AoT-Driven Resource Reservation Based on Associated Network Slice for IIoT Systems

Xiaojing Wen, *Member, IEEE*, Cailian Chen, *Senior Member, IEEE*, Xinping Guan, *Fellow, IEEE*, Cheng Ren, *Member, IEEE*, Yehan Ma, *Member, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—Joint estimation is crucial in the industrial Internet of Things (IIoT) by integrating data from diverse devices to improve monitoring accuracy. Network slicing can meet the heterogeneous needs of devices through logical isolation. However, existing methods often overlook the interaction of multiple slices on estimation performance, leading to potential estimation bias and ineffective resource costs. To address this, we propose an Age of Task (AoT)-driven associated network slicing method tailored for joint estimation scenarios. Specifically, we design an association-oriented slicing architecture for joint estimation that considers both the heterogeneous requirements of individual slices and the interactive effects of multiple slices. We define slice association based on the AoT to quantify the coupling relationship between slicing strategies and estimated performances. Moreover, we develop a dynamic-fitness multivariable particle swarm optimization algorithm to achieve associated slicing. Simulation results show that the associated slicing scheme achieves a flexible balance between timeliness and accuracy.**

***Index Terms*—Age of task (AoT), industrial Internet of Things (IIoT) system, joint estimation, network slicing.**

I. Introduction

RECENT advancements in industrial Internet of Things (IIoT) technology have paved the way for various industrial services, including intelligent manufacturing and production monitoring [1], [2]. To effectively monitor the states of these industrial services, joint estimation integrates observations from different devices, enabling a more comprehensive and accurate understanding of the physical system. For example, the measurement of rotation angles of movable surfaces is crucial in flight control system functional testing (FCSFT) [3]. Contact sensors (such as inclinometers or gyroscopes) and noncontact sensors (such as cameras or laser trackers) are commonly used for this purpose. The former method uses sensors fixed on the wing surface for intensive sampling, which can achieve real-time tracking of angles, but the measurement accuracy is relatively low, with an indicative error of approximately $\pm 0.2^\circ$ [4]. The latter method involves installing calibration points on the wings and collecting angle variations through images. This approach can achieve high-precision estimation with an error not exceeding 0.04° [5]. However, due to its resource-intensive nature, it results in poor real-time performance. Therefore, joint estimation emerges as an effective approach for timely and reliable estimation of aircraft movable surface angles. In this context, a primary challenge arises in achieving the reasonable allocation of limited resources to improve the real-time performance and reliability of joint estimation while ensuring heterogeneous quality of service (QoS) requirements.

Network slicing, as a promising technique, can satisfy different QoS requirements by building multiple logically isolated slices on the existing network infrastructure [6], [7]. In addition, given that IIoT services often have more stringent QoS requirements, a comprehensive industrial network slicing strategy should consider transmission and computing resource allocation. That makes network slicing particularly attractive when combined with edge computing [8], as it enables low-latency access to customized computing resources at the edge [9], effectively addressing the heterogeneous needs of sensors in joint estimation scenarios. In a joint estimation system, where multiple slices contribute to the same observation task, the system performance is influenced by all slice measurements. This interaction, which we define as the interslice association, plays a crucial role in enhancing system performance and guiding the design of effective slice strategies. However, existing slicing methods mainly consider the demand guarantee of a single slice. Even if a few studies consider the interslice influence, this consideration typically remains at the network level, making it difficult to meet the sensing requirements of joint estimation scenarios.

In this case, developing optimal network slicing strategies under joint estimation scenarios presents several challenges. First, most existing slicing architectures overlook the requirements of system sensing performance to allocation strategies, which cannot be directly applied in joint estimation scenarios. Second, the effects of multiple slices on system performance have not been well investigated, and appropriate metrics to characterize their association are lacking. Third, network slicing strategies

Received 3 March 2024; revised 16 August 2024 and 13 September 2024; accepted 24 September 2024. Date of publication 16 December 2024; date of current version 5 March 2025. This work was supported by the National Natural Science Foundation of China under Grant 92167205, Grant 62025305, Grant 61933009, and Grant 62432009. Paper no. TII-24-0962. *(Corresponding author: Cailian Chen.)*

Xiaojing Wen, Cailian Chen, Xinping Guan, and Cheng Ren are with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China, also with the Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China, and also with the Shanghai Engineering Research Center of Intelligent Control and Management, Shanghai 200240, China (e-mail: xiaojingwen@sjtu.edu.cn; cailianchen@sjtu.edu.cn; xpguan@sjtu.edu.cn; rencheng@sjtu.edu.cn).

Yehan Ma is with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: yehanma@sjtu.edu.cn).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TII.2024.3475423

are influenced by various constraints, including resource limitations, slice association, and estimation convergence. Consequently, an optimal network slicing strategy must encompass all these factors to ensure effective joint estimation across multiple types of devices.

As is well-known, estimation accuracy is greatly influenced by the timeliness of the measurement states. Therefore, the age of information (AoI) [10] has been extensively studied as a metric to measure the freshness of available state information. It is particularly suitable for time-sensitive and high-reliability data in IIoT applications [11], [12]. However, most existing works primarily focus on single-source tasks and may not be directly applicable to the joint estimation scenario involving multiple sources. Fortunately, Wen et al. introduced the concept of age of task (AoT) [13] to quantify the freshness of multielement and compute-intensive tasks, which aligns well with the requirements of our specific application scenarios. Recognizing the strengths of AoT in capturing the impact of transmission and computing delays in IIoT applications, we aim to incorporate AoT metrics into the representation of slice association metrics for joint estimation. The contributions of this article are summarized as follows.

1) To address the diverse QoS requirements of different sensors and their collective impact on the system performance in joint estimation scenarios, a novel association-oriented network slicing architecture is proposed, considering the heterogeneous requirements of individual slices and the associated impact of multiple slices.
2) We define the slice association metric from the sensing perspective and utilize the AoT to reveal the coupling relationship between slicing strategies and system estimation performance. The process of associated network slicing is then formulated as a resource reservation optimization problem, aiming to minimize the overall estimation error under coupling constraints.
3) Recognizing the tight coupling relationship between estimation weight factors and slice reservation ratios, we first provide the optimal resource reservation ratio under the fixed weight factors. We then develop a dynamic-fitness multivariable particle swarm optimization (DMPSO) algorithm to achieve their dynamic optimization, thereby achieving the associated network slicing.

The rest of this article is organized as follows. In Section II, related works are introduced. Section III presents the system model and problem formulation, followed by the joint optimization of interslice transmission and computing resource reservation in Section IV. Section V shows the simulation results. Finally, Section VI concludes this article.

II. Related Works

A. Multisensor Joint Estimation

In recent years, joint estimation has emerged as an important approach to enhance sensing accuracy. Some studies focus on information fusion strategies [14], [15]. For example, Guo et al. [14] proposed a visual and dynamic information fusion strategy to acquire parameter information accurately. Senel et al. [15] developed a modular real-time multisensor fusion framework to achieve environment awareness in challenging time-critical applications. Other studies focus on multidevice collaborative computing models [16], [17], [18]. Chen et al. [16] proposed a new intelligent game anti-interference collaborative computing model to enhance cross-dimensional anti-interference performance in multidevice collaborative computing environments. Li et al. [17] introduced a task-driven activity model to capture activity probabilities in multidevice collaboration scenarios, thereby improving activity detection and channel estimation accuracy. However, these studies primarily focus on fusion strategies after obtaining information from multiple sources, neglecting the need to ensure QoS requirements for individual sensors during the information acquisition process. Furthermore, limited transmission and computing resources in IIoT systems pose significant challenges in allocating resources effectively while ensuring heterogeneous QoS requirements and joint estimation reliability.

B. Network Slicing in IIoT Systems

Network slicing has become the basic framework of industrial IoT systems [6], which enables multiple isolated logical networks (i.e., slices) to coexist on a shared physical network infrastructure, has attracted great attention from academia and industry. There are two specific goals: minimizing energy consumption or system cost while satisfying QoS demands [19], [20] and maximizing service capacity or profit under physical network resource constraints [21], [22]. For example, Jovsilo et al. [19] proposed a joint slice selection and edge resource management problem to minimize computational task completion time. Feng et al. [21] developed a framework that optimizes dynamic slice request admission, user association, CPU-cycle frequency, subcarrier allocation, and power allocation while considering spatial and temporal traffic variations. However, these studies primarily focus on ensuring heterogeneous requirements and overlook the impact of slice interactions on the system performance, which prevents them from maximizing their contribution in the joint estimation.

C. Interslices Interactions

Several scholars have investigated interslice interference or effects, conducting relevant research in this area. On the one hand, they have proposed interslice allocation strategies to mitigate interslice interference or interruptions [23], [24]. For instance, Zhou et al. [23] developed a slicing-based resource management framework, which focused on interference for multiple slices to balance adaptivity and overhead of resource management. On the other hand, they have considered interslice resource orchestration or multiplexing to enhance throughput or service rate [25], [26]. For example, Zhang et al. [25] introduced a service-oriented hierarchical soft-slicing framework that improved throughput through opportunistic resource reuse at both intra- and interslice levels. While these studies examine slice interactions, they primarily focus on network performance interference or network resource constraints. The unique aspect of the joint estimation scenario lies in network resource allocation strategies from the perspective of sensing performance. Therefore, a crucial challenge arises in defining slice association metrics that are suitable for joint estimation scenarios and effectively capturing the coupling impact between multiple slices on system performance.

