Hindawi
Journal of Sensors
Volume 2021, Article ID 6680718, 18 pages
https://doi.org/10.1155/2021/6680718

Research Article

Wireless Sensor Network Deployment in Cyberphysical Machine Tool System Based on Optimal Allocation of Memory Buffers

Xiaoyang Zhou

College of Engineering, Harbin University, Harbin, China

Correspondence should be addressed to Xiaoyang Zhou; zxy2013yang@126.com

Received 12 December 2020; Revised 11 January 2021; Accepted 15 January 2021; Published 3 February 2021

Academic Editor: Xavier Vilanova

As an important direction of Industry 4.0, cyberphysical machine tool systems (CPMTS) can realize the deep integration and real-time interaction of physical components and information to optimize manufacturing processes. Wireless sensor network (WSN), an important part of CPMTS, is responsible for data collection and transmission. However, in the process of data transmission, due to memory limitations and noise interference, unreasonable sensor distribution will affect the performance of CPMTS. At the same time, data accuracy will be affected due to the resource constraints of CPMTS. To solve the problems above, this paper firstly presented a single-station transfer model to ensure the layout of sensors in each sink, which can meet the detection capability of fault/monitoring data. Then, by using fuzzy graphs, a multihop-station transfer model and data-collecting model are developed to describe the data flow and memory allocation in the wireless network. Taking noise interference and data position into consideration, a MILP problem is formulated and the optimization solution is obtained by using the "branch and bound" method. Finally, case studies about optimal sensor distribution on the single station and path optimization on the multihop station are presented to illustrate the proposed strategy. The case studies validated that the proposed sensor distribution in a single station can achieve higher detectability with fewer resources, and the optimization path strategy can achieve the best performance in two proposed experiments, compared to the shortest path and noninferior path strategies.

1. Introduction

CPMTS (cyberphysical machining tool systems) with an open architecture are essential to individualized industries (small batch and various types) in modern manufacturing, as their hardware, software, and bus specifications have an open architecture design [1, 2]. Thanks to their advantage in flexibility, adaptability, versatility, and expansibility, end users can realize their customized functions by developing algorithm modules and capture the information flow as they need [3]. For CPMTS, sensors act as one of the core components to provide the vital link between control systems and the physical world, which means sensors and sensing technologies constitute the fundamental basis in the perception layer. In particular, wireless sensor networks (WSNs) have become an active area in researches, as WSNs can overcome some limitations of wired sensor networks such as the complexities of installation, the high failure rate of connectors, and the difficulties in troubleshooting an individual sensor. The performance of CPMTS critically depends on the accuracy and efficiency of sensor measurements on faulty symptoms and monitoring quantity. Improper sensor deployment leads to insufficient or inaccurate measurements which may cause functional abnormalities and even personnel injuries. Although redundantly sensing every physical parameter of a system can reduce information loss, the redundant sensor network may be cursed with an overload on data analysis as well as cost. This is especially critical in CPMTS, since system overload reduces real-time performance, which will result in an abundance of serious consequences. Moreover, due to limited communication bandwidths, redundantly deploying sensors is regarded as invalid.

At present, in the deployment of WSNs, many studies have mainly focused on the lifetime [4–6], quality of service (QoS) [7, 8], and network cost [9–11], since they are the most critical issues. Chanak and Banerjee [12] presented a fault node distribution and management approach based on fuzzy

rules in WSNs, which implement the efficient route towards the base station to reuse faulty nodes. It provides better QoS and network lifetime. Awad [13] presented an analytical framework for the optimal deployment of relay nodes of a wireless network using a two-dimensional Gaussian distribution. Céspedes-Mota et al. [14] applied a multiobjective differential evolution algorithm to jointly optimize the distribution of the wireless sensors. Khalesian and Delavar [15] proposed a constrained Pareto-based multiobjective evolutionary approach (CPMEA) which was aimed at finding Pareto optimal layouts that maximize the coverage and minimize the sensor energy consumption for the sake of prolonging the network lifetime while maintaining full connectivity between each sensor node and the high-energy communication node (HECN). Rahman et al. [16] used two-sensor node deployment strategies: non-corona- and corona-based sensor node deployment strategies to significantly utilize the energy of the nodes and prolong the network lifetime. Research on wireless sensor distribution for manufacturing is very limited. Ko et al. [17] analyzed interference at multiple-sensing nodes in FSN by applying statistical methods to collected data; then, they proposed an interference model to obtain optimal deployment strategies that minimize the influence of interference. Wang et al. [18] proposed the Enhanced Power-Efficient Gathering in Sensor Information System (EPEGASIS) algorithm to alleviate the problem of hot spots from three aspects: determining optimal communication distance to reduce energy consumption, setting the threshold value to protect the dying nodes, and using mobile sink technology to balance the energy consumption among nodes. Collotta et al. [19] proposed an energy efficient method for data fusion based on fuzzy logic to achieve QoS in WSNs. However, this method only aggregates true information, instead of gathering complete data from the WSNs.

There are, however, other concerns beyond those which are ignored. In particular, memory allocation is a critical problem when additional wireless sensors that match customized functions are deployed in CPMTS where out-of-memory errors may exist. Memory allocation has been well studied in deeply embedded applications. Sánchez-Oro et al. [20] proposed a parallel variable neighborhood search algorithm for the dynamic memory allocation problem to solve dynamic memory allocation problems in embedded systems. Goens et al. [21] modeled the application in a data-centric way, by explicitly defining buffers and associating computational tasks that access the buffers within well-specified time intervals, and then presented a layered approach to describe and solve the buffer-allocation problem as well as related subproblems, using mixed-integer linear programming. Soto et al. [22] proposed two mid-term iterative approaches to solve a static version of the allocation problem, which have the best solution quality compared to long-term and short-term approaches. In addition, considering noise sources, WSNs may evolve into a multiple hop network in which data transmission is more difficult. Therefore, a good sensor deployment strategy, which can result in a configuration with the optimal performance while satisfying prespecified resource constraint criteria, must be developed to overcome such problems.

Our goal is to establish a WSN deployment strategy meeting the requirement of CPMTS (as shown in Figure 1) and memory restriction, as a reliable solution for a resource-constrained environment where wireless sensors continuously monitor manufacturing processes. The main contributions of this paper are as follows:

(1) A single-station transfer model that ensures the detection capability of fault and monitoring data in each sink is presented

(2) By using fuzzy graphs, a multihop-station transfer model and a data-collecting model are developed to describe the data flow and memory allocation in the wireless network

(3) Taking noise interference and data position into sensor path optimization, a MILP problem is formulated and the optimization solution is obtained by using the "branch and bound" method

The rest of this paper is organized as follows: Section 2 provides details of model development about the transfer model and the data-collecting model. Then, Section 3 presents the optimal deployment strategy under the MILP. Section 4 illustrates detailed case studies. Section 5 highlights the findings of this research and discusses future work.

2. Model Development

2.1. Single-Station Transfer Model. In manufacturing, the quality of the sensor signal is strict, especially for WSNs where it is necessary to ensure that the sensing data can be obtained accurately and quickly by a numerical control system. Thus, a multihop WSN where multiple sinks are used to ensure data quality and to avoid noise sources is the key to solving this problem. To better describe the flow of information in multihop WSNs, based on the stream of variation (SOV) theory [23], we present a single-station transfer model preferentially, as shown in Figure 2.

$X_i \in \mathbf{U}^{p\text{x}1} (i = 1, 2 \cdots n)$ represent the flow of failure and measurement information between two adjacent depths (depth i and depth $i-1$). $T_i \in \mathbf{U}^{q\text{x}1}$ is the relay information from the sinks at the same depth. The pickup information of sensors in sink i is denoted by $Y_i \in \mathbf{U}^{v\text{x}1}$. Sensor noise, denoted by ξ_i and η_i, are vectors of uncorrelated random variables with zero means. Then, the single-station transfer model in sink i can be expressed as follows:

$$\begin{aligned} X_i &= \mathbf{A}_{i+1} X_{i+1} + \mathbf{B}_i T_i + \zeta_i, \\ Y_i &= \mathbf{C}_i X_i + \eta_i \quad i = 1, 2, \cdots n. \end{aligned} \tag{1}$$

Here, dynamic matrix $\mathbf{A}_{i+1}$ means data attenuation or enhancement during the signal transmission between two adjacent depths. Likewise, dynamic matrix $\mathbf{B}_{i+1}$ indicates the data attenuation or enhancement at the same depth. The system matrix $\mathbf{C}_i$ is determined by the layout structure

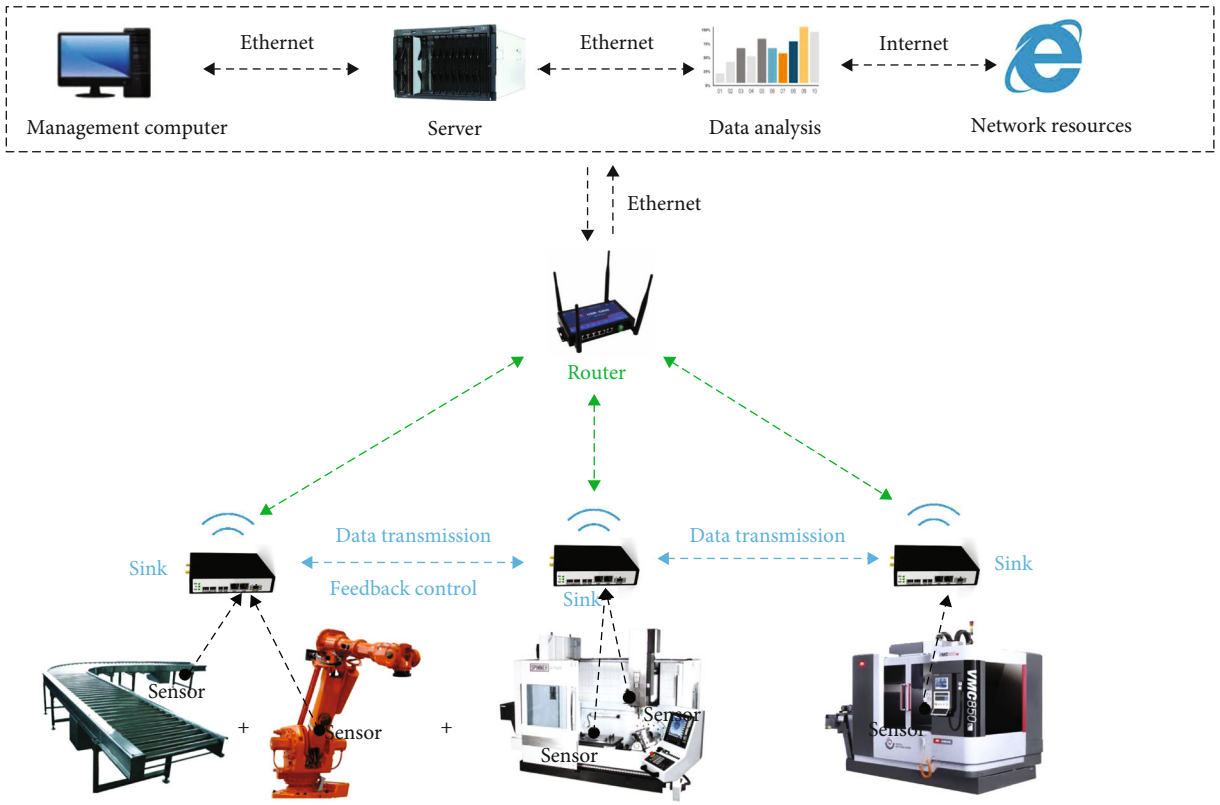


Figure 1: Cyberphysical machining tool system.

