SCIENCE CHINA
Information Sciences

• Supplementary File •

B-CUBE: A Burst-Aware 5G-TSN Architecture and Its Large-Scale Mixed-Flow Scheduling

Yi Xue[1], Jie Jia[1*], Haoyang Mao[1], Jian Chen[1], Wenlong He[1] & Xingwei Wang[1]

[1]*School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China*

Appendix A Network Model of B-CUBE

Appendix A.1 Network Topology and Mixed-Flow Model

Consider a network topology, which consists of the set of industrial devices $\mathcal{D} = \{d_1, d_2, \cdots, d_j, \cdots, d_{N^D}\}$, TSN bridges $\mathcal{B}_{\rm TSN} = \{b_1, b_2, \cdots, b_k, \cdots, b_{N^B}\}$, and 5GS bridges $\mathcal{B}_{\rm 5GS} = \{\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_k, \cdots, \hat{b}_{N^B}\}$, where $N^D = |\mathcal{D}| \in \mathbb{N}^+$ and $N^B = |\mathcal{B}_{\rm TSN}| = |\mathcal{B}_{\rm 5GS}| \in \mathbb{N}^+$ represent the number of industrial devices, TSN bridges, and 5GS bridges, respectively. Further, we define $\mathcal{V} = \mathcal{D} \cup \mathcal{B}_{\rm TSN} \cup \mathcal{B}_{\rm 5GS}$ and $\mathcal{L} \subseteq \mathcal{V} \times \mathcal{V}$, where each element of $\mathcal{L}$ represents a transmission link in the network. Thus, the network topology of B-CUBE can be abstracted as a directed graph $G = (\mathcal{V}, \mathcal{L})$.

Further, we present the mixed-flow model. Let $\mathcal{F} = f_1, f_2, \ldots, f_i, \ldots, f_{N^F}$ denote the set of TT flows, where $N^F = |\mathcal{F}| \in \mathbb{N}^+$. In 5G-TSN, each TT flow f_i periodically generates data frames starting from time slot t_0. Based on a splitting ratio, every data frame is divided into two fragments, $f_i^{\rm TSN}$ and $f_i^{\rm 5GS}$, transmitted via TSN and 5GS bridges, respectively. The periodic TT flows $f_i \in \mathcal{F}$ are featured by source and destination devices, period, frame size, maximum allowed end-to-end delay, injection time offset, and predefined path, given by

$$f_i = \{\rm src, dst, period, size, dd, offset, sr, path\},\ \forall f_i \in \mathcal{F}, \tag{A1}$$

where $f_i.{\rm sr} \in [0, 1]$ denotes the splitting ratio for $f_i^{\rm TSN}$. For instance, if $f_1.{\rm size} = 100{\rm B}$ and $f_1.{\rm sr} = 0.4$, then $f_1^{\rm TSN}$ and $f_1^{\rm 5GS}$ carry 40B and 60B, respectively. Here, $f_i.{\rm sr} = 1$, $f_i.{\rm sr} = 0$, and $0 < f_i.{\rm sr} < 1$ correspond to wired, wireless, and dual-mode transmission, respectively. The path is defined as $f_i.{\rm path} = f_i^{\rm TSN}.{\rm path}, f_i^{\rm 5GS}.{\rm path}$, where each fragment traverses TSN and 5GS bridges of the same nodes. Further, we define the scheduling hyper-period as the least common multiple (LCM) of all TT flows' periods, captures all possible flow generation instances. Thus, only one hyper-period is needed for the flow scheduling. We use $T^{\rm hyper}$ to denote the hyper-period, which can be given by $T^{\rm hyper} = {\rm LCM}(\mathcal{F}.{\rm periods})$. Meanwhile, non-periodic traffic such as unexpected TT flows or safety-critical flows are classified as bursting flows $\bar{\mathcal{F}}$. Bursting flows in B-CUBE are solely routed through the 5G domain, requiring no traffic splitting, and follow a single path of 5GS bridges. Thus, bursting flow $\bar{f}_i \in \bar{\mathcal{F}}$ can be given by

$$\bar{f}_i = \{\rm gt, src, dst, size, dd, path\},\ \forall \bar{f}_i \in \bar{\mathcal{F}}, \tag{A2}$$

where $\bar{f}_i.{\rm gt}$ denotes the generation time of $\bar{f}_i$, which is unknown prior to its occurrence.

Remark 1. In B-CUBE, TSN bridges and 5GS bridges have a one-to-one correspondence at each network node. That is, the TSN bridge b_k at node k is paired with the 5GS bridge $\hat{b}_k$, and they achieve cross-domain time synchronization and resource coordination via the SDN controller. On this basis, TT flows can be transmitted through the wired domain, the 5G domain, or dual-mode according to the splitting ratio $f_i.{\rm sr}$, which is consistent with the three transmission modes of B-CUBE. The fragment $f_i^{\rm 5GS}$ of TT flow f_i is forwarded by 5GS bridges, sharing resources in the 5G domain with bursting flows, and differentiated according to scheduling priorities via the ATM algorithm.

Appendix A.2 Time Slot Configuration

In 5G-TSN, the time slot duration $T_{\rm slot}$ is constrained by both the CQF mechanism and the internal scheduling of the 5GS. For CQF mechanism, all frames in a queue must be forwarded within one slot. Thus, the worst-case forwarding delay of a CQF queue can be given by $T^{\rm CQF} = \frac{Q^{\rm CQF}}{BW} + d^{\rm hop} + \delta$, where $Q^{\rm CQF}$ is the maximum queue capacity. BW is the link bandwidth, $d^{\rm hop}$ is the internal processing and propagation delay of the bridge, and δ is the clock synchronization precision[1)].

* Corresponding author (email: jiajie@mail.neu.edu.cn)

1) δ is the core parameter of cross-domain time synchronization based on IEEE 802.1AS (PTPs), representing the per-hop clock synchronization error (derived from the composition formulas of $T^{\rm CQF}$ and $T^{\rm 5GS}$ in Appendix A.2). In B-CUBE, the wired domain and the 5G domain adopt a unified global δ, and precise cross-domain PTPs synchronization is achieved through the master clock deployed by the SDN controller.

Similarly, each 5GS bridge port pair must complete scheduling within one slot. Thus, the delay of each port pair can be given by $T^{\text{5GS}} = t^{\text{res}}_{\text{DS-TT}} + d^{\text{trans}}_{\text{5GS}} + d^{\text{proc}}_{\text{gNB}} + d^{\text{prop}}_{\text{gNB}} + t^{\text{res}}_{\text{NW-TT}} + \delta$, where $t^{\text{res}}_{\text{DS-TT}}$ and $t^{\text{res}}_{\text{NW-TT}}$ represent the residence time of the TT flow in DS-TT and NW-TT, respectively. $d^{\text{proc}}_{\text{gNB}}$ is the processing delay of the gNB, and $d^{\text{prop}}_{\text{gNB}}$ is the propagation delay between the gNB and the UPF. $d^{\text{trans}}_{\text{5GS}}$ represents the wireless transmission delay inside the 5GS. In order to guarantee the synchronization of cross-domain time slot mapping, T_{slot} should be no less than T^{CQF} and T^{5GS}.