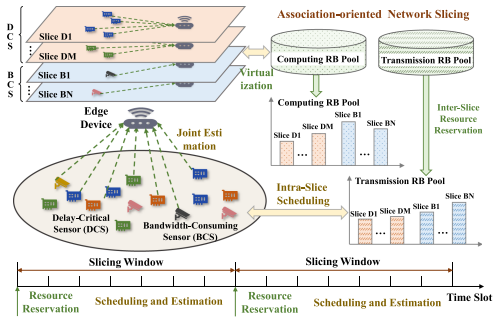


Fig. 1. Association-oriented network slicing architecture.

III. System Models and Problem Formulation

This article aims to address the challenges of heterogeneous demand guarantee and improvement of estimation performance in the joint estimation scenario involving multiple types of devices. To tackle this, this article proposes an association-oriented network slicing architecture that considers both interslice association and intraslice QoS requirements. Based on existing research and actual industrial production needs, the sensors considered in this article are divided into two types: delay-critical sensors (DCSs) and bandwidth-consuming sensors (BCSs). Each type of sensor is divided into multiple slices based on performance requirements. Resources are collaboratively reserved between slices to ensure the timeliness and reliability of observation tasks.

A. Association-Oriented Network Slicing Architecture

The proposed association-oriented network slicing architecture depicted in Fig. 1 consists of $S_{\mathcal{D}}$ DCSs, $S_{\mathcal{B}}$ BCSs, and one edge device, which are divided into M delay-critical slices and N bandwidth-consuming slices collaborating on the same observation task. To effectively capture the system dynamics, this article employs dual time scales. Scheduling and estimation processes are performed at a smaller time scale, corresponding to time slots (discussed in Section III-B), and at the larger time scale, known as slice windows (explained in Section III-C), transmission and computing resource reservation occur. At the start of each slice window, the edge device establishes two virtual resource pools: one for transmission and another for computing resources. Resources are dynamically allocated based on interassociation and intraresource constraints. The slice window changes dynamically with the duration of each task. Within each slice window, the measured values are uploaded and processed by each slice, utilizing the reserved transmission and computing resources. The fusion estimation is carried out by a Kalman filter at the edge device, serving as the basis for subsequent control actions. Without loss of generality, the resource reservation policy is assumed to remain unchanged within a slice window.

Let S_{Dm} represent the number of DCSs in Slice Dm, $\forall m \in M$, where each DCS samples data at a period of T_{Dm}. Similarly, S_{Bn} and T_{Bn} represent the number of BCSs and sampling period in Slice $Bn, \forall n \in N$, respectively. The data packets generated by the DCSs and BCSs, which are transmitted to the edge device via 5G communication, are denoted as $\mathcal{U}_{Dm,i} \triangleq (\lambda_{Dm,i}, X_{Dm,i}), \forall m \in M, i \in S_{Dm}$ and $\mathcal{U}_{Bn,j} \triangleq (\lambda_{Bn,j}, X_{Bn,j}), \forall n \in N, j \in S_{Bn}$, respectively. Here, $\lambda_{Dm,i}$ and $\lambda_{Bn,j}$ represent the input-data size (in bits), while $X_{Dm,i}$ and $X_{Bn,j}$ denote the computing intensity (in CPU cycles per bit) of DCS i in Slice Dm and BCS j in Slice Bn, respectively. This article assumes a homogeneous network, implying that all sensors in the same slice have the same packet size and computing intensity. For ease of notation, this article drops the subscript i or j and employs $\mathcal{U}_{Dm} \triangleq (\lambda_{Dm}, X_{Dm})$ or $\mathcal{U}_{Bn} \triangleq (\lambda_{Bn}, X_{Bn})$ to uniformly represent Slice Dm or Slice Bn. In the following sections, we will provide detailed models for the intraslice system and scheduling at the time slot scale, as well as discuss the details of the interslice association model at the slice window scale.

B. Intraslice System and Scheduling Model

Since all slices observe the same physical process in the joint estimation scenario, they share the same state equation. The dynamics of this hierarchical slicing architecture can be represented by the following discrete-time system:

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) + \mathbf{w}(t) \tag{1}$$

where $\mathbf{x}(t)$ represents the system state at time slot t, $\mathbf{u}(t)$ is the control command generated by the edge device. The state-transition matrix is denoted as A, and the control input matrix is represented by B. The system noise $\mathbf{w}(t)$ follows a Gaussian distribution with zero mean and covariance matrix $\Sigma_w \succ 0$. It is assumed that the control system is controllable but unstable, indicated by $\rho(A) > 1$, and the control pair $(A, \Sigma_w^{\frac{1}{2}})$ is stabilized. The initial states $x(0)$ are randomly chosen from an arbitrary finite-moment distribution with mean $\overline{x}(0)$ and variance $\Sigma_{x(0)} \succ 0$.

As depicted in Fig. 1, each slice utilizes distinct measurement devices, leading to different observation equations, which can be expressed as follows:

$$\mathbf{y}_{Dm}(t) = \mathbf{x}(t) + \mathbf{v}_{Dm}(t), \mathbf{y}_{Bn}(t) = \mathbf{x}(t) + \mathbf{v}_{Bn}(t) \tag{2}$$

where the measurements obtained by Slice Dm and Slice Bn at time slot t are denoted as $\mathbf{y}_{Dm}(t) = \{y_{Dm,i}(t)\}$ and $\mathbf{y}_{Bn}(t) = \{y_{Bn,j}(t)\}$, respectively. The observation noise $\mathbf{v}_{Dm}(t)$ and $\mathbf{v}_{Bn}(t)$ are assumed to follow mutually independent and identically distributed Gaussian distributions, with realizations $\mathbf{v}_{Dm}(t) \sim \mathcal{N}(0, \Sigma_{\mathbf{v}_{Dm}})$ and $\mathbf{v}_{Bn}(t) \sim \mathcal{N}(0, \Sigma_{\mathbf{v}_{Bn}})$, where $\Sigma_{\mathbf{v}_{Dm}}$ and $\Sigma_{\mathbf{v}_{Bn}}$ are positive definite covariance matrices. It is assumed in this article that the measurement accuracy of the BCS is higher than that of the DCS, i.e., $\Sigma_{\mathbf{v}_{Dm}} > \Sigma_{\mathbf{v}_{Bn}}, \forall m \in M, n \in N$.

Next, the scheduling model of these slices is discussed in detail. We use $\tau_{Dm}(t)$, $T^{\mathrm{T}}_{Dm}(t)$, $T^{\mathrm{C}}_{Dm}(t)$ and $\tau_{Bn}(t)$, $T^{\mathrm{T}}_{Bn}(t)$, $T^{\mathrm{C}}_{Bn}(t)$ to denote the sample time, transmission delay, and computation delay of Slice Dm and Slice Bn, respectively. Furthermore, the available bandwidth is evenly split to $\mathrm{M_T}$ independent transmission resource blocks (RBs), and the bandwidth of each RB is B. The transmission RB reservation ratio of Slice Dm is denoted as $\beta_{Dm}(t) \in [0, 1]$, and the transmission rate of each DCS in Slice Dm is denoted by

$$R_{Dm}(t) = \frac{\beta_{Dm}(t)\mathrm{M_T}}{S_{Dm}} B \log_2\left(1 + \frac{\overline{h}_{Dm} p_{Dm}}{N_0 B}\right) \tag{3}$$

where $\overline{h}_{Dm}$ is the average channel gain of DCSs considering dual-scale fading, p_{Dm} is the transmission power of each DCS, and N_0 is the noise power spectral density. To this end, the transmission delay of Slice Dm is expressed as

$$T^{\mathrm{T}}_{Dm}(t) = \begin{cases} \lceil \frac{\lambda_{Dm}}{R_{Dm}(t)} \rceil, & t \mod T_{Dm} = 0 \\ 0, & \text{others} \end{cases} \tag{4}$$

where λ_{Dm} is the packet size of DCSs in Slice Dm. Without loss of generality, $\lceil \cdot \rceil$ here means rounding up to the occupied integer time slots.