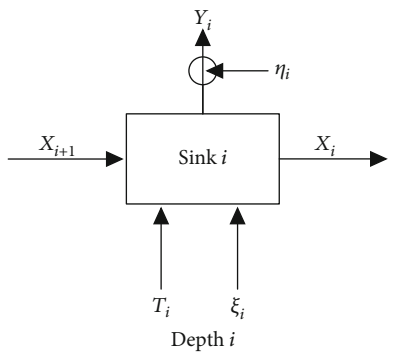


Figure 2: Single-station transfer model in depth i.

of the sensor attached to sink i. As we know, sink i is the base station that transmits the data obtained by its sensors to the next station. So, using fuzzy graph $G_F = (\mathbf{E} \cup \mathbf{R}, \mathbf{B} \cup \mathbf{S} \cup \mathbf{D})$, the relationship among sinks, sensors, and monitoring data/fault can be described in Figure 3.

The fuzzy graph is composed of base station node B, sensor node S, data node D, and detection characteristic edge E as well as data transmission edge R. Data node D, consisting of monitor volume priority (P), fault severity (F), occurrence rate (O), and detection rate (D), can be represented by priority number (PN). The PN value can be calculated by multiplying the ranked P, F, O, and D values from FEMA and algorithm configuration information, as $\mathrm{PN} = P \times F \times O \times D$, $\mathrm{PN} \in \mathbf{PN}$. Here, the higher the PN value is, the higher priority for monitoring and corrective action should be given. Similar to the data node, the sensor node includes the information of signal-noise ratio (SNR), sensitivity (sen), resolution (res), and accuracy (acc). The sensor index (SI) is used to put the above factors together; then, the SI value is $\mathrm{SI} = \mathrm{SNR} \times \mathrm{sen} \times \mathrm{res} \times (1 - \mathrm{acc})$, $\mathrm{SI} \in \mathbf{SI}$. The base node, with the information of the sensor data transmission ratio (SDTR), reflects the information transmission ability between the sensor and the base station. The SDTR value can be calculated as dividing data gain (DG) by sample time (ST) as SDTR = DG/ST. Besides, like the base node discussed above, sensor data index (SDI) including sensing gain (SG) and sample time (ST) is denoted to reflect the detectability between sensors and data. The SDI value can be calculated as SDI = SG/ST. The value discussed above needs to be normalized into comparable values based on the fuzzy membership function [24]. Taking the PN value as an example, we define a membership function f_A as fuzzy subset **A** on the **PN** set. There exists a mapping from generic elements PN on **PN** to fuzzy subset **A**. Then, we can map the PN value for different types of data into comparable values between 0 and 1:

$$f_A : \mathrm{PN} \longrightarrow [0, 1]. \tag{2}$$

The connection path (i.e., from node D to S, then to B) can be divided into two parts. One is considering fuzzy nodes D, S, and edge E; then, the path from D_i to S_j is defined as $p_{ij} = (D_i, S_j)$, where **PDS** is the set of p_{ij}. The other is considering fuzzy nodes B, S, and edge R. The path from S_j to B_z can

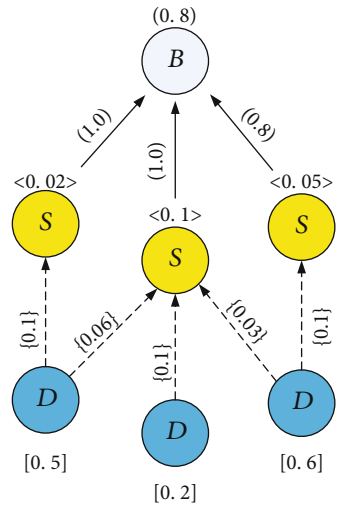


Figure 3: A fuzzy triplicate graph for BSD.

be defined as $q_{jz} = (S_j, B_z)$, where **PSB** is the set of q_{jz}. Then, the connection path CP can be defined as follows:

$$\mathrm{CP} = \mathbf{PDS} \cup \mathbf{PSB}. \tag{3}$$

There may exist more than one path from data node D to sensor node S, as shown in Figure 3, so equation (3) needs to be modified as follows:

$$\mathrm{CP} = (\mathrm{PDS}_1, \mathrm{PDS}_2, \cdots, \mathrm{PDS}_n) \cup \mathrm{PSB}. \tag{4}$$

After confirming the connecting path and calculating values on the B, S, and D nodes, the values of edge elements R and E in the graph are critical to represent the cause-effect relationship from the sensor detectability to data and transmission ability from sensors to the base station. When it comes to edge value determination, multiple attributes like data, sensors, data-sensor relations, and base station and sensor-based relation will affect sensor deployment on data monitoring/fault diagnosis. This paper uses the analytic hierarchy process (AHP) to integrate these properties into edge element values in the fuzzy graph. The detailed calculation process is in reference [24]. We summarize the results that will be used in this paper. The edge value of E and R can be described as follows:

$$\begin{aligned} E &= [\mathrm{PN}, \mathrm{SDI}, \mathrm{SI}] \times [\mathrm{PV}_{\mathrm{PN}}, \mathrm{PV}_{\mathrm{SD}}, \mathrm{PV}_S]^T, \\ R &= [\mathrm{SI}, \mathrm{SDTR}, \mathrm{DG}] \times [\mathrm{PV}_S, \mathrm{PV}_{\mathrm{SDTR}}, \mathrm{PV}_D]^{\mathrm{T}}. \end{aligned} \tag{5}$$

Here, $PV = (pv_1, pv_2, \cdots, pv_n)$ can be obtained by generating the comparison matrix $\mathbf{CM}_n$ to calculate the geometric means of each row and relative priorities. In addition, we define a set **E** that includes all the edge values of E, and detection coefficient $\varphi_i \in \mathbf{SDI}$ is represented to reflect the detection capability of each sensor to the fault/monitoring data. Similarly, a set **R** that includes all edge values of R is defined, and $\pi \in \mathbf{SDTR}$ is used to reflect the transmission capability from the sensor node to the base node. When π is equal to zero, it means that the sensors at sink i cannot monitor data/-fault j or the layout of the sensors is invalid. Otherwise, when π is equal to 1, it means that the sensors at sink i can meet the monitoring requirements with respect to data/fault j. Besides, $\pi \in (0, 1)$ means that transmitting part of the data from sensors to the sink has failed, which means a new layout plan of sensors or sinks needs to be put forward.

2.2. Multihop-Station Transfer Model. To better describe the data flow in WSN of CPMTS, we extend the single-station transfer model. Taking noise interference [16] and heterogeneous sensors into consideration, this paper presents a hierarchical architecture to ensure detectability and diagnosability, as shown in Figure 4. We assume that one single sink can only be connected to a certain sink at the same time, which means a plural connection is not allowed. The transverse direction elements represent the flow of information between different depths in the manufacturing process, and longitudinal direction elements refer to the flow of information at the same depth. The index of sinks at the same depth is denoted by $\mathrm{sink}_i^{(n)}$, $i, n = 0, 1, \cdots k$.

Path connection with the proposed model is determined as the data path from depth i to depth n represented by $\mathbf{W}^T_{i,n}$ and the data path at the same depth represented by $\mathbf{V}^T_{i,n}$. For example, if there exists a data flow from depth i to depth $i-1$, then $\mathbf{W}^T_{i,i-1} = 1$, otherwise is 0. Thus, the input-output model of the system can be obtained as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_k \\ \cdots \\ Y_n \end{bmatrix} = \begin{bmatrix} A_2 W_{2,1} \\ A_3 W_{3,2}, A_2 W_{2,1} \\ \cdots \\ A_{k+1} W_{k+1,k}, A_{k-1} W_{k,k-1}, \cdots, A_2 W_{2,1} \\ \cdots \\ A_{n+1} W_{n+1,n}, \cdots, A_{k+1} W_{k+1,k}, \cdots, A_2 W_{2,1} \end{bmatrix} \cdot \begin{bmatrix} X_2 \\ X_3 \\ \cdots \\ X_{k+1} \\ \cdots \\ X_{n+1} \end{bmatrix} + \begin{bmatrix} B_2 V_{2,1} \\ B_3 V_{3,2}, B_2 V_{2,1} \\ \cdots \\ B_{k+1} V_{k+1,k}, B_{k-1} V_{k,k-1}, \cdots, B_2 V_{2,1} \\ \cdots \\ B_{n+1} V_{n+1,n}, \cdots, B_{k+1} V_{k+1,k}, \cdots, B_2 V_{2,1} \end{bmatrix} \cdot \begin{bmatrix} T_2 \\ T_3 \\ \cdots \\ T_{k+1} \\ \cdots \\ T_{n+1} \end{bmatrix} + \boldsymbol{\theta}. \tag{6}$$

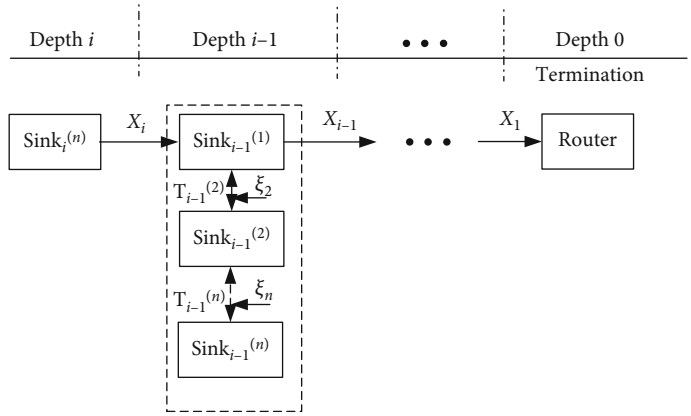


Figure 4: Multihop-station transfer model.