Since the time slot represents the scheduling granularity of the CQF mechanism, all flow periods should be divisible by the predefined time slot size. Meanwhile, considering that T_{slot} also affects the jitter bound of the TT flows, the time slot configuration can be constrained by

$$\begin{cases} (T^{\text{hyper}})\text{mod}(T_{\text{slot}}) = 0, (f_i.\text{period})\text{mod}(T_{\text{slot}}) = 0, \\ T_{\text{slot}} \leqslant \text{GCD}(\mathcal{F}.\text{periods}), \\ T_{\text{slot}} \geqslant \max\left\{T^{\text{CQF}}, T^{\text{5GS}}\right\}. \end{cases} \tag{A3}$$

Then, the set of time slots to be scheduled can be given by $\mathcal{T} = \{t_0, t_1, \cdots, t_{T^{\text{hyper}}/T_{\text{slot}}-1}\}$.

Appendix A.3 5GS Communication Model

Together with the network topology and the mixed-flow model, we further specify the internal communication model of the 5GS bridge. Since the scheduling process within each 5GS bridge is individual and hidden from the external TSN system, it can be abstracted as a multi-user uplink communication model. Each 5GS bridge $\hat{b}_k$ is modeled as a multi-user uplink system comprising a gNB and M^k user equipments (UEs), with $\mathcal{M}^k = \{1, 2, \cdots, m, \cdots, M^k\}$ denoting the UE set. Each UE connects to the DS-TT as an ingress port, while the gNB links to the NW-TT via the 5GC as the egress port. TT flow fragments f_i^{5GS} and bursting flows $\bar{f}_i$ arriving at the bridge are buffered in respective UE-level TT and burst-aware queues, and transmitted over the UE–gNB link. Per-slot transmission capacity is constrained by wireless resource allocation and channel state, with excess fragments being dropped to maintain CQF compliance. Let $Q^k_m(t)$ and $\bar{Q}^k_m(t)$ denote the queue lengths of UE m's TT and burst-aware queues, respectively, in 5GS bridge $\hat{b}_k$ at time slot t. For each 5GS bridge, we consider an uplink OFDMA scenario where each UE occupies an independent channel. The channel gain for UE m in 5GS bridge $\hat{b}_k$ at time slot t is $h^k_m(t) \in \mathbb{C}$, which follows Rayleigh fading and can be modeled as $h^k_m(t) = \sqrt{PL^k_m(t)}\tilde{h}^k_m(t)$, where $PL^k_m(t)$ denotes the distance-dependent path loss of the UE-gNB link, $\tilde{h}^k_m(t) \sim \mathcal{CN}(0, 1)$ denotes the small-scale fading. According to the above definition, the signal of UE m received by gNB at time slot t can be given by $y^k_m(t) = h^k_m(t)\sqrt{p^k_m(t)}s^k_m(t) + \delta^k_m(t)$, where $p^k_m(t)$ and $s^k_m(t)$ denote the transmission power and the transmission signal of UE m, respectively, where $\mathbb{E}[|s^k_m(t)|^2] = 1$. $\delta^k_m(t) \sim \mathcal{CN}(0, \sigma^2)$ represents the additive white Gaussian noise (AWGN) in channel m. Then, the signal-to-noise ratio (SNR) of UE m can be given by $\text{SNR}^k_m(t) = \frac{|h^k_m(t)|^2 p^k_m(t)}{\sigma^2}$. Assume that the total uplink bandwidth of the 5GS bridge is B^{total}Hz, and $B^k_m(t)$ is the bandwidth allocated to channel m, where the bandwidth of each channel is configurable [1]. Then, according to Shannon's theorem, we can obtain the real-time channel capacity of each UE, which can be given by

$$r^k_m(t) = B^k_m(t) \log_2 \left(1 + \text{SNR}^k_m(t)\right). \tag{A4}$$

Appendix B Cross-Domain Mixed-Flow Scheduling

Appendix B.1 Mixed-Flow Queue Arrival States

To address the complexity of cross-domain mixed-flow scheduling in B-CUBE, we first establish clear arrival states for TT and bursting flows to facilitate subsequent problem formulation. Building on the network model in Appendix A, the system incorporates three primary queue types: the TT queue and burst-aware queue at each 5GS bridge's ingress port, and the CQF queue at each bridge's egress port. The corresponding arrival states are defined as follows.

1) TT Flow's Arrival State in TT Queue: Consider a particular 5GS bridge $\hat{b}_k$ at time slot t, we define $U^k_m(f_i, t)$ as the arrival state of flow f_i at the TT queue of UE m. $U^k_m(f_i, t) = 1$ if f_i arrives at the TT queue of UE m, otherwise $U^k_m(f_i, t) = 0$. The condition for $U^k_m(f_i, t) = 1$ can be given by

$$[v^k_m, \hat{b}_k] \in f_i^{\text{5GS}}.\text{path} \ \wedge \ t = \big(f_i.\text{offset} + \alpha f_i.\text{period} + \text{hop}(\hat{b}_k, f_i)\big)\text{mod}\big(T^{\text{hyper}}\big), \tag{B1}$$

where $\forall \alpha \in \left[0, \frac{T^{\text{hyper}}}{f_i.\text{period}} - 1\right]$, v^k_m represents the device or 5GS bridge connected to UE m of $\hat{b}_k$, and $\text{hop}(\hat{b}_k, f_i)$ represents the hop of bridge $\hat{b}_k$ in f_i^{TSN}.path.