Upon the arrival of all measurements from Slice Dm, the edge device performs fusion to extract useful information. To realize the unified management of resources, the available CPU cycles are also evenly divided into $\mathrm{M_C}$ independent computing RBs, with each RB having C CPU cycles. Let the variable $\gamma_{Dm}(t) \in [0,1]$ denote the computing RB reservation ratio allocated to Slice Dm. Refer to [27], [28], [29], the computing delay of Slice Dm at time slot t is defined as

$$T^{\mathrm{C}}_{Dm}(t) = \begin{cases} \lceil \frac{\lambda_{Dm} X_{Dm} S_{Dm}}{\gamma_{Dm}(t) \mathrm{M_C} C} \rceil, & t \mod T_{Dm} = 0 \\ 0, & \text{others.} \end{cases} \tag{5}$$

Similarly, the transmission and computing RB reservation ratios for Slice Bn are given by $\beta_{Bn}(t)$ and $\gamma_{Bn}(t)$, respectively. Consequently, we can derive the expressions for the transmission rate, transmission delay, and computing delay of Slice Bn as follows:

$$R_{Bn}(t) = \frac{\beta_{Bn}(t)\mathrm{M_T}}{S_{Bn}} B \log_2 \left(1 + \frac{\overline{h}_{Bn} p_{Bn}}{N_0 B}\right) \tag{6}$$

$$T^{\mathrm{T}}_{Bn}(t) = \begin{cases} \lceil \frac{\lambda_{Bn}}{R_{Bn}(t)} \rceil, & t \mod T_{Bn} = 0 \\ 0, & \text{others} \end{cases} \tag{7}$$

$$T^{\mathrm{C}}_{Bn}(t) = \begin{cases} \lceil \frac{\lambda_{Bn} X_{Bn} S_{Bn}}{\gamma_{Bn}(t) \mathrm{M_C} C} \rceil, & t \mod T_{Bn} = 0 \\ 0, & \text{others} \end{cases} \tag{8}$$

where $\overline{h}_{Bn}$ is the average channel gain and p_{Bn} is the transmission power of the BCS in Slice Bn.

C. Interslice Association Model

This article is concerned with the estimation performance of the overall system, so the mean squared error (MSE) is used to reveal the coupling relationship between all slices.

We denote the state estimation at the edge device of Slice Dm at time slot t by $\hat{\mathbf{x}}_{Dm}(t)$ and define the corresponding estimation error as $\mathbf{e}_{Dm}(t)$, which is defined as

$$\mathbf{e}_{Dm}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}_{Dm}(t). \tag{9}$$

As mentioned earlier, the timeliness of Slice Dm is better, while the measurement accuracy of Slice Bn is higher. In order to improve the estimation performance, we can perform joint estimation by fusing the state information of all Slice Dm and Slice Bn, which can be represented as

$$\hat{\mathbf{x}}(t) = \sum_{m=1}^{M} \xi_{Dm} \hat{\mathbf{x}}_{Dm}(t) + \sum_{n=1}^{N} \xi_{Bn} \hat{\mathbf{x}}_{Bn}(t) \tag{10}$$

where $\hat{\mathbf{x}}(t)$ is the estimation of task calculated from all estimations $\hat{\mathbf{x}}_{Dm}(t)$ of Slice Dm and $\hat{\mathbf{x}}_{Bn}(t)$ of Slice Bn at the time slot t. In addition, the weight factor $\xi_{Dm}, \xi_{Bn} \in [0,1]$ indicates the trust degree of the edge device to the estimation of Slice Dm and Slice Bn, and $\sum_{m=1}^{M} \xi_{Dm} + \sum_{n=1}^{N} \xi_{Bn} = 1, \forall m \in M, n \in N$. If the system pays more attention to DCSs, the value of ξ_{Dm} is larger than ξ_{Bn}. Otherwise, ξ_{Dm} is less than ξ_{Bn} when the system is more concerned with BCSs. Especially $\xi_{Dm} = 0, \forall m \in M$ or $\xi_{Bn} = 0, \forall n \in N$ for the case that the system uses only one type of measurement for observation. Then, the estimation error of the whole system can be expressed as

$$\mathbf{e}(t) = \sum_{m=1}^{M} \xi_{Dm} \mathbf{e}_{Dm}(t) + \sum_{n=1}^{N} \xi_{Bn} \mathbf{e}_{Bn}(t). \tag{11}$$

From the perspective of state estimation, the MSE is employed to evaluate the estimation performance and given by

$$\begin{aligned} \mathcal{J}(t) &= \mathbb{E}\left\{\|\mathbf{e}(t)\|_2^2\right\} \\ &\le \sum_{m=1}^{M} \xi_{Dm} \mathbb{E}\left\{\|\mathbf{e}_{Dm}(t)\|_2^2\right\} + \sum_{n=1}^{N} \xi_{Bn} \mathbb{E}\left\{\|\mathbf{e}_{Bn}(t)\|_2^2\right\} \\ &= \sum_{m=1}^{M} \xi_{Dm} \mathcal{J}_{Dm}(t) + \sum_{n=1}^{N} \xi_{Bn} \mathcal{J}_{Bn}(t). \end{aligned} \tag{12}$$

D. Problem Formulation

To improve the adaptability of monitoring tasks to system dynamics and resource constraints, a joint design of transmission and computing RB reservation ratio under associated slice constraints is necessary. The objective is to minimize the estimation distortion of the overall task. Within the duration of T time slots in a slicing window, we formulate the following optimization problem, considering network resource limitations and estimation convergence constraints:

$$\mathcal{P}: \min_{\boldsymbol{\beta}(t), \boldsymbol{\gamma}(t)} \frac{1}{T} \sum_{t=1}^{T} \left[\sum_{m=1}^{M} \xi_{Dm} \mathcal{J}_{Dm}(t) + \sum_{n=1}^{N} \xi_{Bn} \mathcal{J}_{Bn}(t) \right] \tag{13a}$$

$$\text{s.t. } 0 \le \beta_{Dm}(t) \le 1, 0 \le \beta_{Bn}(t) \le 1 \quad \forall m \in M, n \in N \tag{13b}$$

$$0 \le \gamma_{Dm}(t) \le 1, 0 \le \gamma_{Bn}(t) \le 1 \quad \forall m \in M, n \in N \tag{13c}$$

$$\sum_{m=1}^{M} \beta_{Dm}(t) + \sum_{n=1}^{N} \beta_{Bn}(t) = 1 \tag{13d}$$

$$\sum_{m=1}^{M} \gamma_{Dm}(t) + \sum_{n=1}^{N} \gamma_{Bn}(t) = 1 \tag{13e}$$

$$\mathbb{E}\left\{T^{\mathrm{T}}_{Dm}(t) + T^{\mathrm{C}}_{Dm}(t) \le T^{\mathrm{th}}_{Dm}\right\} \ge 1 - \frac{1}{\rho(A)^2} \tag{13f}$$

$$\Pr\left\{\left(\sum_{n=1}^{N} \beta_{Bn}(t)\right)\left(\sum_{n=1}^{N} \gamma_{Bn}(t)\right) = 0\right\} \le \frac{1}{\rho(A)^2} \tag{13g}$$

where $\boldsymbol{\beta}(t) = \{\{\beta_{Dm}\}, \{\beta_{Bn}\}\}$, $\boldsymbol{\gamma}(t) = \{\{\gamma_{Dm}\}, \{\gamma_{Bn}\}\}$, $\forall m \in M, n \in N$, and the sum of the logarithmic function $\ln(\cdot)$ in the objective function guarantees both the overall performance and the fairness. Constraints (13b)–(13e) are the resource reservation constraints to ensure that transmission and computing resources allocated to all slices should not exceed the resource capacity. Constraints (13f) and (13g) are the delay conditions and resource conditions to ensure the convergence of Slices Dm and Bn, respectively.[1] Note that $\mathcal{P}$ is a nonlinear programming

[1]To ensure the convergence of the state estimation, the probability of successful state updates should be greater than or equal to $1 - \frac{1}{\rho(A)^2}$. Specifically, for Slice Dm, this implies that the probability of the total delay is less than or equal to the maximum allowable delay T^{th}_{Dm} should be greater than or equal to $1 - \frac{1}{\rho(A)^2}$. Similarly, for all Slices Bn, it is necessary to ensure that the probability of nonscheduling is less than or equal to $\frac{1}{\rho(A)^2}$.

problem that cannot be explicitly expressed in terms of decision variables. Therefore, effective shrinkage and decomposition methods are employed to solve $\mathcal{P}$.