Here, $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_n]$, $\theta_n = V_{n,n-1} \cdot \xi_n + \eta_n$. A detailed diagnosability analysis of equation (6) was reported in [24]. We summarize the results that will be used in this paper. The covariance of equation (6) with respect to Y is as follows:

$$\sum_Y = W \sum_X W^T + V \sum_T V^T + V \sum_\theta V^T. \tag{7}$$

In equation (7), $\sum_Y$ is known for the percentage of the sensor information obtained by the corresponding sink, and the eigenvalue is defined as λ. The transmission capability of the sink i can be represented by μ, where u_{ij} is the transmission capability from the sink i to sink j. The transmission capability coefficient u_{ij} is as follows:

$$u_{ij} = \begin{cases} \dfrac{\lambda i}{\lambda j} & i \neq j \\ 1 & i = j \end{cases} \quad i \geq 1, n \geq j. \tag{8}$$

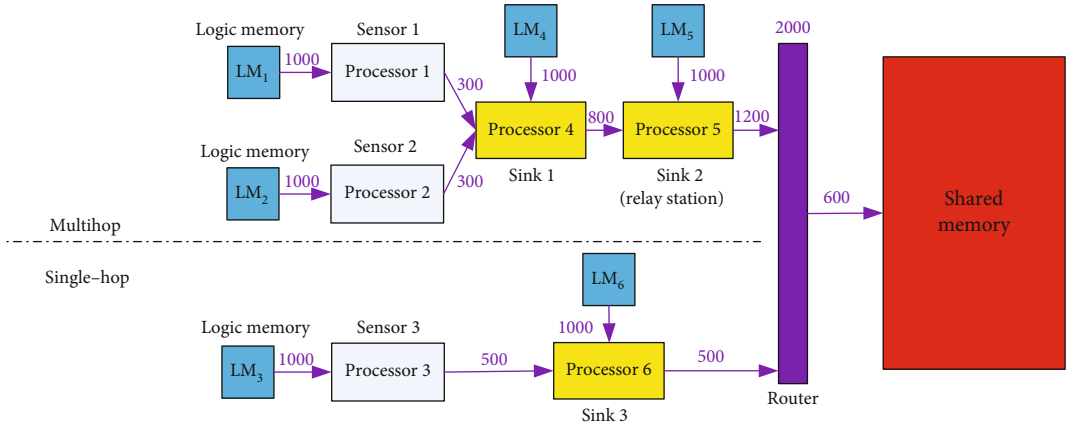
From equation (8), we can conclude that when u_{ij} is equal to zero, sink i cannot transmit data to sink j, or sink i failed to transmit data to the next depth. Otherwise, when u_{ij} is equal to 1, it means that sink i can meet the transmission requirements with respect to sink j. In addition, when u_{ij} is between 0 and 1, it means that transmitting part of the data by sink i has failed or transmission capacity is limited.

2.3. Data-Collecting Model. In CPMTS, the wireless sensor network in a sensing layer undertakes the task of data acquisition. Each sensor transmits a large amount of data to the system per unit time. Although the sensor redundancy arrangement can make the algorithm get better data support, it will greatly increase the system burden, as well as data reading and writing errors. Here, we use a data-centric approach by explicitly defining a buffer and associating the task of accessing the buffer at a specified time interval, and then a data flow model is obtained to maximally balance the bandwidth of channels and memory usage.

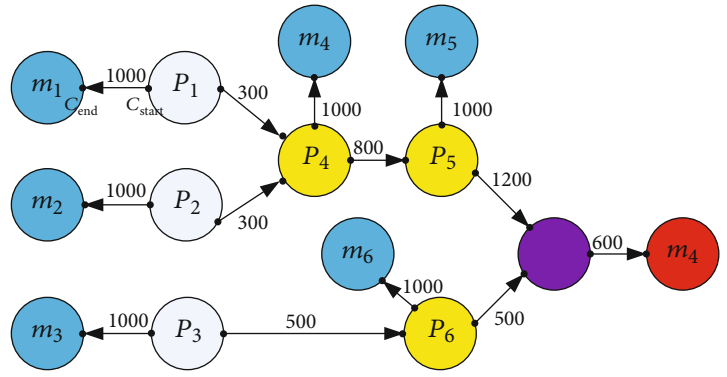
In this paper, we consider hardware architectures of the data-collecting model consisting of processing elements, storage elements, and simple interconnections, as shown in Figure 5(a). As it is commonly done, we represent the actual model as a fuzzy graph $G_{\text{DF}} = (V', E')$, as shown in Figure 5(b). To abstract the hardware, two sets, **M** and **P**, represent the storage elements (logic buffer and shared memory) and process elements, respectively. And then, two nodes c_{start} and c_{end} are defined to represent the interconnection between two elements. Hence, V' in graph G_{DF} can be described as $V' = \mathbf{M} \cup \mathbf{P} \cup \{c_{1,\text{start}}, c_{1,\text{end}}, c_{2,\text{start}}, \cdots\}$. We denote the edges $e' \in \mathbf{E}'$ with the bandwidth denoted by $bw(e')$ of the interconnection in kilobytes per second (kbps). In addition, the direction of the edge expresses the possibility of access. For example, an edge $e = (v'_{\text{start}}, v'_{\text{end}})$ means that v'_{start} with all its accessory elements can access v'_{end}. Notice that access or communication is valid only between two nodes, which allows the simple modeling of access rights via edge directions in the graph while retaining much descriptive power.

A data-collecting or data-transmitting application is modeled as a collection of data buffers and tasks that access these buffers at an arbitrary instance of time. We define a five-tuple (p, lb, d, t_1, t_2) to represent a read or write access with a bandwidth demand of d kbps from the processing unit p to the logic buffer lb (i.e., $lb \in \mathbf{M}$, with a function of size (lb)) to represent the size in byte of each logic buffer. The write or read access starts at t_1 and ends at t_2. Hence, a data flow at any interval of time can be represented as $f_i = (p_i, lb_i, d_i, (t_1)_i, (t_2)_i)$, $f \in \mathbf{F}$. For example, the flow of $f_3 = (p_3, lb_3, 600,5,7)$ indicates that the processing unit p_3 reads or writes 600 kbps of data from buffer lb_3 during time intervals 5 to 7.

For now, our goal is to find a suitable solution to achieve efficient allocation from buffer to shared memory under the restriction of bandwidth and memory size, which means that at every edge the bandwidth capacity is not exceeded by the sum of demands going through it. To simplify the model, we firstly consider it at a single time point. Consider set $\boldsymbol{\Pi}$ of all paths in the fuzzy graph G_{DF}. Since the paths in Figure 5(b) are unidirectional, the data flow path from processor p_i to storage element m_j can be represented as $fp(p_i, m_j) = (p_i, v'_0) \cup (v'_0, v'_1) \cup, \cdots, \cup (v'_k, m_j)$. Note that



(a) Hardware architectures of data collecting model



(b) Corresponding fuzzy graph G_{DF}

FIGURE 5: Data-collecting architectures and corresponding fuzzy graph G_{DF}.

the path from p_2 to m_3 is invalid because no path allows p_2 access to b_3. Then, a binary variable $P_{f,\pi}$ for every pair of flow $f \in \mathbf{F}$ and a path $\pi \in \boldsymbol{\Pi}$ is presented, where $P_{f,\pi} = 1$ indicates that flow f is routed via path π.

Based on the full-time spectrum theory [21], time operator $T_k \subseteq \mathbf{T}$ is introduced to represent time frame in blocks, so all time dependencies can be handled at once instead of a single time point. To ensure that data flow is efficient, for all edges $e' \in \mathbf{E}'$, the amount of data cannot exceed the bandwidth at any time point. Note that in the wireless sensor node layout there exists relay nodes that avoid noise interference and extend the coverage area. As a relay node, the total transmitted data at all time dependencies should be within the bandwidth limitation, which means

$$\sum_{\substack{\pi_k \in \boldsymbol{\Pi}_k \\ e' \in \mathbf{E}'}} \sum_{\substack{f_i, f_j \in \mathbf{F} \\ \pi_{ki} \in fp(p_i, lb_j)^T, \pi_{kj} \in fp(p_j, m)^T}} P_{f_i,\pi_{ki}} |T_k| d_i + P_{f_j,\pi_{kj}} |T_k| d_j \le bw(e'). \tag{9}$$

Every buffer should allocate in exactly one memory, so we define two additional binary variables that represent the allocation. $A_{lb,m}(lb, m \in \mathbf{M})$ indicates that the data in the logic buffer can be transmitted into shared memory, where $A_{lbi,mi} = 1$ means that the data in logic buffer lb_i is allocated in the shared memory m_i. What's more, MI $D_{lbi,lbj}$ indicates that base station s_i transmits data to the relay station s_j. Likewise, $MID_{lbi,lbj}$ means that the data in the logic buffer lb_i is allocated in the logic buffer lb_j of the relay station. Note that the flow of data includes the allocation, which means

$$\sum_{\substack{f \in \mathbf{F} \\ \pi k \in \text{ft}(p_i, lb_i)}} P_{f,\pi_k} = A_{lb,m}. \tag{10}$$

Taking relay station and entire time dependencies into consideration, the condition for explicit mapping between logic buffer lb and shared memory m is

$$\sum_{\substack{m\in\mathbf{M}\\ lb,lbi,lbj\in\mathbf{M},\text{life}(lb)\cap T_k\neq\phi}} A_{lb,m}\left(\text{MID}_{lbi,lbj}+1\right)\neq 0. \tag{11}$$

Here, life(lb) means the logic buffer lb is available at time point T_k. A further condition is that the memory sizes are not exceeded by the sum of the buffers allocated there.

$$\sum_{\substack{m,lbi,lbj\in\mathbf{M}\\ \text{life}(lb)\cap Tk\neq 0}} A_{lbj,m}\text{size}\left(lb_j\right)+\text{MID}_{lbi,lbj}\text{size}(lb_i)\leq \text{size}(m). \tag{12}$$

From equation (9) to equation (12)), it is ensured that at no point in time will the total size of all buffers exceed the size of the available memory. A new condition should be taken into account so that when the buffer is allocated in memory, no matter if there is enough space or not, the position of the buffer will affect the allocation. We introduce the integer variable D_b to represent the first memory address of logical buffer lb. And the first available address of memory can be represented as M_b. To ensure that the allocation is valid, which means the data is cached in shared memory and the logical buffer blocks in shared memory are contiguous, the following limitation needs to be considered. The condition for valid memory allocation is that the first logical buffer address in memory is forward to the first available address of memory, and the total size of the logical buffer in shared memory is less than the shared memory size. Taking the relay station problem into consideration, we can formulate the above problem:

$$\sum_{D_{bi}+\text{size}(b_i)\cap D_{bj}+\text{size}\left(b_j\right)=\phi} D_{bi}+\text{size}(b_i)+\text{size}\left(b_j\right)\leq M_b+\text{size}(m),$$
$$D_{bi}\geq M_b. \tag{13}$$

Notice that overlap between two buffers lb_1 and lb_2 is not allowed in allocation, which means $D_{b1}+\text{size}\left(b_1\right)\leq D_{b2}$.

