensor networks with multiple sinks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar./Apr. 2008, pp. 2367–2372.

[18] B. Paul and M. A. Matin, "Optimal geometrical sink location estimation for two-tiered wireless sensor networks," *IET Wireless Sensor Syst.*, vol. 1, no. 2, pp. 74–84, Jun. 2011.

[19] J. Yan and K. G. Yu, "Base station placement for wireless sensor network positioning system via lexicographical stratified programming," *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 11, pp. 4453–4468, Nov. 2015.

[20] S. L. Hakimi, "Optimum locations of switching centers and the absolute centers and medians of a graph," *Oper. Res.*, vol. 12, no. 3, pp. 450–459, Jun. 1964.

[21] O. Kariv and S. L. Hakimi, "An algorithmic approach to network location problems. II: The *p*-medians," *SIAM J. Appl. Math.*, vol. 37, no. 3, pp. 539–560, Jan. 1979.

[22] N. Megiddo and K. J. Supowit, "On the complexity of some common geometric location problems," *SIAM J. Comput.*, vol. 13, no. 1, pp. 182–196, Jan. 1983.

[23] E. Minieka, "The *m*-center problem," *SIAM Rev.*, vol. 12, no. 1, p. 139, Jun. 1969.

[24] T. F. Gonzalez, "Clustering to minimize the maximum intercluster distance," *Theor. Comput. Sci.*, vol. 38, pp. 293–306, Jan. 1985.

[25] D. S. Hochbaum and D. B. Shmoys, "A best possible heuristic for the *k*-center problem," *Math. Oper. Res.*, vol. 10, no. 2, pp. 180–184, May 1985.

[26] D. Chen and R. Chen, "New relaxation-based algorithms for the optimal solution of the continuous and discrete *p*-center problems," *Comput. Oper. Res.*, vol. 36, no. 5, pp. 1646–1655, May 2009.

[27] C. Contardo, M. Iori, and R. Kramer, "A scalable exact algorithm for the vertex *p*-center problem," *Comput. Oper. Res.*, vol. 103, no. 3, pp. 211–220, Mar. 2019.

[28] D. Chen and R. Chen, "Optimal algorithms for the α-neighbor *p*-center problem," *Eur. J. Oper. Res.*, vol. 225, no. 1, pp. 36–43, Feb. 2013.

[29] Z. Drezner, V. Marianov, and G. O. Wesolowsky, "Maximizing the minimum cover probability by emergency facilities," *Ann. Oper. Res.*, vol. 246, nos. 1–2, pp. 349–362, Nov. 2016.

[30] J. A. Pacheco and S. Casado, "Solving two location models with few facilities by using a hybrid heuristic: A real health resources case," *Comput. Oper. Res.*, vol. 32, no. 12, pp. 3075–3091, Dec. 2005.

[31] W. Pullan, "A memetic genetic algorithm for the vertex *p*-center problem," *Evol. Comput.*, vol. 16, no. 3, pp. 417–436, Sep. 2008.
[32] Y. Levin and A. Ben-Israel, "A heuristic method for large-scale multi-facility location problems," *Comput. Oper. Res.*, vol. 31, no. 2, pp. 257–272, Feb. 2004.
[33] T. Drezner and Z. Drezner, "Sequential location of two facilities: Comparing random to optimal location of the first facility," *Ann. Oper. Res.*, vol. 26, nos. 1–2, pp. 5–18, Nov. 2016.
[34] J. Garcia-Diaz, J. Sanchez-Hernandez, R. Menchaca-Mendez, and R. Menchaca-Mendez, "When a worse approximation factor gives better performance: A 3-approximation algorithm for the vertex *k*-center problem," *J. Heuristics*, vol. 23, no. 5, pp. 349–366, Oct. 2017.
[35] J. Garcia Diaz, R. Menchaca Mendez, R. Menchaca Mendez, and R. Quintero, "A structure-driven randomized algorithm for the *k*-center problem," *IEEE Latin Amer. Trans.*, vol. 13, no. 3, pp. 746–752, Mar. 2015.
[36] M. G. C. Resende, and C. C. Ribeiro, "Greedy randomized adaptive search procedures: Advances, hybridizations, and applications," in *Handbook of Metaheuristics*. Boston, MA, USA: Springer, Mar. 2010, pp. 283–319.
[37] B. Callaghan, S. Salhi, and G. Nagy, "Speeding up the optimal method of Drezner for the *p*-centre problem in the plane," *Eur. J. Oper. Res.*, vol. 257, no. 3, pp. 722–734.
[38] J. Mihelič and B. Robič, "Solving the *k*-center problem efficiently with a dominating set algorithm," *J. Comput. Inf. Technol.*, vol. 13, no. 3, pp. 225–234, Mar. 2005.
[39] M. Albareda-Sambola, L. I. Martínez-Merino, and A. M. Rodríguez-Chía, "The stratified *p*-center problem," *Comput. Oper. Res.*, vol. 108, no. 8, pp. 213–225, 2019.

ZHANGLING DUAN received the B.Sc. degree from the Hefei University of Technology, in 2014, where he is currently pursuing the Ph.D. degree. His main research interests include the Internet of Things (IoT) and wireless sensor networks (WSN).

XING WEI received the Ph.D. degree from the University of Science and Technology of China, in 2009. He is currently an Associate Professor with the Hefei University of Technology. His research interests include deep learning and the Internet of Things engineering, driverless solutions, and so on.

JIANGHONG HAN is currently a Full Professor and a Ph.D. Supervisor with the Hefei University of Technology. His research interests include networking, wireless communications, embedded systems, and wireless sensor networks.

GUOQIANG YANG received the B.Sc. degree from the Anhui University of Finance and Economics. He is currently pursuing the master's degree with the Hefei University of Technology. His research interests include computer vision and wireless sensor networks.

YANG LU was born in 1967. He received the Ph.D. degree from the Hefei University of Technology, in 2002. He is currently a Professor and a Doctoral Supervisor with the the Hefei University of Technology. His main research interests include the Internet of Things (IoT) engineering and distributed control systems.