IV. Interslice Optimal Transmission and Computing Resource Reservation

Due to the significant impact of measurement freshness on estimation performance, it is crucial to characterize the timeliness of the measurements. The traditional AoI, which describes the timeliness of individual devices, is no longer applicable as multiple devices serve the same process in this article. Instead, we adopt the AoT concept proposed in [30], which captures the freshness of a task until the receiving end obtains the status information from all observing devices and completes the corresponding processing. According to the definition of AoT, its value can be obtained by subtracting the generation time slot of the last received states in the latest observation from the current time slot

$$\Delta_{\text{age}}(t) \triangleq t - G(\bar{t}), \bar{t} = \arg\max_{\bar{t}} \left\{\alpha(\bar{t}) \leq t\right\} \tag{14}$$

where $G(t)$ and $\alpha(t)$ represent the generation and arrival time slots of the most recent measurements in the latest observation, respectively. Specifically, the AoT for Slice Dm and Slice Bn can be expressed as follows:

$$\begin{aligned} &\Delta_{Dm}(t) \triangleq t - \tau_{Dm}(\bar{t}), \\ &\bar{t} = \arg\max_{\bar{t}} \left\{\alpha_{Dm}(\bar{t}) \leq t, T^{\mathrm{T}}_{Dm}(\bar{t}) \neq 0\right\} \end{aligned} \tag{15}$$

$$\begin{aligned} &\Delta_{Bn}(t) \triangleq t - \tau_{Bn}(\widetilde{t}), \\ &\widetilde{t} = \arg\max_{\widetilde{t}} \left\{\alpha_{Bn}(\widetilde{t}) \leq t, T^{\mathrm{T}}_{Bn}(\widetilde{t}) \neq 0\right\} \end{aligned} \tag{16}$$

where $\tau_{Dm}(t)$ and $\tau_{Bn}(t)$ are the generation time slot of the last received measurements of Slice Dm and Slice Bn, respectively. The corresponding arrive time slot of these slices are $\alpha_{Dm}(t) = \tau_{Dm}(t) + T^{\mathrm{T}}_{Dm}(t) + T^{\mathrm{C}}_{Dm}(t)$ and $\alpha_{Bn}(t) = \tau_{Bn}(t) + T^{\mathrm{T}}_{Bn}(t) + T^{\mathrm{C}}_{Bn}(t)$.

Assuming the latest received state information of Slice Dm at the edge device is $\mathbf{x}(t - \Delta_{Dm}(t))$, we can compute the state estimation performance as

$$\begin{aligned} \mathbf{e}_{Dm}(t) = &\sum_{i=1}^{\Delta_{Dm}(t)} A^{i-1}\mathbf{w}(t-i) \\ &- A^{\Delta_{Dm}(t)} K_{Dm}(t - \Delta_{Dm}(t))\mathbf{v}_{Dm}(t - \Delta_{Dm}(t)) \end{aligned} \tag{17}$$

where $K_{Dm}(t)$ is the Kalman gain of Slice Dm, which can be computed using the observation noise covariance matrix of DCS $\Sigma_{\mathbf{v}_{Dm}}$ and the covariance matrix between the predicted and true values

$$K_{Dm}(t) = \frac{\mathrm{E}\left[\left(\mathbf{x}(t) - \hat{\mathbf{x}}_{Dm}(t)\right)\left(\mathbf{x}(t) - \hat{\mathbf{x}}_{Dm}(t)\right)^{\top}\right]}{\mathrm{E}\left[\left(\mathbf{x}(t) - \hat{\mathbf{x}}_{Dm}(t)\right)\left(\mathbf{x}(t) - \hat{\mathbf{x}}_{Dm}(t)\right)^{\top}\right] + \Sigma_{\mathbf{v}_{Dm}}}.$$

Similarly, we can obtain the estimation error of Slice Bn

$$\begin{aligned} \mathbf{e}_{Bn}(t) = &\sum_{i=1}^{\Delta_{Bn}(t)} A^{i-1}\mathbf{w}(t-i) \\ &- A^{\Delta_{Bn}(t)} K_{Bn}(t - \Delta_{Bn}(t))\mathbf{v}_{Bn}(t - \Delta_{Bn}(t)) \end{aligned} \tag{18}$$

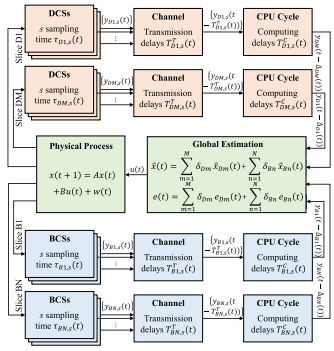


Fig. 2. Block diagram of the estimation under the associated network slicing.

where $K_{Bn}(t)$ is the Kalman gain of Slice Bn, which can be calculated by

$$K_{Bn}(t) = \frac{\mathrm{E}\left[\left(\mathbf{x}(t) - \hat{\mathbf{x}}_{Bn}(t)\right)\left(\mathbf{x}(t) - \hat{\mathbf{x}}_{Bn}(t)\right)^{\top}\right]}{\mathrm{E}\left[\left(\mathbf{x}(t) - \hat{\mathbf{x}}_{Bn}(t)\right)\left(\mathbf{x}(t) - \hat{\mathbf{x}}_{Bn}(t)\right)^{\top}\right] + \Sigma_{\mathbf{v}_{Bn}}}.$$

The whole estimation process, considering transmission and computing delays, is illustrated in Fig. 2 within the context of associated network slicing. It is important to highlight that the quantities $\mathcal{J}_{Dm}(t)$ and $\mathcal{J}_{Bn}(t)$ are influenced by $\Delta_{Dm}(t)$ and $\Delta_{Bn}(t)$. This observation further motivates our exploration of the relationship between AoT and estimation performance when resources are limited. Let $\mathcal{T} = \mathrm{Tr}(A^T A \Sigma_{\mathbf{w}})$, then $\mathcal{J}_{Dm}(t)$ can be expressed as

$$\begin{aligned} \mathcal{J}_{Dm}(t) &\leq \mathbb{E}\Bigg\{ \sum_{i=1}^{\Delta_{Dm}(t)-1} \mathrm{Tr}\left(\left(A^T A\right)^i \Sigma_{\mathbf{w}}\right) \\ &\quad + \mathrm{Tr}\left(A^T A\right)^{\Delta_{Dm}(t)} \frac{(\Sigma_{\mathbf{w}})^2 \Sigma_{\mathbf{v}_{Dm}}}{(\Sigma_{\mathbf{w}} + \Sigma_{\mathbf{v}_{Dm}})^2} \Bigg\} \\ &= \frac{(\mathcal{T})^{\mathbb{E}\{\Delta_{Dm}(t)\}} - 1}{\mathcal{T} - 1} + \underbrace{\mathrm{Tr}\left(A^T A\right)^{\mathbb{E}\{\Delta_{Dm}(t)\}} \overbrace{\frac{(\Sigma_{\mathbf{w}})^2 \Sigma_{\mathbf{v}_{Dm}}}{(\Sigma_{\mathbf{w}} + \Sigma_{\mathbf{v}_{Dm}})^2}^{\mathcal{H}_{Dm}}}_{\mathcal{G}_{Dm}} \\ &= \frac{(\mathcal{T})^{\mathbb{E}\{\Delta_{Dm}(t)\}}}{\mathcal{T} - 1} + \mathcal{G}_{Dm} \triangleq \mathcal{F}_{Dm}. \end{aligned}$$

Based on the important properties of the spectral radius, we know that the spectral radius of a matrix is not greater than any of its induced norms, i.e., $\rho(A) \leq \|A\|$. In this case, since $\mathcal{T}$ can be represented as the Frobenius norm of matrix A, i.e., $\|A\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}(A^T A)}$, we can conclude that $(\mathcal{T})^{\mathbb{E}\{\Delta_{Dm}(t)\}} \gg \rho(A) \geq 1$. Therefore, the third step naturally follows from this inequality. Similarly, the MSE of Slice Bn is

$$\begin{aligned} \mathcal{J}_{Bn}(t) &\leq \frac{(\mathcal{T})^{\mathbb{E}\{\Delta_{Bn}(t)\}}}{\mathcal{T} - 1} + \mathcal{G}_{Bn} \triangleq \mathcal{F}_{Bn} \\ \mathcal{G}_{Bn} &= \mathrm{Tr}\left(A^T A\right)^{\mathbb{E}\{\Delta_{Bn}(t)\}} \overbrace{\frac{(\Sigma_{\mathbf{w}})^2 \Sigma_{\mathbf{v}_{Bn}}}{(\Sigma_{\mathbf{w}} + \Sigma_{\mathbf{v}_{Bn}})^2}^{\mathcal{H}_{Bn}}}. \end{aligned} \tag{19}$$

The objective function $\mathcal{P}$ can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{m=1}^{M}\xi_{Dm}\ln\mathcal{F}_{Dm}+\sum_{n=1}^{N}\xi_{Bn}\ln\mathcal{F}_{Bn}\right]$$

where time accumulated objective function makes it challenging to solve this problem effectively. For the convenience of tackling this problem, we focus on the subproblem at each time slot with the objective function

$$\mathcal{P}1:\min_{\boldsymbol{\beta}(t),\boldsymbol{\gamma}(t)}\sum_{m=1}^{M}\xi_{Dm}\ln\mathcal{F}_{Dm}+\sum_{n=1}^{N}\xi_{Bn}\ln\mathcal{F}_{Bn}$$

$$\text{s.t. (13b)}-\text{(13g)} \tag{20a}$$

where $\Delta_{Dm}(t)$ and $\Delta_{Bn}(t)$ are the only time-dependent variable. Note that $\mathcal{F}_{Dm}$ and $\mathcal{F}_{Bn}$ are increasing with respect to $\mathbb{E}\{\Delta_{Dm}(t)\}$ and $\mathbb{E}\{\Delta_{Bn}(t)\}$ when the spectral radius of plant A is larger than one.