3. Optimal Sensor Distribution Strategy

3.1. Sensor Path Optimization. At present, wireless sensor optimal distribution strategy focuses on two aspects: one is to minimize the number of sensors in order to control overall cost. The other is to prolong the lifetime of the wireless sensor network. However, relay stations, which play an important role in data transmission, control the flow of data, thereby, directly affecting the quality of data transmission. A reliable data transmission path can reduce noise interference and improve the accuracy of data transmission. Here, based on data flow, we optimize the data path from one arbitrary base node to the router. To simplify the problem, we use a 4×4 square grid space to represent the multihop-station transfer model, where (i, j) is the index of each segment. The 4×4 square grid space is shown in Figure 6, with the depth from D_0 to D_3.

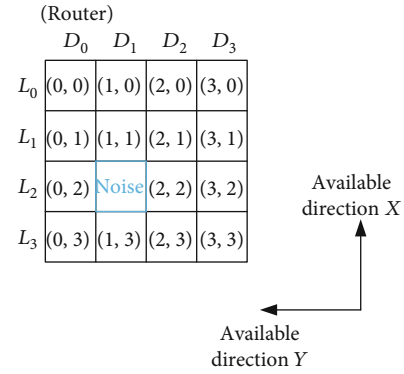


Figure 6: A 4×4 square grid space.

Assumptions in the sensor path optimization are as follows:

(1) Data can only be transmitted in one direction at the same time, horizontal or vertical, between depths or within layers

(2) There is only one base station (i.e., sink) located at the center of each grid segment, and the distance between two arbitrary base stations is within communication range R_c

(3) Interference by a noise source affects every sensor node individually. Thus, the level of noise is additive and the total interference effect in each grid segment is proportional to the number of sensors in the segment

(4) Noise source is located at (1,2)

(5) Sensors in each grid segment satisfy requirements of detection and diagnosis

The goal of our optimization is to make the sensor path shortest and avoid the impact of noise, so that data can be transferred accurately and quickly. To achieve this, three cases need to be taken into consideration. The first case is memory overload (see Figure 7(a)): during the transmission of data from a base station to a relay station, there may be situations in which the relay station cannot receive the transmission data completely due to the overload of data. The second case is insufficient bandwidth (See Figure 7(b)): in this situation, due to bandwidth limitation, data cannot be transmitted to the relay station completely. The last case is electromagnetic interference (see Figure 7(c)): when the path contains strong electromagnetic interference, it is impossible to ensure the accuracy and completeness of data, which is forbidden in CPMTS.

For manufacturing systems that are less affected by electromagnetism, we can figure that the shortest path from depth 3 to depth 0 is toward the lateral propagation. For

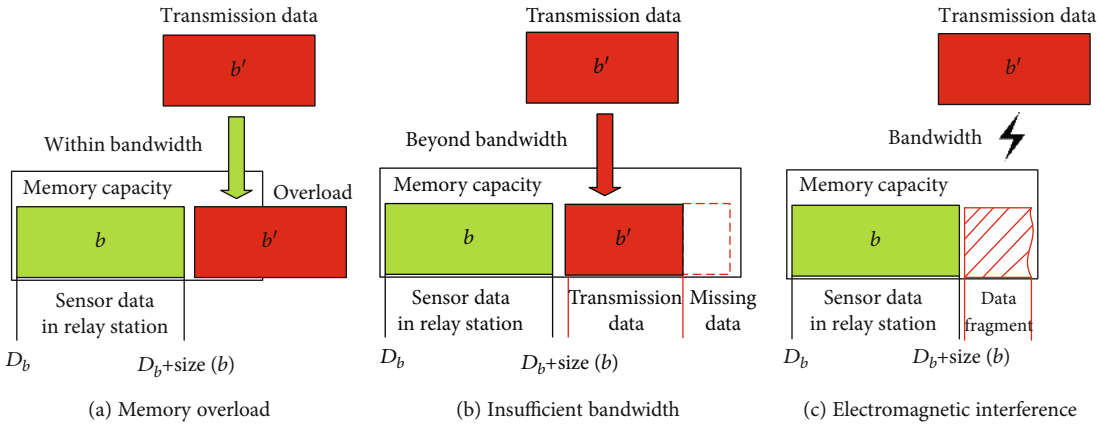


Figure 7: Three possible cases in data transmission.

example, the shortest path from segment (3,0) to segment (0,0) is $(3,0) \cup (2,0) \cup (1,0) \cup (0,0)$. However, the cost in sensor distribution is one of the primary targets, and too many routers in depth 0 can cause poor system robustness. So the optimization can be expressed to find a valid path to minimize resource consumption. To achieve this, a 4×4 directed graph matrix $\mathbf{G}$ is defined, where $\mathbf{G}_{i,j}$ means the amount of data gathered by sensors. Note that $\mathbf{G}_{i,j} = 0$ means strong electromagnetic interference in this segment that forbids data from passing through. A path matrix $\mathbf{D}_{4\times4}$ is defined to represent the collection segments from the beginning $(D_1 - D_3)$ to the end (D_0), where $D_{i,j}$ means the total amount of data transferred by the relay segment. Taking the segment (3,3) with the amount of 400 kB of data as an example, its corresponding $D(3,3)$ is 400, because segment (3,3) cannot be a relay station. If segment (i, j) is a relay station, its value consists of sensor data in the relay station and transmission data from other base stations. Note that $D_{i,j} = 0$ means that this segment is not a part of the selected path.

A valid route should satisfy the following conditions:

(1) Since the data can only be propagated along the $-X$ direction and the $+Y$ direction, in order to make one arbitrary segment belong to the path, it should satisfy the following conditions: the segment has the transfer data and the path matrix of this segment is not 0

$$G_{i,j} \neq 0 \wedge D_{i,j} \neq 0 \quad \exists 0 \leq i, j < 4. \tag{14}$$

(2) λ_b is introduced to represent the portion of bandwidth that the base station is allowed to use from the actual bandwidth. Note that the stability of CPMTS is the key to ensure processing; however, the excessive bandwidth utilization and data overload will not only reduce system robustness but will also reduce machining quality. At the same time, the data transmitted by the base station should be within the bandwidth restriction

$$G_{i,j} \leq \lambda_b bw_{i,j}(e), \tag{15}$$

where $\lambda_b = 0.8$ ensures that the robustness of the system can be guaranteed on the basis of transmission efficiency

(3) Whether a path can connect with the next segment depends on its capacity, which means

$$\sum_{m=1}^{3} G_{i,j} + G_{i-m,j} \leq lb_{i-m} \left\| \sum_{n=1}^{3} G_{i,j} + G_{i,j-n} \leq lb_{i-m}. \right. \tag{16}$$

(4) If the path consists of many segments, more time will be consumed. Note that, for the data acquisition task, the total of consumed time should be under the restriction of the sampling time. A 4×4 time matrix $\mathbf{T}'$ is introduced to represent the time that the current segment transmits data to the next segment. Let $\mathbf{D}'$ represent the normalized matrix of $\mathbf{D}$, where $D'_{i,j} = 1$ means there exists a segment that is a part of a valid path

$$\mathbf{D}'\mathbf{T}' \leq t, \tag{17}$$

where t is the sampling time. Note that the route may not be unique, so considering the robustness of the system and the lifetime of the base station in each segment, the optimal solution with the minimum

average memory occupancy rate and bandwidth occupancy ratio can be expressed as

$$\begin{aligned} &\text{Minimise} && \frac{1}{n}\cdot\sum_i\sum_j\frac{D_{i,j}}{lb_{i,j}} \\ &&& \frac{1}{n}\cdot\sum_i\sum_j\frac{D_{i,j}}{bw_{i,j}(e)} \\ &\text{s.t.} && \frac{D_{i,j}}{lb_{i,j}}\le 1 \\ &&& \frac{D_{i,j}}{bw_{i,j}(e)}\le 1. \end{aligned} \tag{18}$$

When electromagnetic interference occurs, due to the strong radiation energy in the interference source center, the segment with the electromagnetic interference source cannot transmit data to the other segment. The redundancy arrangement of sensors is usually used to reduce the influence of electromagnetic interference on the surrounding area of the interference source [20]. Thus, to ensure the quality of data transmission, the path of data flow should go through the area with less interference. The level of interference is proportional to the radiation energy of the noise source. Since energy is inversely proportional to the square of the distance, interference level at (i, j) is

$$\mathrm{NL}(i,j)=\frac{n\cdot g_{i,j}}{(i-2)^2+(j-2)^2}\alpha^{n\cdot g_{i,j}}. \tag{19}$$

Here, $g_{i,j}\in(0,1]$ is sensor density at segment (i, j), n is the number of sensors, and $\alpha\in(0,1]$ is noise reduction parameters. Hence, the segment in an optimized path should be under allowable noise interference $\mathrm{NL}_{\max}$.

$$\mathrm{NL}(i,j)\le\mathrm{NL}_{\max}\quad i,j\in[1,4]. \tag{20}$$

Due to the existence of electromagnetic interference and redundant arrangement of sensors in segments, a valid route that meets equations (14) to (20) may not exist. Normally, changing segment location can address this issue; however, it costs too much and the integrity of the system is vulnerable to disruption. In this paper, an extended matrix $\mathbf{G}'_{(i+k),(j+s)}$ is intended to create a new path for the disabled segment, where k is expansion depth and j is the extension layer. To obtain a valid route under the original basis, we map the element in $\mathbf{G}_{i,j}$ into the extend matrix $\mathbf{G}'_{(i+k),(j+s)}$,where $g_{i,j}=g'_{i+k,j}$. Then, we modify the path $\mathbf{D}'$ by determining the starting and ending points of the path. To ensure the shortest path, discriminant matrix $\mathbf{One}_{i+k,j+s}$ is introduced, where the rank of $\mathbf{One}_{i+k,j+s}$ is 1. So, the minimum of $\mathbf{D}'\cdot\mathbf{One}$ corresponding to $\mathbf{D}'$ is the shortest route.