2) Bursting Flow's Arrival State in Burst-Aware Queue: Different from TT flows, bursting flows are non-periodic. Meanwhile, affected by wireless scheduling, bursting flows may generate residence time within the routed 5GS bridge (during the residence time, $U^k_m(\bar{f}_i, t) = 1$ holds). Define $\bar{t}^{\text{res}}_{i,j}$ as the residence time of $\bar{f}_i$ at the j-th hop 5GS bridge, the condition for $U^k_m(\bar{f}_i, t) = 1$ can be given by

$$[v^k_m, \hat{b}_k] \in \bar{f}_i.\text{path} \ \wedge \ t \in \left(\left[\sum_{j=1}^{\text{hop}(\hat{b}_k, \bar{f}_i)-1} (\bar{t}^{\text{res}}_{i,j} + 1), \sum_{j=1}^{\text{hop}(\hat{b}_k, \bar{f}_i)-1} (\bar{t}^{\text{res}}_{i,j} + 1) + \bar{t}^{\text{res}}_{i,\text{hop}(\hat{b}_k, \bar{f}_i)}\right]\right)\text{mod}\big(T^{\text{hyper}}\big). \tag{B2}$$

3) TT Flow's Arrival State in CQF Queue: For the sake of illustration, we consider a particular TSN bridge b_k, and the related definitions and formulas also apply to 5GS bridges. Let E^k represent the number of egress ports of bridge b_k, and define $A_k^e(f_i, t)$ as the arrival state of flow f_i at the e-th egress port of bridge b_k at time slot t ($e \in [1, E^k]$). $A_k^e(f_i, t) = 1$ if flow f_i arrives at the e-th egress port of bridge b_k at time slot t, otherwise $A_k^e(f_i, t) = 0$. The condition for $A_k^e(f_i, t) = 1$ can be given by

$$[b_k, b_l(e)] \in f_i^{\text{TSN}}.\text{path} \ \wedge \ t = \big(f_i.\text{offset} + \alpha f_i.\text{period} + \text{hop}(b_k, f_i)\big)\text{mod}\big(T^{\text{hyper}}\big), \tag{B3}$$

where $b_l(e)$ represents the TSN bridge connected to the e-th egress port of bridge b_k.

4) Bursting Flow's Arrival State in CQF Queue of 5GS Bridge: According to (B2) and (B3), The condition for $A_k^e(\bar{f}_i, t) = 1$ can be given by

$$[\hat{b}_k, \hat{b}_l(e)] \in \bar{f}_i.\text{path} \ \wedge \ t = \Big(\sum_{j=1}^{\text{hop}(\hat{b}_k, \bar{f}_i)-1} (\bar{t}_{i,j}^{\text{res}} + 1) + \bar{t}_{i,\text{hop}(\hat{b}_k, \bar{f}_i)}^{\text{res}}\Big)\text{mod}\big(T^{\text{hyper}}\big). \tag{B4}$$

Appendix B.2 Optimization Objective

Considering that online scheduling in the 5G domain ultimately determines the scheduling status of mixed flows, and the need to ensure higher priority for bursting flows, we define the optimization objective as maximizing the AoI-weighted schedulable flow number in 5G domain. Consider a uplink burst-aware queue $\bar{Q}_m^k(t)$ for a particular 5GS bridge $\hat{b}_k$. Let $\alpha_m^k(t)$ denote its AoI at time slot t, defined as the number of time slots elapsed since the most recent bursting data frame sent from queue $\bar{Q}_m^k(t)$ to the gNB until time slot t. Let $g_m^k(t)$ denote the transmission time of the most recent bursting data frame from queue Q received at the gNB during time slot t. Then, $\alpha_m^k(t)$ can be computed as $\alpha_m^k(t) = t - g_m^k(t)$. In practice, $\alpha_m^k(t)$ exhibits a zigzag-like function which increases linearly with a slope of 1 and drops to 1 whenever queue $\bar{Q}_m^k(t)$ transmits a new data frame. Further, define $\phi_m^k(t) \in \{0, 1\}$ as the transmission decision for queue $\bar{Q}_m^k(t)$, determined by the real-time wireless scheduling strategy. $\phi_m^k(t) = 1$ indicates that queue $\bar{Q}_m^k(t)$ transmits bursting data frames to the gNB at time slot t, otherwise $\phi_m^k(t) = 0$. Then, the update process for $\alpha_m^k(t)$ can be given by

$$\alpha_m^k(t) = a_m^k(t-1) \times (1 - \phi_m^k(t-1)) + 1, \ t \in \mathcal{T} - \{t_0\}. \tag{B5}$$

where $\alpha_m^k(t_0) = 1$.Further, let $\bar{\mathcal{F}}_m^k(t) = \{\bar{f}_i | U_m^k(\bar{f}_i, t) = 1, \bar{f}_i \in \bar{\mathcal{F}}\}$ and $\mathcal{F}_m^k(t) = \{f_i | U_m^k(f_i, t) = 1, f_i \in \mathcal{F}\}$ denote the sets of flows in the burst-aware queue $\bar{Q}_m^k(t)$ and TT queue $Q_m^k(t)$ at time slot t, respectively, the optimization objective can be given by

$$\max \sum_{t\in\mathcal{T}} \sum_{\hat{b}_k \in \mathcal{B}_{5\text{GS}}} \sum_{m \in \mathcal{M}^k} \Big(\sum_{S_i \in \mathcal{S}_m^k(t)} S_i + \alpha_m^k(t) \times \sum_{\bar{S}_i \in \bar{\mathcal{S}}_m^k(t)} \bar{S}_i \Big). \tag{B6}$$

where $S_i \in \{1, 0\}$ and $\bar{S}_i \in \{1, 0\}$ denote the scheduling state of TT flow f_i and bursting flow $\bar{f}_i$, respectively. $S_i = 1$ if f_i is successfully scheduled, otherwise $S_i = 0$. The same applies to $\bar{S}_i$. $\mathcal{S}_m^k(t) = \{S_i | f_i \in \mathcal{F}_m^k(t)\}$, and $\bar{\mathcal{S}}_m^k(t) = \{\bar{S}_i | \bar{f}_i \in \bar{\mathcal{F}}_m^k(t)\}$.

Appendix B.3 Core Constraints

To clarify the problem formulation, we further provide the definition of core constraints for cross-domain mixed-flow scheduling. We first focus on the offset of TT flows, which should be bounded by its period and deadline, as delayed injection timing may cause exceeding the deadline. Thus, the offset constraint can be given by

$$0 \leqslant f_i.\text{offset} \leqslant \text{offset}_i^{\max}, \forall i \in [1, N^F], \tag{B7}$$

where $\text{offset}_i^{\max} = \min\{\frac{f_i.\text{period}}{T_{\text{slot}}}, \frac{f_i.\text{dd}}{T_{\text{slot}}} - \text{hop}_i\} - 1$, hop_i denotes the number of hops in the path of f_i.