Then, we further analyze the relationship between estimation performance and AoT, obtaining

$$\begin{aligned}&\sum_{m=1}^{M}\xi_{Dm}\ln\mathcal{F}_{Dm}+\sum_{n=1}^{N}\xi_{Bn}\ln\mathcal{F}_{Bn}\\ &\le\sum_{m=1}^{M}\xi_{Dm}\mathbb{E}\left\{\Delta_{Dm}(t)\right\}\left(\ln\mathcal{T}+\ln A^{T}A+\ln\mathcal{H}_{Dm}\right)\\ &\quad+\sum_{n=1}^{N}\xi_{Bn}\mathbb{E}\left\{\Delta_{Bn}(t)\right\}\left(\ln\mathcal{T}+\ln A^{T}A+\ln\mathcal{H}_{Bn}\right)\\ &=\sum_{m=1}^{M}\xi_{Dm}\mathbb{E}\left\{\Delta_{Dm}(t)\right\}\mathcal{K}_{Dm}+\sum_{n=1}^{N}\xi_{Bn}\mathbb{E}\left\{\Delta_{Bn}(t)\right\}\mathcal{K}_{Bn}.\end{aligned} \tag{21}$$

Furthermore, the original problem is upper bounded by the linear function of the average AoT. This reveals the impact of the coupling relationship between different slices on the system performance. Based on this observation, we provide Definition 1 to formally define the association coupling relationship between slices and its influence on the system.

Definition 1: For the case where time-sensitive and bandwidth-consuming slices serve the same task, the slice association metric is a linear function related to the slice AoT

$$\sum_{m=1}^{M}\xi_{Dm}\mathbb{E}\left\{\Delta_{Dm}(t)\right\}\mathcal{K}_{Dm}+\sum_{n=1}^{N}\xi_{Bn}\mathbb{E}\left\{\Delta_{Bn}(t)\right\}\mathcal{K}_{Bn} \tag{22}$$

which reveals the coupling effect of network resource allocation strategies on the system performance and is conducive to better support for heterogeneous sensing services.

Therefore, we will pay an effort to reduce the average AoT with limited resources to enhance the estimation performance. We consider the zero-wait periodic sampling strategy ($T_{Dm}=T_{Dm}^{\mathrm{T}}+T_{Dm}^{\mathrm{C}}, T_{Bn}=T_{Bn}^{\mathrm{T}}+T_{Bn}^{\mathrm{C}}$) and the AoT diagram is shown in Fig. 3. The $\Delta_U(t)$ curve can be divided into a parallelogram of area $(T_U^{\mathrm{T}}(t)+T_U^{\mathrm{C}}(t))^2$ and a triangle of area $(T_U^{\mathrm{T}}(t)+T_U^{\mathrm{C}}(t))^2/2$. This area, averaged over time slots, captures the average AoT associated with Slice U. Hence, the average AoT of each slice can be expressed as

$$\mathbb{E}\left\{\Delta_U(t)\right\}=\frac{3}{2}(T_U^{\mathrm{T}}(t)+T_U^{\mathrm{C}}(t)) \tag{23}$$

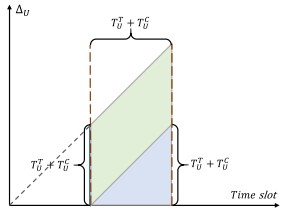


Fig. 3. AoT evolution in terms of T_U^{T} and T_U^{C}.

which makes it possible for us to decouple optimization problem $\mathcal{P}1$ into two subproblems of transmission RB reservation and computing RB reservation.

In particular, the optimization subproblem of transmission RB reservation at the tth time slot is given by

$$\mathcal{P}2:\min_{\boldsymbol{\beta}(t)}\sum_{m=1}^{M}\xi_{Dm}T_{Dm}^{\mathrm{T}}(t)\mathcal{K}_{Dm}+\sum_{n=1}^{N}\xi_{Bn}T_{Bn}^{\mathrm{T}}(t)\mathcal{K}_{Bn} \tag{24a}$$

$$\text{s.t. (13b), (13d), (13f)}-\text{(13g)}. \tag{24b}$$

Obviously, the problem $\mathcal{P}2$ is convex, which can be easily proved by deriving the second derivative of the objective function (24a) greater than or equal to zero. To tackle this problem, we utilize Karush–Kuhn–Tucker (KKT) conditions and the corresponding Lagrangian dual problem to obtain its optimal closed-form solution.

Theorem 1: Given the system described by (1) and (2) with the state-transition matrix A, the AoT of Slice Dm Δ_{Dm} and Slice Bn Δ_{Bn} defined in (15) and (16), the optimal transmission RB reservation ratio β^* can be expressed as follows:

$$\beta_u^*=\frac{iS_u}{\mathrm{M_T}}, i=\arg\left\{\frac{iS_u}{\mathrm{M_T}}\le\tilde{\beta}_u\le\frac{(i+1)S_u}{\mathrm{M_T}}\right\} \tag{25}$$

where $\tilde{\beta}_u$ is the theoretically optimal ratio, i.e.,

$$\tilde{\beta}_u=\frac{\sqrt{\xi_u\lambda_u S_u\mathcal{K}_u/\Phi_u}}{\sum_{j\in\{DM\cup BN\}}\sqrt{\xi_j\lambda_j S_j\mathcal{K}_j/\Phi_j}}\quad\forall u\in\{DM\cup BN\} \tag{26}$$

and in which $\Phi_j=\log(1+\frac{\overline{h}_j p_j}{N_0 B}),\forall j\in\{DM\cup BN\}$.

Proof: First, constraints (13d) and (13e) can be omitted since we assumed the computing RB reservation ratio γ could satisfy those. From the above, we have that $\mathcal{P}2$ is convex, and thus, its optimal solution must satisfy the KKT conditions. In order to formulate the corresponding Lagrangian dual problem, we assume nonnegative Lagrange multiplier vectors $\boldsymbol{\alpha_D},\boldsymbol{\alpha_B},\boldsymbol{\eta_D},\boldsymbol{\eta_B}$ for constraints in (24b), respectively. Next, we define the Lagrangian dual problem corresponding to $\mathcal{P}2$ as $\max_{\boldsymbol{\alpha_D},\boldsymbol{\alpha_B},\boldsymbol{\eta_D},\boldsymbol{\eta_B}}\min_{\boldsymbol{\beta}}\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\alpha_D},\boldsymbol{\alpha_B},\boldsymbol{\eta_D},\boldsymbol{\eta_B})$, where the Lagrangian dual problem is given by

$$\begin{aligned}&\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\alpha_D},\boldsymbol{\alpha_B},\boldsymbol{\eta_D},\boldsymbol{\eta_B})\\ &=\sum_{m=1}^{M}\xi_{Dm}T_{Dm}^{\mathrm{T}}(t)\mathcal{K}_{Dm}+\sum_{n=1}^{N}\xi_{Bn}T_{Bn}^{\mathrm{T}}(t)\mathcal{K}_{Bn}\\ &\quad+\sum_{m=1}^{M}\alpha_{Dm}(\beta_{Dm}-1)+\sum_{n=1}^{N}\alpha_{Bn}(\beta_{Bn}-1)\\ &\quad-\sum_{m=1}^{M}\eta_{Dm}\beta_{Dm}-\sum_{n=1}^{N}\eta_{Bn}\beta_{Bn}.\end{aligned}$$

Then, we express the KKT conditions as follows:

$$\begin{aligned}
&\textbf{stationarity}: \sum_{m=1}^{M}\alpha_{Dm}+\sum_{n=1}^{N}\alpha_{Bn}-\sum_{m=1}^{M}\eta_{Dm}-\sum_{n=1}^{N}\eta_{Bn}\\
&=\frac{\sum_{m=1}^{M}\xi_{Dm}\lambda_{Dm}\mathcal{K}_{Dm}S_{Dm}}{\beta_{Dm}^2\mathrm{M_T}\log\left(1+\frac{\overline{h}_{Dm}p_{Dm}}{N_0B}\right)}+\frac{\sum_{n=1}^{N}\xi_{Bn}\lambda_{Bn}\mathcal{K}_{Bn}S_{Bn}}{\beta_{Bn}^2\mathrm{M_T}\log\left(1+\frac{\overline{h}_{Bn}p_{Bn}}{N_0B}\right)}
\end{aligned} \tag{27}$$

$$\begin{aligned}
&\textbf{pr.feasibility}: \ (24b)\\
&\textbf{du.feasibility}: \ \boldsymbol{\alpha_D},\boldsymbol{\alpha_B},\boldsymbol{\eta_D},\boldsymbol{\eta_B}\succeq 0\\
&\textbf{co.slackness}: \ \alpha_{Dm}(\beta_{Dm}-1)=0,\alpha_{Bn}(\beta_{Bn}-1)=0\\
&\quad -\eta_{Dm}\beta_{Dm}=0,-\eta_{Bn}\beta_{Bn}=0 \quad \forall m\in M,n\in N.
\end{aligned} \tag{28}$$