Note that isolated islands may occur during path optimization, which means that there is no valid path for the data in this segment to be effectively transferred or the resource utilization of some segment in path optimization is almost full. For instance, as shown in Figure 6, when electromagnetic interference occurs in segment (1,2) where the redundant arrangement of sensors is applied in the segment nearby, no valid path exists. At this time, in order not to change the overall structure of the sensors, we extend the current segment to the next layer and take depth expansion, so that the current segment can reach depth D_0 with the shortest effective route. If it is difficult to transmit data between routers in depth D_0, an additional router is used in the extended segment in D_0. Finally, if the current extension layer does not meet the requirements, keep repeating the above steps to m layer. Since there are no sensors in each segment of the extended layer and electromagnetic interference is low, m is equal to 1.

3.2. MILP Formulation. From the equations discussed above, we can see that sensor path optimization is a multiobjective optimization problem. However, in multiobjective issues, optimization of a target may lead to deterioration of other objectives. When the objective function is in a conflict state, there will be no optimal solution to make all the objective functions reach the maximum or minimum value at the same time, so we can only seek the noninferior solutions (or Pareto solutions). As bandwidth utilization within a certain range has less impact on the system in the process of data flow, the "Constraint Model" is used to get optimization solutions from Pareto solutions, where average bandwidth usage is limited to 80%. Then, this objective can be entered into the conditional constraint group as a constraint condition. Now, the problem for sensor path optimization is related to the well-known "multi-commodity flow problem" (MCFP) in optimization. In contrast to the MCFP, however, we do not want to allow multiple simultaneous paths through the graph, since this would imply sending the data fragmented. Instead, mixed-integer linear programming (MILP) is used for the following reasons: on one hand, the MILP formulation allows users to flexibly extend the formulation gradually in order to solve the complete sensor path optimization problem. On the other hand, the formulation of constraints is very canonical, and can thus serve as a basis for future work for finding heuristics by clearly defining which constraints should be met. Based on all the definitions and equations given previously, we can formulate a MILP problem, as shown in Figure 8, to optimize the sensor path.

There are several ways to solve the MILP problem. One is linear programming (LP) relaxation, which uses linear programming to solve the integer programming problem [11]. However, it attempts to use the approximate procedure of simply applying the simplex method to the LP relaxation and then rounding the noninteger, which will lead to an uncertain optimal integer solution. For this, we use another method called "branch and bound" to systematically enumerate all candidate solutions, then discard large subsets of fruitless candidates, by using upper and lower estimated bounds of the quantity being optimized. The key to the remarkable efficiency of the branch and bound method lies in removing some feasible solutions from a linear programming problem which will make it easier to solve.

$$\text{Minimise} \quad \frac{1}{n}\cdot\sum_{i}\sum_{j}\frac{D_{i,j}}{lb_{i,j}} \quad \text{with} \quad \text{respect} \quad \text{to:} \quad \frac{D_{i,j}}{lb_{i,j}} \le 1$$

$$\text{For all } s_i \in \mathbf{S_q},\ m_j \in \mathbf{M_q} \quad : \quad u_{ij}=1 \qquad \text{---(8)}$$

$$\text{For all } f \in \mathbf{F}, T_k \subseteq \mathbf{T}, e' \in \mathbf{E}' :$$

$$\sum_{\substack{\pi_k \in \Pi_k \\ e' \in \mathbf{E}'}} \sum_{\substack{f_i, f_j \in \mathbf{F} \\ \pi_{ki} \in fp(p_i, lb_j)^T, \pi_{kj} \in fp(p_j, m)^T}} P_{f_i, \pi_{ki}} |T_k| d_i + P_{f_j, \pi_{kj}} |T_k| d_j \le bw(e') \qquad \text{---(9)}$$

$$\forall lb \in \mathbf{LB} :$$

$$\sum_{\substack{m \in M \\ lb, lbi, lbj \in M,\ \text{life}(lb) \cap T_k \ne \phi}} A_{lb,m}(\text{MID}_{lbi,lbj} + 1) \ne 0 \qquad \text{---(11)}$$

$$\sum_{\substack{m, lbi, lbj \in M \\ \text{life}(lb) \cap Tk \ne 0}} A_{lbj,m}\text{size}(lb_j) + \text{MID}_{lbi,lbj}\text{size}(lb_i) \le \text{size}(m) \qquad \text{---(12)}$$

$$\sum_{D_{bi}+\text{size}(b_i) \cap D_{bj}+\text{size}(b_j) \ne \phi} D_{bi} + \text{size}(b_i) + \text{size}(b_j) \le M_b + \text{size}(m) \qquad \text{---(13)}$$

$$D_{bi} \ge M_b$$

$$\exists 0 \le i, j < 4, G_{i,j} \ne 0 \wedge D_{i,j} \ne 0$$

$$G_{i,j} \le \lambda_b bw_{i,j}(e) \qquad \text{---(15)}$$

$$\sum_{m=1}^{3} G_{i,j} + G_{i-m,j} \le lb_{i-m} \,\|\, \sum_{n=1}^{3} G_{i,j} + G_{i,j-n} \le lb_{i-m} \qquad \text{---(16)}$$

$$\forall D(i,j) \ne 0:$$

$$\mathbf{D}'\mathbf{T}' \le t \qquad \text{---(17)}$$

$$\exists\ \text{NL}(i,j) \ne 0:$$

$$\text{NL}(i,j) \le \text{NL}_{\max} \quad i, j \in [1,4] \qquad \text{---(20)}$$

$$\frac{1}{n}\cdot\sum_{i}\sum_{j}\frac{D_{i,j}}{bw_{i,j}(e)} \le 80\%$$

Figure 8: MILP problem.

3.3. Optimization Strategy. The basic ideas of optimization strategy can be expressed as follows: first, on the basis of ensuring that the sensors of each base station can accurately obtain the monitoring quantity and signal quantity, the number and position of sensors in the base station are optimized. Next, the flow paths of sensing data in each base station are optimized so that the gathered data can be transmitted along the shortest path with fewer system resources. Then, system resource utilization is optimized to increase the robustness and integrity of CPMTS. The details of the optimization strategy are shown in Figure 9.

Step 1. Import the single-station transfer model to optimize the number and location of sensors in each base station using equation (1) to equation (5). (Complexity: $O(n)$.)

Step 2. Importing the multihop-station transfer model to develop the multihop WSN. (Complexity: $O(n)$.)

Step 3. If $\pi_{ij} = 1$ and $u_{ij} = 1$ (i.e., sensors at sink i can meet the monitoring requirements with respect to monitoring data/fault j, and data can be transmitted between sinks), import base station location information. (Complexity: $O(1)$.)

Step 4. According to base station location information, generate data flow paths. The maximum number of tool paths is $\text{Log}_m n$, because the maximum number of data flow paths is equal to the number of group levels minus one. Note that group levels are $\text{Log}_m n + 1$. (Complexity: $\text{Log}_m n$.)

Step 5. Taking electromagnetic interference into consideration, modify the generated path to ensure the accuracy of data. (Complexity: $O(1)$.)

Step 6. Using the "branch and bound" method to solve the MILP problem shown in Figure 8, the optimization path is obtained. (Complexity: $O(n^2)$.)

From the steps discussed above, the complexity of optimization algorithm is $O(n^2)$. Similarly, the space complexity can be obtained as $O(1)$.

4. Case Study

4.1. Experimental Set-up. The proposed wireless sensor deployment strategy is illustrated by a cyberphysical network system, and the corresponding layout information is shown in Table 1. Notice that strong electromagnetic interference appears in layer 2 of depth 2 so that no sensor is deployed.

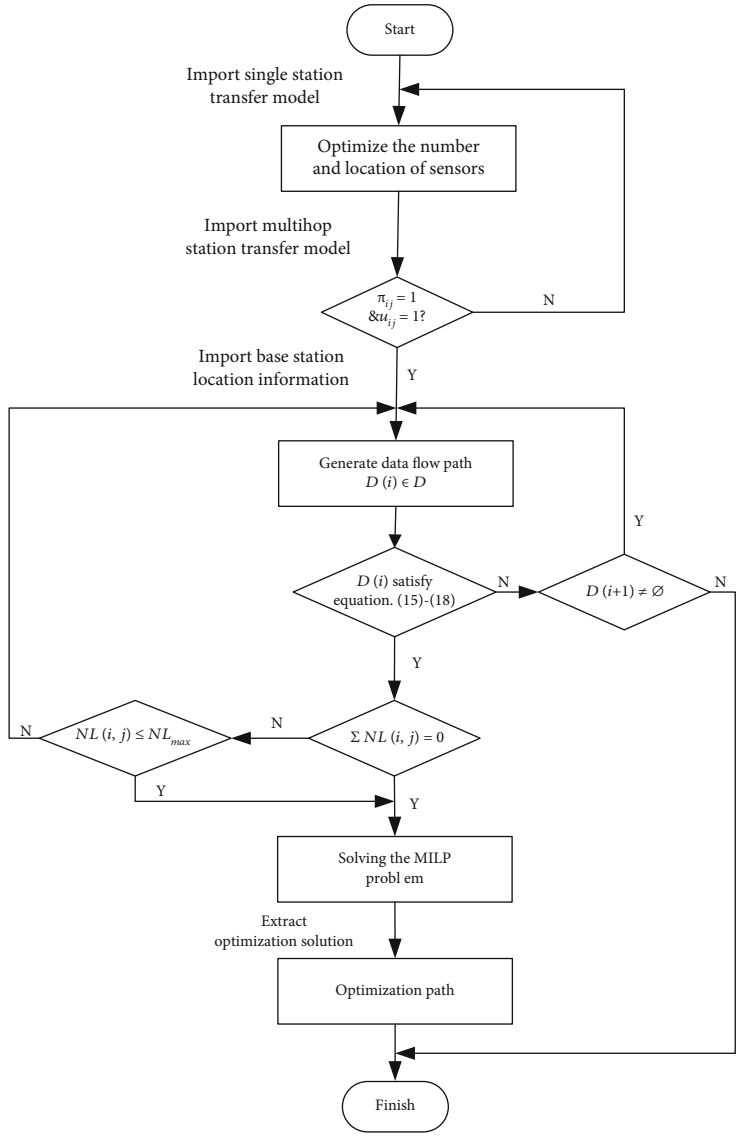


Figure 9: Flow chart of optimization strategy.

Table 1: Layout information.

Depth Layer	D0	D1	D2	D3
L0	Router	Sink 1*	Sink 4*	Sink 6*
L1	—	Sink 2*	Sink 5**	Sink 7*
L2	—	Sink 3**	Sink 9***	Sink 8**
L3	—	—	—	—

—: without sink. *Interference level.

Affected by strong electromagnetic interference in sink 9, sink 3 and sink 8 in layer 2 of depth 1 and depth 3, and sink 5 in layer 1 of depth 2, respectively, suffer from medium interference. No sinks are in layer 3 which means layer 3 can be considered as an extended layer. The distance between layers is within 10 m and the distance between depths is within 10 m.