Further, considering the egress port of each TSN bridge, where the number of schedulable frames is constrained by the CQF queue resource Q^{CQF}. According to (B3), under a given f_i.sr, the CQF queue resource constraints for TSN bridges can be given by

$$\sum_{i=1}^{N^F} \Big(f_i.\text{sr} \times f_i.\text{size} \times A_k^e(f_i, t)\Big) \leqslant Q^{\text{CQF}}, \forall k \in [1, N^B], \forall e \in [1, E^k], \forall t \in \mathcal{T}. \tag{B8}$$

Regarding 5GS bridges, the CQF queue must account for the simultaneous arrival of TT and bursting flows. Consequently, its queue resource constraint can be given by

$$\begin{aligned} &\sum_{i=1}^{N^F} \Big(S_i \times (1 - f_i.\text{sr}) \times f_i.\text{size} \times A_k^e(f_i, t)\Big) + \sum_{\bar{f}_i \in \bar{\mathcal{F}}} \Big(\bar{S}_i \times \bar{f}_i.\text{size} \times A_k^e(\bar{f}_i, t)\Big) \leqslant Q^{\text{CQF}}, \\ &\forall k \in [1, N^B], \forall e \in [1, \hat{E}^k], \forall t \in \mathcal{T}. \end{aligned} \tag{B9}$$

Moreover, the residence time $\bar{t}_{i,j}^{\text{res}}$ of a bursting flow is constrained by its deadline to ensure timely forwarding. For a specific bursting flow $\bar{f}_i$ arriving at time slot t, its residence times at preceding nodes have been observed. Thus, the residence time constraint at the currently arrived 5GS bridge $\hat{b}_k$ can be given by

$$0 \leqslant \bar{t}_{i,\text{hop}(\hat{b}_k, \bar{f}_i)}^{\text{res}} \leqslant \bar{f}_i.\text{dd} - \sum_{j=1}^{\text{hop}(\hat{b}_k, \bar{f}_i)-1} (\bar{t}_{i,j}^{\text{res}} + 1), \ \forall \bar{f}_i \in \bar{\mathcal{F}}. \tag{B10}$$

Finally, wireless resource constraints must be considered to ensure correctness of online scheduling. Consider a particular 5GS bridge $\hat{b}_k$ at time slot t, the queue state $Q_m^k(t)$ and $\bar{Q}_m^k(t)$ is necessary for online wireless scheduling, which can be given by

$$\begin{cases} Q_m^k(t) = \sum_{f_i \in \mathcal{F}_m^k(t)} \big((1 - f_i.\mathrm{sr}) \times f_i.\mathrm{size}\big), \\ \bar{Q}_m^k(t) = \sum_{\bar{f}_i \in \bar{\mathcal{F}}_m^k(t)} \bar{f}_i.\mathrm{size}. \end{cases} \tag{B11}$$

For each 5GS bridge, the wireless scheduling should guarantee that the occupied bandwidth does not exceed the total bandwidth B^{total} while maximally ensuring synchronous transmission of mixed flows within a single time slot. Therefore, the wireless resource constraints can be given by

$$\sum_{m \in \mathcal{M}^k} B_m^k(t) \leqslant B^{\mathrm{total}}, \tag{B12}$$

and

$$B_m^k(t) \geqslant \frac{\sum_{f_i \in \mathcal{F}_m^k(t)} \big((1 - f_i.\mathrm{sr}) \times f_i.\mathrm{size}\big) \times S_i + \sum_{\bar{f}_i \in \bar{\mathcal{F}}_m^k(t)} \bar{f}_i.\mathrm{size} \times \bar{S}_i}{\log_2\big(1 + \mathrm{SNR}_m^k(t)\big) \times d_{\mathrm{5GS}}^{\mathrm{trans}}}, \forall k \in [1, N^B], \forall t \in \mathcal{T}. \tag{B13}$$

Appendix B.4 Problem Formulation

Recalling the architecture and network model of B-CUBE, the wireless scheduling process within each 5GS bridge is independent and *black-boxed* to the external TSN system. Meanwhile, according to (B13), wireless resource constraints are related to $\mathcal{F}$.offset and $\mathcal{F}$.sr. Therefore, there is a dependency between wireless scheduling and the solution of wired scheduling. Based on the above analysis, we formulate the cross-domain mixed-flow scheduling problem of B-CUBE as a bilevel optimization problem [2], where the lower-level problem maximizes the throughput of the wired domain by optimizing the offsets and splitting ratio of TT flows, which effectively relieves the resource requirements of the 5G domain (constraint (B13)) that contributes to the schedulability. Based on the obtained wired scheduling strategy, the upper-level problem optimizes the online wireless scheduling strategy within each 5GS bridge to maximize the AoI-weighted schedulable flow number. Thus, the problem can be formulated as[2)]

$$\begin{aligned} \text{(P0)} \quad & \max_{\mathcal{B}^k(t), \mathcal{F}.\mathrm{offset}, \mathcal{F}.\mathrm{sr}, \mathcal{S}, \bar{\mathcal{S}}} \sum_{t \in \mathcal{T}} \sum_{\hat{b}_k \in \mathcal{B}_{\mathrm{5GS}}} \sum_{m \in \mathcal{M}^k} \Big(\sum_{S_i \in \mathcal{S}_m^k(t)} S_i + \alpha_m^k(t) \times \sum_{\bar{S}_i \in \bar{\mathcal{S}}_m^k(t)} \bar{S}_i \Big) && \text{(B14)} \\ & \text{s.t. } \{\mathcal{F}.\mathrm{offset},\ \mathcal{F}.\mathrm{sr}\} \in \mathop{\arg\max}_{\mathcal{F}.\mathrm{offset},\ \mathcal{F}.\mathrm{sr}} \big\{O(\mathcal{F}) | (B7), (B8), 0 \leqslant f_i.\mathrm{sr} \leqslant 1,\ \forall f_i \in \mathcal{F}\big\}, \\ & \quad S_i \in \{0, 1\},\ \forall i \in [1, N^F], && \text{(B14a)} \\ & \quad \bar{S}_i \in \{0, 1\},\ \forall \bar{f}_i \in \bar{\mathcal{F}}, && \text{(B14b)} \\ & \quad (B12), (B13), \end{aligned}$$

where $\mathcal{B}^k(t) = \{B_m^k(t) | m \in \mathcal{M}^k\}$ $(\forall k \in [1, N^B], \forall t \in \mathcal{T})$, $\mathcal{S} = \{S_i | i \in [1, N^F]\}$, $\bar{\mathcal{S}} = \{\bar{S}_i | \forall \bar{f}_i \in \bar{\mathcal{F}}\}$, and

$$O(\mathcal{F}) = \sum_{i=1}^{N^F} \big(f_i.\mathrm{period} \times f_i.\mathrm{size} \times f_i.\mathrm{sr}\big). \tag{B15}$$