We proceed with finding $\boldsymbol{\beta}$. First, from the KKT dual feasibility condition $\boldsymbol{\alpha_D},\boldsymbol{\alpha_B},\boldsymbol{\eta_D},\boldsymbol{\eta_B}\succeq 0$ and the complementary slackness condition (28), we obtain that $\boldsymbol{\alpha_D}=\boldsymbol{\alpha_B}=\boldsymbol{\eta_D}=\boldsymbol{\eta_B}=0$ must hold as otherwise $\boldsymbol{\beta}=0$ would lead to infinite value of the objective function. Then, we obtain the expression (26) for ratio $\tilde{\boldsymbol{\beta}}$ from the KKT stationarity condition (27) and complementary slackness condition (28). Since this article considers a homogeneous environment and a transmission RB can only be allocated to one sensor in a time slot, the slice AoT can only be changed if the number of RBs allocated to Slice Dm is an integer multiple of S_{Dm}. Furthermore, the resource reservation ratio needs to be quantified by $S_{Dm}/\mathrm{M_T}$, which proves the Theorem 1. ■

Similarly, the subproblem of computing resource reservation can be expressed as

$$\mathcal{P}3: \min_{\boldsymbol{\gamma}(t)} \sum_{m=1}^{M}\xi_{Dm}T_{Dm}^{\mathrm{C}}(t)\mathcal{K}_{Dm}+\sum_{n=1}^{N}\xi_{Bn}T_{Bn}^{\mathrm{C}}(t)\mathcal{K}_{Bn} \tag{28a}$$

$$\text{s.t. } (13c),(13e)-(13g). \tag{28b}$$

Theorem 2: Consider system (1) and (2) with the state-transition matrix A, the AoT of Slice Dm Δ_{Dm} and Slice Bn Δ_{Bn} defined in (15) and (16), the optimal computing RB reservation ratio γ^* can be expressed as

$$\gamma_u^*=\frac{iS_u}{\mathrm{M_C}}, i=\arg\left\{\frac{iS_u}{\mathrm{M_C}}\le\tilde{\gamma}_u\le\frac{(i+1)S_u}{\mathrm{M_T}}\right\} \tag{29}$$

where the theoretically optimal reservation ratio is

$$\tilde{\gamma}_u=\frac{\sqrt{\xi_u\lambda_u S_u\mathcal{K}_u X_u}}{\sum_{j\in\{DM\cup BN\}}\sqrt{\xi_j\lambda_j S_j\mathcal{K}_j X_j}} \quad \forall u\in\{DM\cup BN\}. \tag{30}$$

Proof: Using the same approach presented in the proof of Theorem 1, Theorem 2 can be proven. Due to the limitation of length, there is no more tautology here. ■

So far, we have demonstrated that the original problem can be decomposed into two interconnected resource reservation problems that can be solved sequentially. The next step involves optimizing $\xi_s,\forall s\in\{Dm\cup Bn\},m\in M,n\in N$ to minimize the overall MSE for all slices. It is important to note that the weight factors influence the slice reservation ratio as per Theorems 1 and 2, and the resource reservation policy impacts the slice estimation performance and, consequently, the weight factor allocation. This interdependent relationship leads to a dynamically changing objective function, making it challenging to solve this problem using traditional centralized exhaustive methods. To address this, we develop a DMPSO algorithm to solve this problem efficiently.

Algorithm 1: Dynamic-Fitness Multi-Variable PSO Algorithm for Associated Network Slicing (ANS).

Input: The data packs and the available network RBs;
Output: The inter-slice transmission and computing RB reservation ratio $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}^*$ and the slice weight factor $\boldsymbol{\xi}$ for all slices;

1 Set population parameters and randomly generate the initial population, where the historical best fitness of the group is $\mathcal{F}_{gbest}^1$;
2 Initialize iterative number $i=1$;
3 **repeat**
4 **for** *The population number* $j=1:N_{pop}$ **do**
5 Update the velocity $pop_v^i(:,j)$ and position $pop_x^i(:,j)$ and perform boundary processing;
6 Substitute the updated position $\mathcal{F}_{pop}^i(j)$ into Theorems 1 and 2 to judge constraints and calculate the individual fitness;
7 **if** $\mathcal{F}_{pop}^i(j)<\mathcal{F}_{ibest}^i(j)$ **then**
8 $ibest(:,j)=pop_x(:,j)$;
9 $\mathcal{F}_{ibest}^i(j)=\mathcal{F}_{pop}^i(j)$;
10 **end**
11 **if** $\mathcal{F}_{ibest}^i(j)<\mathcal{F}_{gbest}$ **then**
12 $gbest=ibest(:,j)$;
13 $\mathcal{F}_{gbest}=\mathcal{F}_{ibest}^i(j)$;
14 **end**
15 Update iterative number $i=i+1$;
16 **end**
17 **until** $i>ger\|\frac{1}{2}\sum_{i=1}^{2}\left|\mathcal{F}_{gbest}^i-\mathcal{F}_{gbest}^{i-1}\right|<\varepsilon$;
18 Plug $\boldsymbol{\xi}=gbest$ into (25) and (30) to get the optimal resource allocation ratio;
19 Return reservation ratios $\boldsymbol{\beta}^*$, $\boldsymbol{\gamma}^*$ and weight factors $\boldsymbol{\xi}$.

In the following, we elaborate on the DMPSO in detail, presented in Algorithm 1. In the initialization phase, relevant parameters such as population number N_{pop}, spatial dimension dim, and maximum iteration number ger are initialized, where the number of slices determines $\text{dim}=M+N$. Next, the initial population position and velocity are randomly generated, and then the historical best position and fitness of individuals and groups are initialized, as shown in Step 1. Then, the particle swarm iterates by updating the velocities and positions while performing boundary processing, as depicted in Step 5. Following this, the constraints are evaluated, and the individual fitness of updated positions is calculated in Step 6. By comparing the fitness values, the historical best fitness of individuals and groups is updated, as outlined in Steps 7 to 14. The error function is defined as $\frac{1}{2}\sum_{i=1}^{2}|\mathcal{F}_{\text{gbest}}^i-\mathcal{F}_{\text{gbest}}^{i-1}|<\varepsilon$, where $\mathcal{F}_{\text{gbest}}=\sum_{m=1}^{M}\xi_{Dm}(T_{Dm}^{\mathrm{T}}(t)+T_{Dm}^{\mathrm{C}}(t))\mathcal{K}_{Dm}+\sum_{n=1}^{N}\xi_{Bn}(T_{Bn}^{\mathrm{T}}(t)+T_{Bn}^{\mathrm{C}}(t))\mathcal{K}_{Bn}$ changes dynamically with the population position, and $\mathcal{F}_{\text{gbest}}^i$ and $\mathcal{F}_{\text{gbest}}^{i-1}$ are the estimation cost of iteration i and $i-1$, respectively. The termination condition is satisfied when the error function is smaller than a predefined tolerance $\varepsilon=10^{-4}$. Moreover, the computational complexity of Algorithm 1 is $\mathcal{O}(\text{ger}\cdot\text{dim}\cdot N_{\text{pop}})$.

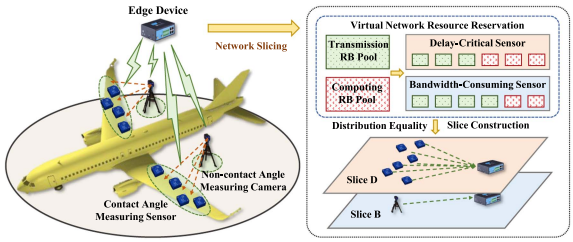


Fig. 4. Resource reservation under associated slices in the FCFST.

V. Simulation Studies

In this section, we assess the performance of the proposed ANS scheme and its DMPSO algorithm by analyzing simulation results in the context of the aileron angle control process in FCSFT. This scenario serves as a typical example of the coexistence of heterogeneous monitoring devices. As shown in Fig. 4, the wing angle is simultaneously monitored by multiple contact angle measuring sensors (inclinometer) and one noncontact angle measuring sensor (camera). Considering the joint constraints of resource availability and slice association, the edge device utilizes association-oriented network slicing to establish virtual resource pools for reserving transmission and computing resources. Specifically, the inclinometers and camera transmit their monitored numerical and image data through the reserved resources to obtain angle information. Subsequently, fusion estimation techniques are employed at the edge devices to reduce angle deviations and improve overall estimation accuracy. The estimated values are used for visualization and updating control decisions.

We consider a network topology covering a rectangular area $[0, 30] \times [0, 100]$. One edge device, one camera, and ten inclinometers (SAT100T-D02, 5 V dc power) are placed within this region. We adopt an empirical industrial wireless channel model consisting of large-scale path loss, shadow fading, and small-scale fading. Let $\mathrm{PL}(d) = \mathrm{PL}_0 + 10n\log_{10}(d) + \kappa$ denote the large-scale fading in dB [31], [36], where PL_0 is the path loss at the reference distance 1 m, and n is the path loss exponent. The shadowing coefficient κ is a zero-mean Gaussian random variable. The small-scale fading distribution between sensors and the edge device is $\mathcal{N}(0, 1)$. The key simulation parameters are listed in Table I, and their values are based on existing experimental results.