To simplify the experimental system, only one CNC with an open structure, shown in Figure 10, is relevant to sink 1, and other sinks (from sink 2 to sink 8) are not directly embedded in the numerical control system, which means sensor data from sink 2 to sink 8 only guarantee data volume in data paths but cannot participate in linkage control. The sensors consisting of hall sensors, temperature sensors, and acceleration sensors that are arranged in this experiment provide data support for the intelligent algorithm module of the open CNC system to expand the thermal characteristic detection of the machine tool spindle and the condition detection of the cutting tool. The sink used in the experiment is LINKQI PBOX1121 which meets the IEEE802.11 standard with 128 M memory and a baud rate of 9600.

4.2. Optimal Sensor Distribution on a Single Station. Taking the VMC-C50 machine tool (shown in Figure 10) monitoring thermal characteristic and tool conditions in the finishing end mill process of HSS (high speed steel) as an example

Figure 10: VMC-C50 CNC with open structure.

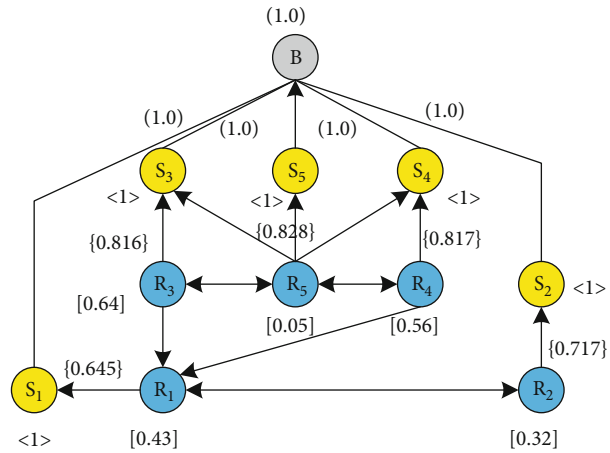


Figure 11: Initial fuzzy bipartite graph.

verifies the effectiveness of a sensor-optimized distribution in a single station (i.e., sink 1). Spindle speed is 3000 r/min, feed is 0.03 mm/z, cutting depth is 0.3 mm, and sampling frequency is 1000 Hz. By using trapezoid membership functions to normalize the sensor nodes, fault nodes, and base nodes, the initial fuzzy bipartite graph is shown in Figure 11, and characteristics of sensors and fault/monitoring volume are listed in Table 2. In addition, the transfer coefficient and the detection coefficient of each sensor node are listed in Table 3.

With the fuzzy graph and transfer coefficient, the optimization results of sensor deployment in the single station are listed in Table 4. In Table 4, there are four sensor selection strategies when electromagnetic interference is low. We can figure out that the maximum value of the detection coefficient is 0.4625, corresponding to unobservability which is -10. Although increasing the number of sensor points can further reduce unobservability, it cannot improve the detection coefficient, which may cause sensor redundancy. Moreover, we can find that sensor nodes S_1, S_2, and S_5 are fixed. According to transfer coefficient π in Table 3, spindle torque overload (R_1) and spindle overheat (R_2) are mainly based on a single sensor node (S_1 and S_2, respectively), and in order for S_1 to have certain monitoring capability for the remaining failure/monitoring quantity, sensor nodes S_1 and S_2 are required. What is more, sensor node S_5 is also a required node because it has the best detection capability for tool conditions. Hence, we choose eight sensors (S_1, S_2 (2), S_3 (2), S_4 (2), and S_5) to gain the best result. When electromagnetic interference occurs, there are two sensor selection strategies for different levels of interference, which ensure the average detection coefficient is 0.4 and unobservability is -10.

4.3. Path Optimization Experiment on Multihop Station

4.3.1. Path Optimization with Valid Solutions. With the help of optimization results on a single station, the sensor distribution strategy of the rest of the base stations is similar. The detection capability of each base station on fault/monitoring quantity is shown in Table 5. Note that sink 9 is located in layer 2 of depth 2, where strong electromagnetic interference occurs, so wireless sensing signals cannot be transmitted.

Considering that data flow from sink 8 to the router is the most representative, we optimize the data path in sink 8. The MILP problem proposed in Section 3.2 was programmed in LINGO, and we can get the optimal solution of the data path: sink 8-sink-7-sink6-sink 4-sink1-router. In contrast, we choose a Pareto solution (sink 8-sink 7-sink 5-sink 2-sink 1-router) and the shortest path solution (sink 8-sink 9-sink 3-router). In this paper, we evaluate the performance of different paths from four aspects: average memory utilization, average bandwidth occupancy, data accuracy, and average latency. The spindle speed of open CNC is 3000 r/min, the total processing time is 120 min, and the permissible bandwidth is 20 kb·s^{-1}. In addition, we assume that the sensors in each sink are enabled, respectively, from D3 to D1 and L2 to L0. The average memory utilization and average bandwidth occupancy for each solution are shown in Figures 12 and 13.

From Figure 12, we can see that in the initial phase, only sensors in sink 8 are enabled on all three paths, so the average memory utilization is the same. With sensors in other depths being enabled, due to a reduction of the influence of electromagnetic interference, the noninferior solution path adopts redundant configuration, which increases the average memory usage rate obviously. Although the average memory utilization of the optimal path still increases, the rising rate compared to the noninferior solution path is much slower, which ensures the accuracy of data transmission. Since the shortest path contains fewer sensors, the average memory usage is minimal. Note that although both the noninferior solution path and the optimal solution path are under memory constraint, the memory usage of the noninferior solution in sink 1 is approximately 90% compared to 80% of the optimal solution. Memory usage of the noninferior solution path in sink 1 will greatly increase the load of an open CNC system, which reduces the real-time performance of the system.

Figure 13 shows that the average bandwidth occupancy of each path can meet the requirement. The bandwidth curve

Table 2: Characteristics of sensors and fault/monitoring volume.

Component	Fault/monitoring volume	Object	Severity	Occurrence rate	Detection rate	Sensor
Spindle	Spindle torque overload	R_1	8	0.6	9	S_1
	Spindle overheat	R_2	8	0.5	8	S_2
Cutting tool	Tool worn	R_3	8	1.0	8	S_3
	Tool break	R_4	8	1.0	7	S_4
	Tool chatter	R_5	7	0.1	8	S_5

Table 3: Transfer coefficient π and detection coefficient φ.

Fault	Hall sensor S_1		Temperature sensor S_2		Acceleration sensor S_3		Acceleration sensor S_4		Acceleration sensor S_5	
	φ	π	φ	π	φ	π	φ	π	φ	π
R_1	0.645	1	—	—	—	—	—	—	—	—
R_2	0.645	10^{-5}	0.717	1	—	—	—	—	—	—
R_3	0.517	10^{-5}	—	—	0.816	1	0.816	1	0.817	1
R_4	0.472	1	0.706	10^{-5}	0.813	1	0.813	1	0.818	1
R_5	0.410	1	0.713	10^{-5}	0.827	1	0.825	1	0.828	1

Table 4: Optimization results.

Average required detectability	Sensor select	Achieved unobservability	Interference
0.1	S_1, S_2, S_5	-2	Low
0.2	S_1, S_2, S_3, S_4, S_5	-5	Low
0.3	S_1, S_2, S_3 (2), S_4 (2), S_5	-8	Low
0.4	S_1, S_2 (2), S_3 (2), S_4 (2), S_5	-10	Low
0.4	S_1 (2), S_2 (2), S_3 (2), S_4 (2), S_5 (2)	-10	Medium
0.4	S_1 (2), S_2 (3), S_3 (4), S_4 (4), S_5 (2)	-10	High

Table 5: Fault/monitoring volume detection capability on each sink.

	Sink 1	Sink 2	Sink 3	Sink 4	Sink 5	Sink 6	Sink 7	Sink 8	Sink 9
R_1	**o**	**o**	**o**	**o**	**o**	**o**	—	—	—
R_2	**o**	**o**	**o**	**o**	**o**	**o**	—	—	—
R_3	**o**	—	**o**	—	**o**	—	**o**	**o**	—
R_4	**o**	—	**o**	—	**o**	—	**o**	**o**	—
R_5	**o**	—	**o**	—	**o**	—	**o**	**o**	—

of the optimal solution is smoother than the noninferior path, which means bandwidth usage mutation that increases packet loss and affects data transmission will seldom occur. In addition, the lower bandwidth allowance of the noninferior path prejudices the expansion of WSN networks. Note that the shortest path still obtains the best result among the three paths, regardless of the electromagnetic interference.

Due to the influence of electromagnetic interference, packet loss and data errors may occur. In this paper, the problem data and lost data are judged by the parity bit in the data packet, and the comparison between total data and data loss/error data is shown in Figure 14.

We can see that the optimized path avoids electromagnetic interference, which reduces the average delay of data transmission at all levels of base stations and improves the accuracy of data transmission. Moreover, only 0.001% of the data that accumulated to 1310400 B in 120 minutes lost packets, while no data errors occurred. For the noninferior path, there are several base stations (sink 5, sink 8) with medium interference in the path, so the packet loss rate reaches 11% and contains 1.35% error data. It can be seen that the optimization path with less electromagnetic interference performs better in data accuracy and packet loss. Since different paths contain different numbers of sensors, the total

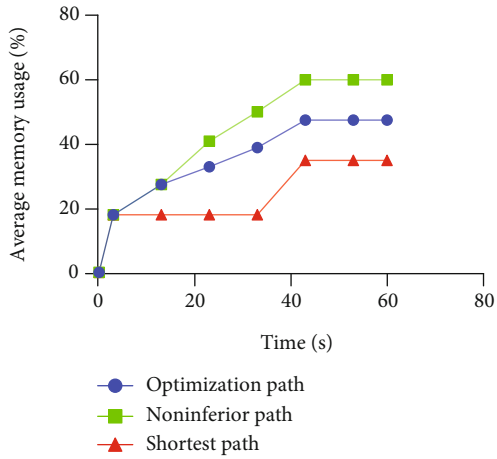


Figure 12: Average memory usage of three paths.

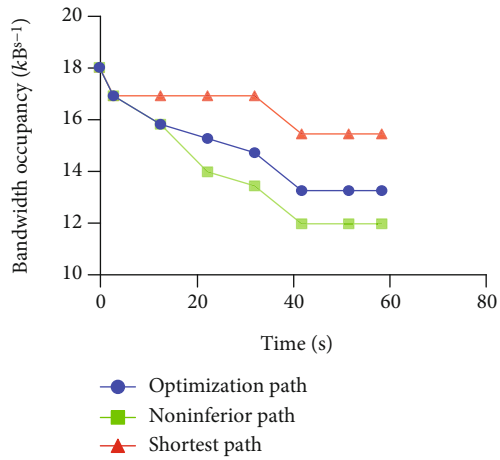


Figure 13: Bandwidth occupancy of three paths.