It can be observed that both the lower/upper-level problems of (P0) are MINLP with deeply coupled variables, making the problem NP-hard and difficult to find a high-quality solution with polynomial time complexity in large-scale scheduling scenarios. To cope with this, and considering the *black box* of wireless scheduling and the fact that no linking variables are included in problem (P0) (i.e., the lower-level problem does not depend on the solution of the upper-level problem [3]), we decouple it into two sub-problems, wired and wireless scheduling, and sequentially solve them with reduced complexity while guaranteeing the reasonableness of the scheduling scheme. According to the definition of problem (P0), the wired scheduling problem can be formulated as

$$\begin{aligned} \text{(P1)} \quad & \max_{\mathcal{F}.\mathrm{offset}, \mathcal{F}.\mathrm{sr}} \quad O(\mathcal{F}) && \text{(B16)} \\ & \text{s.t.} \quad 0 \leqslant f_i.\mathrm{sr} \leqslant 1,\ \forall f_i \in \mathcal{F}, && \text{(B16a)} \\ & \qquad (B7), (B8). \end{aligned}$$

Further, each 5GS bridge should optimize its internal wireless resource allocation online. Consider a particular 5GS bridge $\hat{b}_k$ at time slot t, the wireless scheduling problem can be formulated as

$$\begin{aligned} \text{(P2)} \quad & \max_{\mathcal{B}^k(t), \mathcal{S}_m^k(t), \bar{\mathcal{S}}_m^k(t)} \sum_{m \in \mathcal{M}^k} \Big(\sum_{S_i \in \mathcal{S}_m^k(t)} S_i + \alpha_m^k(t) \times \sum_{\bar{S}_i \in \bar{\mathcal{S}}_m^k(t)} \bar{S}_i \Big) && \text{(B17)} \\ & \text{s.t.} \quad (B14a), (B14b), (B12), (B13). \end{aligned}$$

2) Since the variables involved in constraint (B9) and (B10) have to be observed after wireless scheduling, these constraints are not considered in problem (P0). Meanwhile, we consider this constraint in step 7 of **Algorithm C3** to guarantee the rationality of the whole scheduling scheme.

Appendix C Algorithm Design

In this section, we first conduct a comprehensive analysis of the scale of problems (P1) and (P2), ensuring a thorough understanding of the issues at hand. We then propose the problem reformulation schemes to reduce the complexity. Further, we propose a solution framework for B-CUBE cross-domain mixed-flow scheduling, as shown in **Algorithm C1**.

Algorithm C1 Solution framework for cross-domain mixed-flow scheduling in B-CUBE

Require: Network and flow parameters;
Ensure: Wired scheduling strategy $\{\mathcal{F}.\text{offset}, \mathcal{F}.\text{sr}\}$, wireless scheduling strategy $\mathcal{B}^k(t)$, and flow scheduling state $\mathcal{S}$;
1: Execute the JOSO algorithm (**Algorithm C2**) to obtain the wired scheduling strategy $\{\mathcal{F}.\text{offset}, \mathcal{F}.\text{sr}\}$;
2: **for** t **in** $\mathcal{T}$ **do**
3: **for** $\hat{b}_k$ **in** $\mathcal{B}_{5\text{GS}}$ **do**
4: Generate $\mathcal{F}_m^k(t)$ and $\bar{\mathcal{F}}_m^k(t)$ based on $\{\mathcal{F}.\text{offset}, \mathcal{F}.\text{sr}\}$ and the 5G domain scheduling results from previous slot;
5: Execute the online ATM algorithm (**Algorithm C3**) to obtain the real-time wireless scheduling strategy $\mathcal{B}^k(t)$;
6: **end for**
7: **end for**
8: Generate the final scheduling result $\{\mathcal{S}, \bar{\mathcal{S}}\}$ for the current hyper-period based on $\{\mathcal{F}.\text{offset}, \mathcal{F}.\text{sr}, \mathcal{B}^k(t)\}$;

The JOSO algorithm works in SDN controller, and solves wired scheduling strategy based on TT flow features and network topology provided by CUC and CNC. The ATM algorithm works in the 5GC of each 5GS bridge and performs online optimization of wireless scheduling strategy based on the results of the JOSO algorithm. According to the scheduling strategy of all 5GS bridges in the current slot, real-time flow scheduling status can be obtained, which will be used as the environment variable of the next slot ATM algorithm until the end of the hyper-period. Finally, the solution framework will obtain the cross-domain mixed-flow scheduling strategy. The specific solutions are detailed in the following subsections.

Appendix C.1 Joint Offset and Splitting Optimization

According to the problem formulation, problem (P1) includes both non-linear and integer constraints, where the search space for offsets is $\prod_{i=1}^{N^F} f_i.\text{period}$. Therefore, it is a MINLP problem, rendering it NP-hard [4]. To reduce complexity while pursuing a near-optimal solution, we introduce a l_0-norm-based reformulation that transforms (P1) into an equivalent non-linear programming (NLP) problem, which is then solved by a combined SL0 and MM optimization scheme. Since the scheduling strategies of offsets and splitting ratio are mutually influential, we adopt a combined rather than decoupled reformulation scheme by designing a unified variable $\boldsymbol{x}$ to represent the joint scheduling strategy of the two. Define $\boldsymbol{x} \in \mathbb{R}^{N\times 1}$, where $N = \sum_{i=1}^{N^F}(offset_i^{\max}+1)$. The dimension of $\boldsymbol{x}$ represents all feasible offset scheduling strategies, and each element value x_n of $\boldsymbol{x}$ represents the splitting rule on the corresponding offset setting. Define $\text{lower}_i = \sum_{j=0}^{i-1}(\text{offset}_j^{\max}+1)+1$ and $\text{upper}_i = \sum_{j=0}^{i}(\text{offset}_j^{\max}+1)$ (set $\text{offset}_0^{\max} = -1$ for validity) for each TT flow f_i, and then f_i.offset and f_i.sr can be determined from a subset of $\boldsymbol{x}$, which can be given by $\boldsymbol{x}_i = \{x_n | x_n \in \boldsymbol{x}, n \in [\text{lower}_i, \text{upper}_i]\}$. To guarantee the uniqueness of solutions, we introduce a l_0-norm constraint for the elements of x_i, which can be given by

$$\sum_{x_n \in \boldsymbol{x}_i} ||x_n||_0 \leqslant 1, \ \forall i \in [1, N^F]. \tag{C1}$$

According to (C1), there is at most one non-zero element x_{n^*} in $\boldsymbol{x}_i$, then f_i.offset and f_i.sr can be given by

$$f_i.\text{offset} = n^* - \text{lower}_i, \quad f_i.\text{sr} = x_{n^*}. \tag{C2}$$

In particular, when $\boldsymbol{x}_i$ is a zero vector, set the corresponding $f_i.\text{offset} = 0$ and $f_i.\text{sr} = 0$ to avoid ambiguities. Here is an example for intuitive understanding: Assume that the parameter settings of two TT flows f_1 and f_2 are shown in Table C1, and according to Appendix B.3, their upper bounds of offsets are $\text{offset}_1^{\max} = 3$ and $\text{offset}_2^{\max} = 2$, respectively. At this time, $N = 7$, $\boldsymbol{x} \in \mathbb{R}^{7\times 1}$, $\text{lower}_1 = 1$, $\text{upper}_1 = 4$, $\text{lower}_2 = 5$, and $\text{upper}_2 = 7$. Therefore, the offset and splitting strategies of f_1 and f_2 can be determined by the subsets $\boldsymbol{x}_1 = \{x_1, x_2, x_3, x_4\}$, and $\boldsymbol{x}_2 = \{x_5, x_6, x_7\}$ of $\boldsymbol{x}$, respectively. When we adopt $\boldsymbol{x}$ from (C3) as the wired scheduling scheme, according to (C2), we can obtain the unique offset and splitting ratio strategy as (C4).