A. Performance Evaluations

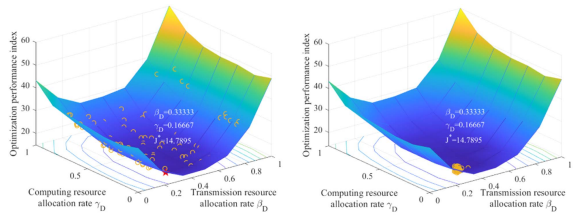
In the simulation, we consider the pitch angle flight control system for a case study. The system model parameters are shown as follows [37]:

$$A = \begin{bmatrix} \frac{-(7\mathcal{K})}{10} - 0.9 & \frac{-(21\mathcal{K})}{50} - 1.18 & -0.68 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{31}$$

where $\mathcal{K}$ is an adjustable parameter that represents different surfaces. In this study, we focus on the ailerons as the surface of interest and set $\mathcal{K} = 1.2$. The control input matrix is denoted as $B = [1, 0, 0]^\top$, and the pitch angle flight control system is utilized to control aileron deflection within the range of $-60°$ to $60°$. For the inclinometer, the simulated voltage change is converted into the inclination angle using a single pendulum's force in the gravitational field. The covariance matrix of the measurement noise related to sensor linearity is set to $\Sigma_{\mathbf{v}_D} = 0.2$ [4]. In the case of noncontact measurement, the covariance matrix of the measurement noise is set to $\Sigma_{\mathbf{v}_B} = 0.04$ [5], in accordance with the inclination calculation principles mentioned earlier.

TABLE I
Main Parameters

Parameters	Values
The number of transmission or computing RBs	30
Bandwidth of each channel B	900 kHz
Capacity of each computing unit C	6×10^8 CPU cycles [31]
Size of each sensor package λ_D	12.1 Kb
Size of a frame of 1920 by 1080 λ_B	1.97 Mb
computing intensity of inclinometer X_D	500 CPU cycles/bit
computing intensity of camera X_B	1000 CPU cycles/bit
Noise power spectral density N_0	-174 dBm/Hz [32]
Path loss exponent in industrial environment κ	4 dB [33]
Path loss at the reference distance PL_0	56.7 dB [34]
Path loss exponent in industrial environment n	4.31 [35]
Maximum transmit power of sensor	10 dBm
Duration of one time slot	10 ms
Number of time slots	200



(a) (b)

Fig. 5. Optimization performance evaluation of DMPSO algorithm. (a) Initial distribution of particle. (b) Final iteration of particle.

To assess the optimization performance of our proposed DMPSO algorithm, we employ the Monte Carlo simulation as a benchmark and analyze the trend of estimation performance variations under different weight factors and slice reservation ratios. As illustrated in Fig. 5, for both scenarios, the optimization performance indicators initially decrease and then increase as the transmission and computing reservation ratios of Slice D (denoted by β_D and γ_D, respectively) rise. This pattern indicates the presence of an optimal value for the interslice transmission and computing RB reservation ratios that minimize the overall estimation error. The optimal value is highlighted with a red star in Fig. 5(a).

To determine the theoretical optimal value, we utilize the DMPSO algorithm to perform particle swarm iterations, yielding optimal weight factors, which are then substituted into (26) and (31), resulting in $\tilde{\beta}_D = 0.3117$ and $\tilde{\gamma}_D = 0.1789$. After quantization, the optimal transmission and computing RB reservation ratios of Slice D are $\beta_D^* = 0.33333$ and $\gamma_D^* = 0.16667$, respectively. These values align closely with the red star obtained experimentally. As shown in Fig. 5(b), the particle swarm converges around the red star, further validating the optimality of our proposed theorem. The convergence of the DMPSO algorithm is effectively demonstrated in Fig. 6(a), where it reaches convergence in about 35 iterations, a rate that is both efficient and satisfactory.

Furthermore, to verify the advantage of using dynamic weight factors in the DMPSO algorithm, we compare it with the random weight factor and average weight factor algorithms. As depicted in Fig. 6(b), although the estimation error variance of all algorithms increases with the number of slices, our proposed algorithm consistently outperforms the others. This superior performance is attributed to the dynamic adjustment of slice

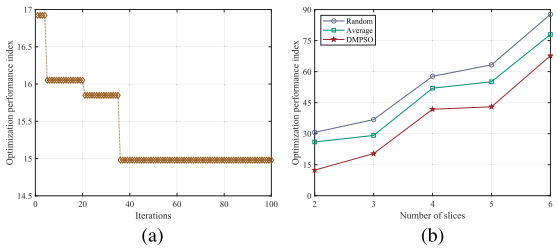


Fig. 6. (a) Convergence performance evaluation of DMPSO algorithm. (b) Performance evaluation with varying numbers of slices.

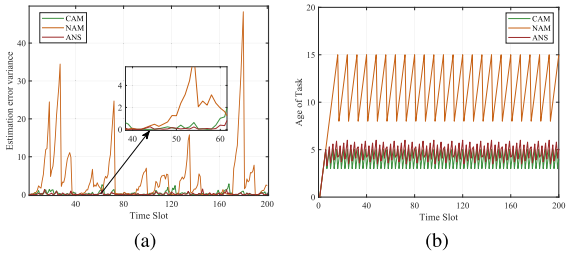


Fig. 7. Real-time performance evaluation with different estimation schemes. (a) Estimation error variance trend. (b) AoT trend.

weights during resource reservation, which maintains the heterogeneous QoS requirements of each slice and meets estimation convergence criteria, even as the number of slices changes.

B. Performance Comparisons Among Different Schemes

The DMPSO algorithm enhances traditional particle swarm optimization to simultaneously achieve optimal solutions for multidimensional variables for the ASN scheme. To emphasize slice association features rather than the technical specifics of the DMPSO algorithm, we refer to our proposed approach as the "ASN scheme" in the following discussion. We then provide a comprehensive evaluation of its performance across various dimensions. First, we assess the effectiveness of joint estimation within the ANS scheme. We compare it against the contact angle measurement (CAM) scheme [4] and the noncontact angle measurement (NAM) scheme [5], both of which rely exclusively on inclinometers and cameras for angle measurement, utilizing all available resources.

The real-time performance evaluation with different schemes is illustrated in Fig. 7. Comparing with Fig. 7(a) and (b), we can see that although the AoT performance of the ANS scheme is between the CAM and NAM schemes, its estimation performance outperforms the remaining two schemes. This demonstrates that the ANS scheme balances the advantages of high-precision observations from the camera and dense sampling of the inclinometer, mitigating the impact of observation noise and temporal effects on the estimation error, as described by (17) and (18). These observations are further supported by the average performance shown in Fig. 8(a) and (b) over 200 time slots.

To accomplish the estimation task, the two slices must allocate RBs effectively while adhering to the nonsharing principle of network resources. The effectiveness of associated reservations in the proposed ANS scheme is also evaluated. Refer to related research [23], [24], [25], [26], the following benchmark algorithms that ignore slice association are considered for performance comparison.

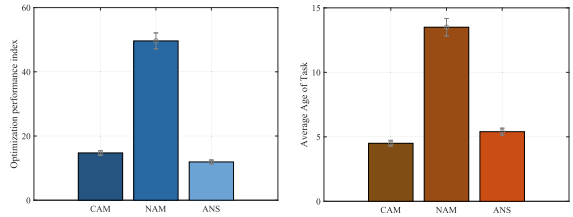


Fig. 8. Average performance evaluation with different estimation schemes. (a) Average estimation error. (b) Average AoT.

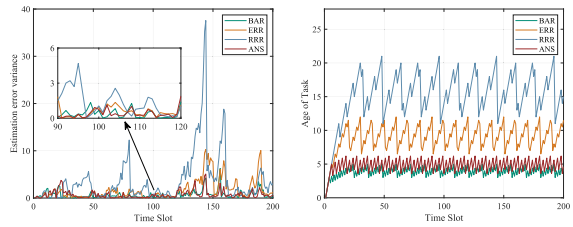


Fig. 9. Real-time performance estimation with different slicing schemes. (a) Estimation error variance trend. (b) AoT trend.

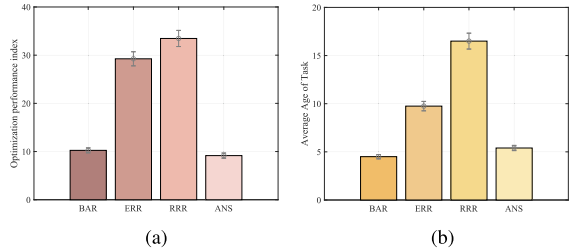


Fig. 10. Average performance evaluation with different slicing schemes. (a) Average estimation error. (b) Average AoT.

1) *Best AoT reservation (BAR) scheme:* The network resources are assigned to the slice with high AoT after meeting the heterogeneous QoS requirements.
2) *Equal resource reservation (ERR) scheme:* The network resources are equally assigned to all slices after meeting the heterogeneous QoS requirements.
3) *Random resource reservation (RRR) scheme:* The network resources are randomly assigned to all slices after meeting the heterogeneous QoS requirements.