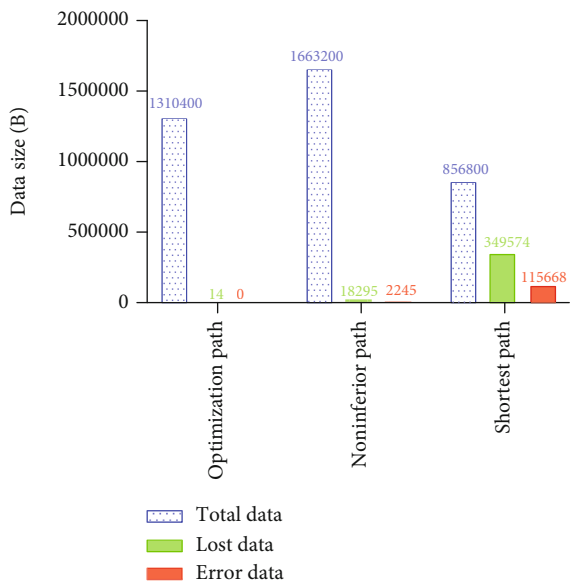


Figure 14: Comparison of packet loss data and error data.

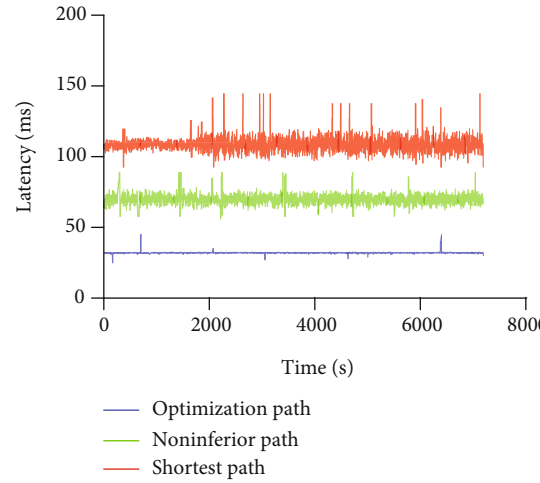


Figure 15: Latency comparison of three paths.

data amount also varies. Although the shortest path performs well in terms of average memory utilization and bandwidth occupancy, it also has the highest packet loss rate and error messages. For the shortest path, the relay station must be arranged on sink 9, where strong electromagnetic interference occurs. It gets the worst performance with 40.8% data loss and 13.5% error data.

In Figure 15, the comparison of the latency of three paths is presented. The average latency in the optimization path is 32 ms, with the highest latency of 45 ms and the lowest latency of 25 ms, and the latency curve is smooth which ensures the stability of data transmission. Note that there are a few latency oscillations in the curve, and the reason why latency oscillation occurs is that white noise occurs irregularly and medium interference exists between sink 8 and sink 7. For the noninferior path, as the amount of transmitted data (1663200 B) is greater than the optimization path, the latency will be higher even in the absence of interference. In addition, the relay station with medium interference is chosen in the noninferior path, which will affect the data transmission among three sinks (sink 7, sink 5, and sink 2). Hence, the average latency in the noninferior path is 70 ms, with the highest latency of 89 ms and the lowest latency of 58 ms. Note that the vibration of the noninferior path is greater than the optimization path, and there are several areas with strong amplitude vibration that is the mean reason for data loss and error data. With the number of base stations subject to electromagnetic interference increasing, the delay between each base station increases and more error data occurs. This phenomenon is well reflected in the shortest path. As the path is through strong electromagnetic interference, the average latency is the highest at 109 ms, with the highest latency of 145 ms and the lowest latency of 93 ms.

4.3.2. Path Optimization with Isolated Islands. Isolated islands may occur during path optimization, which means that there is no valid path for the data in one segment to be effectively transferred or the resource utilization of one segment in the path is almost full. In this paper, a strong noise source is manually added in depth 3 of layer 0 to simulate

Table 6: Layout information of isolated islands.

Depth Layer	D0	D1	D2	D3
L0	Router	Sink 1*	Sink 4**	Sink 6***
L1	Sink N4*	Sink 2*	Sink 5**	Sink 7**
L2	—	Sink 3**	Sink 9***	Sink 8**
L3	Router 3	Sink N1*	Sink N2**	Sink N3*

the isolated island situation. The layout information is shown in Table 6.

Due to the additional noise source, the sinks in depth 2 of layer 0 and depth 3 of layer 1 are affected by the new noise source, which make them both in medium interference level. Further, a redundant sensor configuration should be applied to sink 4 and sink 7. In this situation, there is only one valid path (sink 8-sink 7-sink 5-sink 2-sink 1-router) to transmit data. However, the memory usage of sink 1 in this path is almost 100%, which will cause more unpredictable problems when data transmission is in progress. Note that we call the path discussed above the "noninferior path" and put it in a comparison experiment. According to Section 3.1, we can get two optimization paths by extending graph matrix **G**. The first path is called the extensional opt path by adding three sinks and a router (i.e., sink N1, sink N2, sink N3, and router 3) in extended layer 3, and the data in route 3 is transmitted to the router via Ethernet. The extensional opt path is sink 8-sink N3-sink N2-sink N1-router 3-router. We get the second path called optimization path by adding a new sink in depth 0 of layer 1 to transmit data to the router. The optimization path is sink 8-sink 7-sink 5-sink 2-sink 4-router. Note that no sensors are in the additional sinks. In this experiment, we evaluate four paths (noninferior path, extensional opt path, optimization path, and the shortest path) from four aspects. The comparison of average memory usage is shown in Figure 16, the average bandwidth occupancy is shown in Figure 17, the data accuracy is shown in Figure 18, and the average latency is shown in Figure 19.

From Figure 16, we can figure out that the average memory usage of the extensional opt path achieves the best results because no additional sensors are used in the extended layer. However, the cost in the extensional opt path will increase greatly, and an extension path may not be generated due to space constraints. Still, compared to the extensional opt path, the shortest path gets a better result. Memory usage of the optimization path and the noninferior path are high because both of the two paths transmit data through many sinks with a redundant sensor configuration. Note that the memory usage of sink 1 in the noninferior path is almost 100%, which will greatly affect data accuracy, latency, and so on. In contrast, the optimization path effectively reduces memory usage by avoiding a busy station, which makes it possible to transmit data accurately and efficiently.

Similarly, we can find that in Figure 17 the extensional path and the shortest paths perform well in bandwidth occupancy. Since the optimization path avoids busy relay stations, the bandwidth occupancy is acceptable in spite of transmitting data through sinks with a redundant sensor configuration. The bandwidth occupancy of the noninferior path performs worst but is still within the bandwidth restriction.

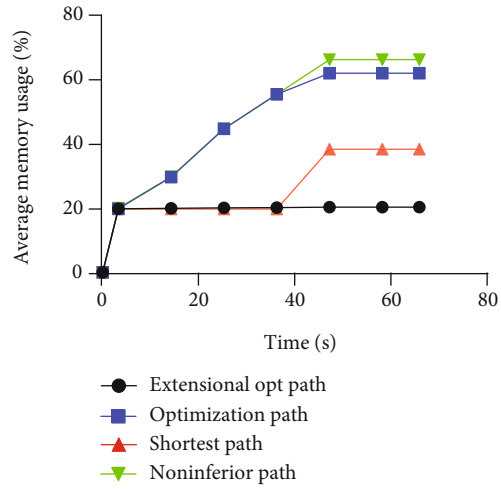


Figure 16: Average memory usage of four paths.

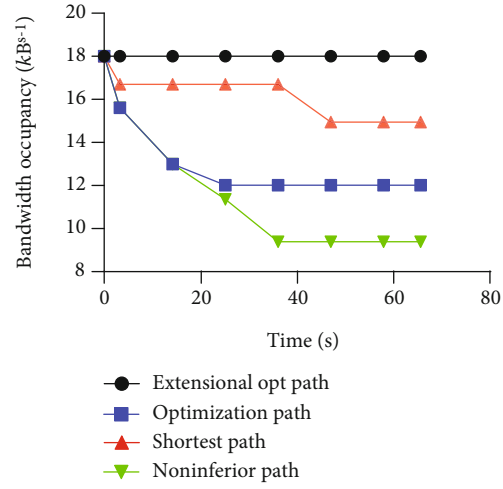


Figure 17: Bandwidth occupancy of four paths.

In Figure 18, no data packet loss and error data occur in the extensional opt path due to its shorter distance from sink 8 to the router and the lower influence of electromagnetic interference; thus, it achieves the best performance. Although the optimization path avoids strong electromagnetic interference and busy relay stations, it still has 7% data packet loss and 1.1% error data, with the total data accumulated to 1260000 B in 120 minutes. This is because the path involves three medium interference segments which will reduce the accuracy of data transmission. Even so, this is the optimal route without adding shielding devices or changing the overall sensor layout structure. The noninferior path and the shortest path contain more data packet loss and error data. For the noninferior path, it also contains three medium interference segments, and more importantly, the transmission data has to wait for memory release due to the full memory

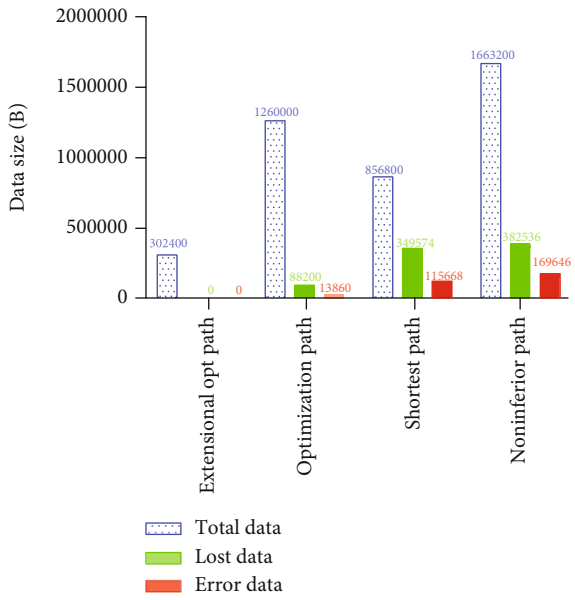


Figure 18: Comparison of packet loss data and error data.