Table C1 Prameter settings of TT flows

id	Deadline	Period	Hops
f_1	$6 * T_{\text{slot}}$	$7 * T_{\text{slot}}$	3
f_2	$5 * T_{\text{slot}}$	$6 * T_{\text{slot}}$	3

$$\begin{cases} \boldsymbol{x} &= [0,\ 0.2,\ 0,\ 0,\ 0,\ 0,\ 0.3], \\ \boldsymbol{x}_1 &= [0,\ 0.2,\ 0,\ 0], \\ \boldsymbol{x}_2 &= \qquad\qquad\quad [0,\ 0,\ 0.3]. \end{cases} \tag{C3}$$

$$\begin{cases} f_1.\text{offset} = 1, & f_1.\text{sr} = 0.2, \\ f_2.\text{offset} = 2, & f_2.\text{sr} = 0.3. \end{cases} \tag{C4}$$

In summary, vector $\boldsymbol{x}$ contains the offsets and splitting ratio for all TT flows. Further, to characterize the real-time traffic condition of each CQF queue, we define the coefficient matrix $\mathbf{C} \in \mathbb{R}^{M\times N}$, where $M = T^{\text{hyper}} \times \sum_{k=1}^{N^B} E^k$. Define $\boldsymbol{c}_n$ ($n \in [1, N]$) as the column vector of $\mathbf{C}$, which represents the queue resource occupation status when f_i.offset $= n - \text{lower}_i$ ($n \in [\text{lower}_i, \text{upper}_i]$). Each element $c_{m,n}$ in $\boldsymbol{c}_n$ can be determined by the corresponding flow arrival state $A_k^e(f_i, t)$. Specifically, when f_i.offset $= n - \text{lower}_i$, $A_k^e(f_i, t)$ can be obtained from (B3), and further $c_{m,n}$ can be given by

$$c_{m,n} = f_i.\text{size} \times A_k^e(f_i, t),\ \forall n \in [\text{lower}_i, \text{upper}_i], \tag{C5}$$

where $m = \sum_{k'=1}^{N^B} \left[E^{k'} \times (t-1)\right] + \sum_{k'=1}^{k-1} \left(E^{k'}\right) + e$. According to (C5), $c_{m,n} = f_i$.size if flow f_i arrives at the e-th egress port of bridge b_k at time slot t (i.e., $A_k^e(f_i, t) = 1$), otherwise $c_{m,n} = 0$. Let $\boldsymbol{z} = \mathbf{C} * \boldsymbol{x}$, which represents the traffic status for all CQF queues at each time slot. Based on the above definition and analysis, problem (P1) can be reformulated as

$$\begin{aligned}
\text{(P1-1)}\quad & \max_{\boldsymbol{x}} \ \sum_{m=1}^{M} z_m && \text{(C6)}\\
& \text{s.t.}\quad 0 \leqslant x_n \leqslant 1,\ \forall x_n \in \boldsymbol{x}, && \text{(C6a)}\\
& \qquad\ \ 0 \leqslant z_m \leqslant Q^{\text{CQF}},\ \forall m \in [1, M], && \text{(C6b)}\\
& \qquad\ \ (C1),
\end{aligned}$$

where $\sum_{m=1}^{M} z_m = \sum_{i=1}^{N^F} \left(f_i.\text{period} \times f_i.\text{size} \times f_i.\text{sr}\right)$ ($z_m \in \boldsymbol{z}$), constraint (C6a) is equivalent to the constraint (B14a), constraint (C6b) is equivalent to the constraint (B8), and the constraint (B7) is guaranteed by the dimension of $\boldsymbol{x}$. Thus, (P1-1) is an equivalent problem of (P1). Note that the constraint (C1) is a non-smooth non-convex constraint. Thus we approximate the l_0-norm with the SL0 mechanism [5], which can be given by $g_\delta(x_n) \triangleq 1 - \exp\left(-\frac{x_n}{\delta}\right)$ ($\forall x_n \in \boldsymbol{x}$), where $\delta > 0$ is a parameter that controls the smoothness of the approximation. Then, constraint (C1) can be relaxed as

$$\sum_{x_n \in \boldsymbol{x}_i} g_\delta(x_n) \leqslant 1,\ \forall i \in [1, N^F]. \tag{C7}$$

Since $g_\delta(x_n)$ is a concave function, inequality (C7) is a non-convex constraint. Thus, we apply MM to deal with it. Consider a particular iteration $l \geqslant 1$, the convex upper bound of $g_\delta(x_n)$ can be obtained via a first-order Taylor expansion, which can be given by $\hat{g}_\delta(x_n) \triangleq g_\delta\left(x_n^{(l-1)}\right) + \frac{1}{\delta}\exp\left(-\frac{x_n^{(l-1)}}{\delta}\right)\left(x_n - x_n^{(l-1)}\right)$. Therefore, inequality (C7) can be reformulated as

$$\sum_{x_n \in \boldsymbol{x}_i} \hat{g}_\delta(x_n) \leqslant 1,\ \forall i \in [1, N^F]. \tag{C8}$$

Based on the above derivations, the problem (P1-1) at the l-th iteration can be reformulated as

$$\begin{aligned}
\text{(P1-2)}\quad & \max_{\boldsymbol{x}} \ \sum_{m=1}^{M} z_m && \text{(C9)}\\
& \text{s.t.}\quad (C6a), (C6b), (C8),
\end{aligned}$$

which is convex and can be solved by CVX [6]. With these optimizations, our proposed JOSO algorithm is summarized in **Algorithm C2**. Since the objective value is non-decreasing after each iteration and has a non-negative upper bound, **Algorithm C2** can converge to a stationary point of problem (P1). According to [7], **Algorithm C2** is guaranteed to converge with a worst-case complexity of $\mathcal{O}(T_{\max}\sqrt{M + N + N^F}(M + 2N + N^F)N^2)$.