Fig. 9 illustrates the real-time performance, including the estimation error variance and AoT, at different time slots with different estimation algorithms. As shown in Fig. 9(b), the proposed ANS scheme significantly outperforms the ERR and RRR algorithms, but performs slightly lower than the BAR solution. This can be attributed to the ANS scheme considering not only the QoS requirements and resource constraints but also the potential interaction among slices to guarantee balanced resource reservation with good estimation performance. Although the BAR algorithm can achieve good cost performance at some time slots in Fig. 9(a), the proposed ANS scheme consistently outperforms the benchmark algorithms with the lowest cost. This performance improvement is attributed to the consideration of interslice association in resource allocation,

which optimally balances the tradeoff between estimation accuracy and timeliness. To further demonstrate the performance of different algorithms, the average performance of the estimation error variance and AoT is also provided, as shown in Fig. 10.

VI. Conclusion

In this article, we introduced a novel association-oriented network slicing architecture to address the coupling effect of different QoS devices on overall performance in joint estimation systems. By analyzing the influence of AoT on slice estimation performance, we demonstrated that the slicing problem, aiming to minimize the system estimation cost, can be optimized using a linear function of the average AoT. To achieve this, we proposed a DMPSO algorithm to determine the optimal estimation weights and reservation decisions. This approach has led to improved system performance and resource utilization, enhancing the efficiency and effectiveness of industrial monitoring, smart manufacturing, and production control processes. Despite these advancements, the current approach primarily considers homogeneous intraslice resources. However, real-world industrial scenarios often involve varying resource demands due to factors such as geographical locations or environmental conditions. To better align with these complexities, future research will focus on optimizing resource orchestration in heterogeneous network environments and addressing the challenges of workload management across multiedge devices.

References

[1] D. A. Chekired, L. Khoukhi, and H. T. Mouftah, "Industrial IoT data scheduling based on hierarchical fog computing: A key for enabling smart factory," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4590–4602, Oct. 2018.

[2] F. Tao, J. Cheng, and Q. Qi, "IIHub: An industrial Internet-of-Things hub toward smart manufacturing based on cyber-physical system," *IEEE Trans. Ind. Informat.*, vol. 14, no. 5, pp. 2271–2280, May 2018.

[3] C. Ren, C. Chen, and X. Wen, Y. Ma, S. Zhu, and X. Guan, "Joint design of communication and computing for digital-twin-enabled aircraft final assembly," *IEEE Internet Things J.*, vol. 10, no. 18, pp. 15872–15886, Sep. 2023.

[4] L. Xie, S. Liu, L. Yang, and Z. Yang, "A novel rudder surface deflection measurement algorithm based on inclination sensor," in *Proc. CSAA/IET Int. Conf. Aircraft Utility Syst.*, 2022, pp. 346–350.

[5] Y. Wang, F. Yang, D. Shan, Q. Fang, and Y. Wei, "Robust vision-based method for wing deflection angle measurement with defocus images," *Measurement*, vol. 189, 2022, Art. no. 110501.

[6] S. Wijethilaka and M. Liyanage, "Survey on network slicing for Internet of Things realization in 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 957–994, Second Quarter 2021.

[7] X. Li et al., "Network slicing for 5G: Challenges and opportunities," *IEEE Internet Comput.*, vol. 21, no. 5, pp. 20–27, Sep. 2017.

[8] W. Shi, J. Cao, and Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

[9] B. Xiang, J. Elias, F. Martignon, and E. Di Nitto, "Joint network slicing and mobile edge computing in 5G networks," in *Proc. IEEE Int. Conf. Commun.*, 2019, pp. 1–7.

[10] S. Kaul, M. Gruteser, and V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. Annu. IEEE Commun. Soc. Conf. Sens., Mesh Ad Hoc Commun. Netw.*, 2011, pp. 350–358.

[11] M. Li, C. Chen, H. Wu, X. Guan, and X. Shen, "Age-of-information aware scheduling for edge-assisted industrial wireless networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5562–5571, Aug. 2021.

[12] I. Kadota et al., "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE ACM Trans. Netw.*, vol. 26, no. 6, pp. 2637–2650, Dec. 2018.

[13] X. Wen, C. Chen, M. Li, L. Lyu, and X. Guan, "Age-of-Task aware sampling rate optimization in edge-assisted industrial network systems," in *Proc. IEEE Glob. Commun. Conf.*, 2021, pp. 1–6.

[14] H. Guo et al., "A fusion estimation of the peak tire-road friction coefficient based on road images and dynamic information," *Mech. Syst. Signal. Process.*, vol. 189, 2023, Art. no. 110029.

[15] N. Senel, K. Kefferpütz, K. Doycheva, and G. Elger, "Multi-sensor data fusion for real-time multi-object tracking," *Processes*, vol. 11, no. 2, p. 501, Feb. 2023.

[16] M. Chen, A. Liu, N. N. Xiong, H. Song, and V. C. Leung, "SGPL: An intelligent game-based secure collaborative communication scheme for metaverse over 5G and beyond networks," *IEEE J. Select. Areas Commun.*, vol. 42, no. 3, pp. 767–782, Mar. 2024.

[17] Y. Li, S. Chen, W. Meng, and C. Li, "Correlation aided joint activity detection and channel estimation for multi-device collaborative massive access," *IEEE Internet Things J.*, vol. 11, no. 10, pp. 18394–18409, May 2024.

[18] J. Chi, T. Qiu, F. Xiao, and X. Zhou, "ATOM: Adaptive task offloading with two-stage hybrid matching in MEC-enabled industrial IoT," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4861–4877, May 2024.

[19] S. Jošilo and G. Dán, "Joint wireless and edge computing resource management with dynamic network slice selection," *IEEE/ACM Trans. Netw.*, vol. 30, no. 4, pp. 1865–1878, Aug. 2022.

[20] L. Ji et al., "Dynamic network slicing orchestration for remote adaptation and conjuration in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 4297–4307, Jun. 2022.

[21] J. Feng, Q. Pei, F. R. Yu, X. Chu, J. Du, and L. Zhu, "Dynamic network slicing and resource allocation in mobile edge computing systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7863–7878, Jul. 2020.

[22] S. Messaoud, A. Bradai, O. B. Ahmed, P. T. A. Quang, M. Atri, and M. S. Hossain, "Deep federated Q-learning-based network slicing for industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5572–5582, Aug. 2021.

[23] C. Zhou et al., "AI-assisted slicing-based resource management for two-tier radio access networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 9, no. 6, pp. 1691–1706, Dec. 2023.

[24] Y. Hou, K. Zhang, X. Liu, G. Chuai, W. Gao, and X. Chen, "An inter-slice RB leasing and association adjustment scheme in O-RAN," *IEEE Trans. Netw. Service Manage.*, vol. 21, no. 1, pp. 402–417, Feb. 2024.

[25] S. Zhang, H. Luo, and J. Li, W. Shi, and X. Shen, "Hierarchical soft slicing to meet multi-dimensional QoS demand in cache-enabled vehicular networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2150–2162, Mar. 2020.

[26] Y. Cui, X. Yang, P. He, D. Wu, and R. Wang, "O-RAN slicing for multi-service resource allocation in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 9272–9283, Jul. 2024.

[27] B. Cao, Z. Li, and X. Liu, Z. Lv, and H. He, "Mobility-aware multiobjective task offloading for vehicular edge computing in digital twin environment," *IEEE J. Select. Areas Commun.*, vol. 41, no. 1, pp. 3046–3055, Oct. 2023.

[28] W. Fan et al., "Joint task offloading and resource allocation for vehicular edge computing based on V2I and V2V modes," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 4277–4292, Apr. 2023.

[29] A. Bozorgchenani et al., "Multi-objective computation sharing in energy and delay constrained mobile edge computing environments," *IEEE Trans. Mobile Comput.*, vol. 20, no. 10, pp. 2992–3005, Oct. 2021.

[30] X. Wen et al., "Age-of-task aware co-design of sampling, scheduling and control for industrial IoT systems," *IEEE Internet Things J.*, vol. 11, no. 3, pp. 4227–4242, Feb. 2024.

[31] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4729–4743, Sep. 2015.

[32] C. Chen, L. Lyu, S. Zhu, and X. Guan, "On-demand transmission for edge-assisted remote control in industrial network systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4842–4854, Jul. 2020.

[33] E. Tanghe et al., "The industrial indoor channel: Large-scale and temporal fading at 900, 2400, and 5200 MHz," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2740–2751, Jul. 2008.

[34] T. Jiang et al., "3GPP standardized 5G channel model for IIoT scenarios: A survey," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 8799–8815, Jun. 2021.

[35] C. Wang, J. Bian, J. Sun, W. Zhang, and M. Zhang, "A survey of 5G channel measurements and models," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3142–3168, Fourth Quarter 2018.

[36] L. Lyu and C. Chen, S. Zhu, and X. Guan, "5G enabled codesign of energy-efficient transmission and estimation for industrial IoT systems," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2690–2704, Jun. 2018.

[37] D. McLean, "Automatic flight control systems," *Meas. Control*, vol. 36, no. 6, pp. 172–175, 2003.