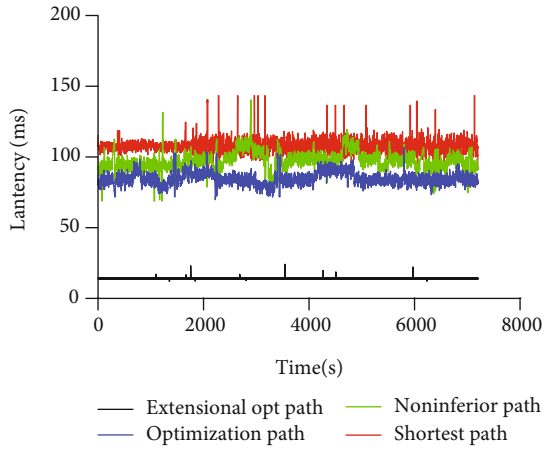


Figure 19: Latency comparison of four paths.

usage in sink 1, which cause more data packet loss and error data.

In Figure 19, the average latency in the extensional opt path is 15 ms, with the highest latency of 25 ms and the lowest latency of 13 ms, and the latency curve is smooth. Note that only a few latency oscillations occur due to the white noise. The latency curves of the optimization path (average latency: 86 ms; highest latency: 108 ms; lowest latency: 73 ms) and noninferior path (average latency: 95 ms; highest latency: 133 ms; lowest latency: 65 ms) are more shaky than the shortest path because the two paths transmit data through three medium interference segments and the interference segments form a complex electromagnetic interference field. Note that the transmission data of sink 1 in the noninferior path has to wait for memory release, which increases the latency and causes more oscillation. The shortest path still performs the worst in latency because of the strong electromagnetic interference.

5. Conclusion

Wireless sensor optimization deployment is an important research issue for CPMTS, and the accuracy and stability of transmitted fault/monitoring data are the keys to the system performance. This paper investigates a strategy for WSN distribution in CPMTS based on model development and path optimization. In the proposed model, a single-station transfer model is first represented to ensure that the layout of sensors in each sink can meet the detection capability of fault and monitoring data. By using the fuzzy graph, the multihop-station transfer model and the data-collecting model are developed to describe the data flow in a wireless network. To achieve an optimization path, noise interference and data position are taken as restrictions to optimize the path. Then, a MILP problem is formulated and the optimization solution is obtained by using the "branch and bound" method. Finally, case studies about optimal sensor distribution on a single station and path optimization on a multihop station are presented to illustrate the proposed strategy. The case study on the single station validates that the proposed sensor distribution in a single station can achieve higher detectability with fewer resources, which contributes to the increase of the real-time performance. The case studies on the multihop station validates that when there exist valid solutions, the optimization path proposed by this paper can achieve the best performance with a minimum average latency of 32 ms and a data loss of 0.001% compared to other paths. Although the shortest path performs best in the usage of bandwidth and memory, it is strongly affected by electromagnetic interference, which causes a high average latency of 109 ms, a data loss of 40.8%, and error data of 13.5% that cannot meet the requirement. Although the noninferior path satisfies the restriction of data loss and average latency, the optimization path has more available bandwidth and memory that ensures the scalability and stability of CPMTS. When isolated islands occur in the path, the proposed extensional opt path and optimization path that avoid high electromagnetic interference and busy relay stations still achieve the best result.

For future research, the authors suggest further investigations from two aspects: (1) extend the proposed strategy with an energy consumption model to further optimize the sensor distribution in CPMTS, and (2) integrate this sensor deployment strategy into a more complicated system so that it can implement the strategy more effectively and adjust the sensor deployment accordingly.

Nomenclature

G_F:	Fuzzy graph for BSD
U:	Parameter matrix
$\mathbf{A}_{i+1}$:	Dynamic matrix
$\mathbf{B}_{i+1}$:	Dynamic matrix
$\mathbf{C}_i$:	System matrix
PN:	Set of priority number (PN) value

SI: Set of sensor index (SI) value
A: Fuzzy subset
PDS: Set of paths from node D_i to node S_j
PSB: Set of paths from node S_j to node B_z
CP: Connection path
R, **E**: Sets of edge elements in fuzzy graph
SDTR: Sensor data transmission ratio
SDI: Sensor data index
$\mathbf{W}^T{}_{i,n}$: Path connection matrix from depth i to depth n
$\mathbf{V}^T{}_{i,n}$: Path connection matrix in the same depth
G_{DF}: Fuzzy graph for data-collecting architectures
M: Set of storage elements
P: Set of process elements
$\mathbf{E}'$: Set of edge elements e' in data-collecting architectures fuzzy graph
F: Set of data flows
$\boldsymbol{\Pi}$: Set of paths in data-collecting architectures fuzzy graph
$P_{f,\pi}$: Binary variable
T: Set of time operators
$A_{lb,m}$: Binary variable
$\mathrm{MID}_{lbi,lbj}$: Binary variable
D_b: Integer variable
G: 4×4 directed graph matrix
D: 4×4 path matrix
$\mathbf{T}'$: Time matrix
$\mathbf{D}'$: Normalized matrix of **D**
$\mathbf{G}'$: Extended matrix of **G**
One: Discriminant matrix.

Data Availability

The (figure) data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the Young Doctor Scientific Research Foundation of Harbin University (HUDF201712).

References

[1] C. Zhang, P. Jiang, K. Cheng, X. W. Xu, and Y. Ma, "Configuration design of the add-on cyber-physical system with CNC machine tools and its application perspectives," *Procedia Cirp*, vol. 56, pp. 360–365, 2016.

[2] K. Wang, C. Zhang, X. Xu, S. Ji, and L. Yang, "A CNC system based on real-time Ethernet and Windows NT," *The International Journal of Advanced Manufacturing Technology*, vol. 65, no. 9-12, pp. 1383–1395, 2013.

[3] Y. Yusof and K. Latif, "A novel ISO 6983 interpreter for open architecture CNC systems," *International Journal of Advanced Manufacturing Technology*, vol. 80, no. 9-12, pp. 1777–1786, 2015.

[4] Zhao Cheng, M. Perillo, and W. B. Heinzelman, "General network lifetime and cost models for evaluating sensor network deployment strategies," *IEEE Transactions on Mobile Computing*, vol. 7, no. 4, pp. 484–497, 2008.

[5] J. Wang, X. Gu, W. Liu, A. K. Sangaiah, and H. J. Kim, "An empower Hamilton loop based data collection algorithm with mobile agent for WSNs," *Human Centric Computing & Information Sciences*, vol. 9, no. 1, 2019.

[6] D. P. Kumar, T. Amgoth, and C. S. R. Annavarapu, "Machine learning algorithms for wireless sensor networks: a survey," *Information Fusion*, vol. 49, pp. 1–25, 2019.

[7] S. Toumpis and L. Tassiulas, "Optimal deployment of large wireless sensor networks," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 2935–2953, 2006.

[8] J. Wang, C. Ju, Y. Gao, A. K. Sangaiah, and G. J. Kim, "A PSO based energy efficient coverage control algorithm for wireless sensor networks," *Computers, Materials and Continua*, vol. 56, no. 3, pp. 433–446, 2018.

[9] K. Xu, Q. Wang, H. Hassanein, and G. Takahara, "Optimal wireless sensor networks (WSNs) deployment: minimum cost with lifetime constraint," in *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, pp. 454–461, Montreal, Que., Canada, August 2005.

[10] S. Chakraborty, S. Chakraborty, S. Nandi, and S. Karmakar, "Impact of redundant sensor deployment over data gathering performance: a model based approach," *Journal of Network and Computer Applications*, vol. 67, pp. 26–42, 2016.

[11] F. S. Hillier and G. J. Lieberman, "Introduction to Operations Research and Revised CD-ROM 8," *McGraw-Hill Science/Engineering/Math*, McGraw-Hill, 2005.

[12] P. Chanak and I. Banerjee, "Fuzzy rule-based faulty node classification and management scheme for large scale wireless sensor networks," *Expert Systems with Applications*, vol. 45, pp. 307–321, 2016.

[13] F. H. Awad, "Optimization of relay node deployment for multisource multipath routing in wireless multimedia sensor networks using Gaussian distribution," *Computer networks*, vol. 145, pp. 96–106, 2018.

[14] A. Céspedes-Mota, G. Castañón, A. F. Martínez-Herrera, L. E. Cárdenas-Barrón, and A. M. Sarmiento, "Differential evolution algorithm applied to wireless sensor distribution on different geometric shapes with area and energy optimization," *Journal of Network and Computer Applications*, vol. 119, no. 1, pp. 14–23, 2018.

[15] M. Khalesian and M. R. Delavar, "Wireless sensors deployment optimization using a constrained Pareto-based multiobjective evolutionary approach," *Engineering Applications of Artificial Intelligence*, vol. 53, pp. 126–139, 2016.

[16] A. U. Rahman, A. Alharby, H. Hasbullah, and K. Almuzaini, "Corona based deployment strategies in wireless sensor network: a survey," *Journal of Network & Computer Applications*, vol. 64, pp. 176–193, 2016.

[17] H. S. Ko, H. Lim, W. Jeong, and S. Y. Nof, "A statistical analysis of interference and effective deployment strategies for facility-specific wireless sensor networks," *Computers in Industry*, vol. 61, no. 5, pp. 472–479, 2010.

[18] J. Wang, Y. Gao, C. Zhou, R. S. Sherratt, and L. Wang, "Optimal coverage multi-path scheduling scheme with multiple mobile sinks for WSNs," *Computers, Materials and Continua*, vol. 61, no. 3, pp. 695–711, 2020.

[19] M. Collotta, G. Pau, and A. V. Bobovich, "A fuzzy data fusion solution to enhance the QoS and the energy consumption in wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 2017, 10 pages, 2017.

[20] J. Sánchez-Oro, M. Sevaux, A. Rossi, R. Martí, and A. Duarte, "Solving dynamic memory allocation problems in embedded systems with parallel variable neighborhood search strategies," *Electronic Notes in Discrete Mathematics*, vol. 47, pp. 85–92, 2015.

[21] A. Goens, J. Castrillon, M. Odendahl, and R. Leupers, "An optimal allocation of memory buffers for complex multicore platforms," *Journal of Systems Architecture.*, vol. 66, no. 1, pp. 69–83, 2016.

[22] M. Soto, A. Rossi, and M. Sevaux, "Iterative approaches for a dynamic memory allocation problem in embedded systems," *European Journal of Operational Research*, vol. 231, no. 1, pp. 34–42, 2013.

[23] Yu Ding, Pansoo Kim, D. Ceglarek, and Jionghua Jin, "Optimal sensor distribution for variation diagnosis in multistation assembly processes," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 4, pp. 543–556, 2003.

[24] Z. Wu, S. J. Hsieh, and J. Li, "Sensor deployment based on fuzzy graph considering heterogeneity and multiple-objectives to diagnose manufacturing system," *Robotics and Computer-Integrated Manufacturing*, vol. 29, no. 1, pp. 192–208, 2013.