Algorithm C2 JOSO algorithm

Require: Network and flow parameters, δ, η_1, η_2, and converge threshold th;
Ensure: Wired scheduling strategy $\{\mathcal{F}.\text{offset}, \mathcal{F}.\text{sr}\}$;

```
Initial random feasible solution x^(0), and calculate matrix C by (C5);
Set l = 1, T_max = 50, and flag = 0;
while l ⩽ T_max AND flag == 0 do
    Solve the problem (P1-2) and obtain x^(l);
    Calculate Δ = | Σ_{m=1}^{M} (z_m^(l) − z_m^(l−1)) |;
    if Δ ⩽ th then
        flag = 1;
    end if
    l = l + 1;
end while
According to x^(l), calculate F.offset and F.sr by (C2);
return F.offset, F.sr;
```

Appendix C.2 AoI-weighted Transmittable Flow Ratio Maximization

The wired scheduling strategy obtained by JOSO relieves the resource requirements of the 5G domain (constraint (B13)) while determining the TT flow arrival state information of the 5GS bridges (Eq. (B1)). Based on this, we further optimize the real-time wireless scheduling problem (P2) to maximize the AoI-weighted schedulable flow number. The key to this problem lies in the handling of the binary scheduling state S. In order to solve it efficiently, we map it to a continuous transmittable flow ratio to eliminate the integer constraints and obtain an equivalent convex form. Specifically, define $a_m^k(t) \in [0,1]$ and $\bar{a}_m^k(t) \in [0,1]$ as the transmittable flow ratio of UE m's TT and burst-aware queue, respectively. Define $e_m^k(t) = \frac{Q_m^k(t)}{\log_2\left(1+\mathrm{SNR}_m^k(t)\right)\times d_{\mathrm{5GS}}^{\mathrm{trans}}}$ and $\bar{e}_m^k(t) = \frac{\bar{Q}_m^k(t)}{\log_2\left(1+\mathrm{SNR}_m^k(t)\right)\times d_{\mathrm{5GS}}^{\mathrm{trans}}}$, constraint (B13) can be rewritten as

$$B_m^k(t) \geqslant e_m^k(t) \times a_m^k(t) + \bar{e}_m^k(t) \times \bar{a}_m^k(t), \tag{C10}$$

Then, problem (P2) can be reformulated as

$$\begin{aligned}
\text{(P2-1)} \quad & \max_{\mathcal{B}^k(t),\mathcal{A}^k(t),\bar{\mathcal{A}}^k(t)} \sum_{m\in\mathcal{M}^k} \left(\lfloor N_{m,t}^k \times a_m^k(t)\rfloor + \lfloor \alpha_m^k(t) \times \bar{N}_{m,t}^k \times \bar{a}_m^k(t)\rfloor\right) && \text{(C11)}\\
\text{s.t.} \quad & 0 \leqslant a_m^k(t) \leqslant 1,\ \forall m \in \mathcal{M}^k, && \text{(C11a)}\\
& 0 \leqslant \bar{a}_m^k(t) \leqslant 1,\ \forall m \in \mathcal{M}^k, && \text{(C11b)}\\
& (B12), (C10),
\end{aligned}$$

where $\mathcal{A}^k(t) = \{a_m^k(t) | m \in \mathcal{M}^k\}$, $\bar{\mathcal{A}}^k(t) = \{\bar{a}_m^k(t) | m \in \mathcal{M}^k\}$, $N_{m,t}^k = |\mathcal{F}_m^k(t)|$, and $\bar{N}_{m,t}^k = |\bar{\mathcal{F}}_m^k(t)|$. According to the description of problem (P2) in Appendix B.4, the optimization objective of problem (P2-1) is equivalent to that of problem (P2). Meanwhile, the integer constraint (B14a) is considered in the rounding process of the objective function, thus problem (P2-1) is equivalent to problem (P2). Further, we relax the rounding function to obtain a convex form of this problem. For the sake of simplicity in subsequent formulas, we define $\boldsymbol{a} = [a_1^k(t), \cdots, a_{M^k}^k(t), \bar{a}_1^k(t), \cdots, \bar{a}_{M^k}^k(t)]^T$, $\boldsymbol{n} = [N_1^k(t), \cdots, N_{M^k}^k(t), \alpha_1^k(t) \times \bar{N}_1^k(t), \cdots, \alpha_{M^k}^k(t) \times \bar{N}_{M^k}^k(t)]^T$, and $\boldsymbol{e}_{m,t}^k = [e_1^k(t), \cdots, e_{M^k}^k(t), \bar{e}_1^k(t), \cdots, \bar{e}_{M^k}^k(t)]^T$, then problem (P2-1) can be formulated as

$$\begin{aligned}
\text{(P2-2)} \quad & \max_{\mathcal{B}^k(t),\boldsymbol{a}} \ \boldsymbol{n}^T\boldsymbol{a} && \text{(C12)}\\
\text{s.t.} \quad & 0 \leqslant a_m \leqslant 1,\ \forall a_m \in \boldsymbol{a}, && \text{(C12a)}\\
& B_m^k(t) \geqslant e_m \times a_m + e_{m+M^k} \times a_{m+M^k},\ \forall a_m \in \boldsymbol{a}, \forall e_m \in \boldsymbol{e}, m \in [1, M^k], && \text{(C12b)}\\
& (B12).
\end{aligned}$$

Remark 2. *Rationality of AoI for Adapting to Bursting Flows.* Bursting flows are aperiodic and safety-critical flows (e.g., equipment fault alarms). Their core requirement is information freshness rather than merely *meeting the deadline*. If a bursting flow (such as a fault signal) arrives late, even without exceeding the deadline, industrial accidents may still occur due to outdated information. AoI can continuously quantify the *freshness* of information from generation to reception and accurately distinguish the scheduling priorities of bursting flows. Meanwhile, jitter constraint is an indicator for the transmission stability of periodic TT flows. Since bursting flows have no fixed period, jitter has no practical physical meaning and cannot be used to evaluate their transmission performance. In contrast, AoI can effectively characterize the transmission timeliness of aperiodic flows. Therefore, AoI is well suited to meet the key requirements of bursting flows.

Remark 3. *Relationship Between AoI and Delay Constraint.* In B-CUBE, AoI is used as the optimization weight, while delay acts as a hard constraint. In the residence time constraint in Appendix B.3 (Eq. (B10)), strict delay constraints have been set for bursting flows to ensure that their end-to-end delay does not exceed industrial requirements. AoI is incorporated into the scheduling optimization objective (Eq. (B6)) as a weighting factor, which prioritizes the scheduling of bursting flows with higher AoI (more outdated information). This achieves *maximizing the information freshness of bursting flows on the premise of satisfying hard delay constraints.*

It can be observed that problem (P2-2) is convex and satisfies Slater's condition. Thus, the KKT conditions are sufficient and necessary to obtain the optimal solution [8]. The Lagrangian function of problem (P2-2) can be given by

$$\begin{aligned}
\mathcal{L} = & \sum_{m=1}^{2M^k} n_m \times a_m + \beta_m (1 - a_m) + \lambda_m a_m + \xi_m \left(B_m^k(t) - e_m \times a_m - e_{m'} \times a_{m'}\right)\\
& + \rho\left(B^{\mathrm{total}} - \textstyle\sum_{m\in\mathcal{M}^k} B_m^k(t)\right),
\end{aligned} \tag{C13}$$

where β_m, λ_m, ξ_m, and ρ are Lagrange multipliers, $m' = (m + M^k) \mathrm{mod} (2M^k)$. Further, the KKT conditions for the problem (P2-2) can be given by

$$\frac{\partial\mathcal{L}}{\partial a_m} = n_m - \beta_m + \lambda_m - \xi_m e_m = 0,\ \frac{\partial\mathcal{L}}{\partial B_m^k(t)} = \xi_m - \rho = 0, \tag{C14a}$$

$$\beta_m (1 - a_m) = 0,\ \lambda_m a_m = 0, \tag{C14b}$$

$$\xi_m \left(B_m^k(t) - e_m \times a_m - e_{m'} \times a_{m'} \right) = 0,\ \rho \left(B^{\text{total}} - \sum\nolimits_{m \in \mathcal{M}^k} B_m^k(t) \right) = 0, \tag{C14c}$$

$$\beta_m \geqslant 0,\ \lambda_m \geqslant 0,\ \xi_m \geqslant 0,\ \rho \geqslant 0,\ (C11a),\ (B12),\ (C10). \tag{C14d}$$

By solving the above KKT conditions, the optimal solution of the problem (P2-2) can be obtained as

$$a_m^* = \begin{cases} 1 & ,\ \rho < \frac{n_m}{e_m}, \\ \frac{B^{\text{total}} - \sum\limits_{\tilde{m} \in \tilde{\mathcal{M}}^k} e_{\tilde{m}}}{e_m} & ,\ \rho = \frac{n_m}{e_m}, \\ 0 & ,\ \rho > \frac{n_m}{e_m}, \end{cases} \tag{C15}$$

and

$$B_m^{k*}(t) = e_m \times a_m^* + e_{m'} \times a_{m'}^*, \tag{C16}$$

where $\tilde{\mathcal{M}}^k = \{m \mid \rho < \frac{n_m}{e_m},\ \forall m \in [1, 2M^k]\}$, and

$$\rho = \begin{cases} 0 & ,\ \sum_{m=1}^{2M^k} e_m \leqslant B^{\text{total}}, \\ \frac{n_{m_j}}{e_{m_j}} & ,\ \sum\limits_{i=1}^{j-1} e_{m_i} < B^{\text{total}} \leqslant \sum\limits_{i=1}^{j} e_{m_i}, \end{cases} \tag{C17}$$

where $m_j \in \hat{\mathcal{M}}^k$ $(j \in [1, 2M^k])$, $\hat{\mathcal{M}}^k$ is the set consisting of all of indexes in descending order of $\frac{n_m}{e_m}$.

Algorithm C3 ATM algorithm

Require: Wired scheduling strategy and system parameters of 5GS bridge b_k^L at time slot t;
Ensure: Optimal wireless scheduling strategy $\mathcal{B}^k(t)$ and flow scheduling state $\mathcal{S}$;
1: Initialize $\mathcal{B}^k(t) = \phi$;
2: **for** m **in** $\mathcal{M}^k$ **do**
3: Generate vector $\boldsymbol{a}$, $\boldsymbol{n}$, and $\boldsymbol{e}$;
4: Calculate a_m^* and $B_m^{k*}(t)$ by (C15) and (C16);
5: $\mathcal{B}^k(t) = \mathcal{B}^k(t) \cup \{B_m^{k*}(t)\}$;
6: **end for**
7: Update $\mathcal{S}$ based on $\boldsymbol{a}^*$ under the constraint of (B9) and (B10);
8: return $\{\mathcal{B}^k(t), \mathcal{S}\}$;

Remark 4. By analyzing the structure of the optimal solution, it can be seen that the difference between the objective function values of problems (P2-1) and (P2-2) corresponding to (C15) and (C16) is less than 1. Therefore, there is no other better solution to problem (P2-1), i.e., (C15) and (C16) are the optimal solutions of problem (P2-1). Since problems (P2-1) and (P2) are equivalent, (C15) and (C16) are also the optimal solutions to problem (P2).

With these optimizations, our proposed ATM algorithm is summarized in **Algorithm C3**. The time complexity of the ATM algorithm is approximated by $\mathcal{O}(M^k)$.

Remark 5. The proposed scheme decouples the cross-domain mixed-flow joint scheduling into two sub-problems: wired scheduling and wireless scheduling. The solution of the wired scheduling serves as the input to the wireless scheduling, with no bidirectional coupling. The search space of traditional global joint scheduling algorithms spans the joint space of wired and wireless domains, involving highly complex variable coupling. In contrast, the search space of the proposed scheme is the superposition of two subspaces, thus significantly reducing the solution complexity. Furthermore, in terms of algorithms, the JOSO algorithm transforms the NP-hard MINLP problem (P1) into a convex optimization problem (P1-2) by combining the SL0-MM method, achieving a polynomial-complexity solution. It provides an essential complexity advantage over exponential-complexity SMT [9] algorithms and quasi-polynomial complexity heuristic algorithms such as Tabu [10].In the experiments in Appendix D, we also quantitatively analyze the time cost and scheduling result quality of different algorithms, and the experimental conclusions are consistent with the above analysis. Meanwhile, the ATM algorithm eliminates integer constraints by mapping binary states to continuous transmissible flow ratios. As a result, it derives the optimal solution analytically, combined with the KKT conditions, enabling efficient online scheduling with linear complexity in the 5G domain. Therefore, the proposed scheme can effectively reduce the computational complexity of large-scale mixed-flow scheduling.

Appendix D Numerical Results

This section evaluates the effectiveness of the proposed architecture and algorithms using three common industrial topologies (ring, linear, and mixed) to assess scalability. Table D1 summarizes the number of bridges and devices. Key parameters include: $Q^{\text{CQF}} = 10 \times$ MTU per CQF queue; $B^{\text{total}} = 500$MHz, $\sigma^2 = -80$dB, and $P_{\max} = 30$dBm per 5GS bridge. The gNB-UE distance d is uniformly distributed in $[80\text{m}, 100\text{m}]$, where $C_0 = -30$dB denotes the path loss at a reference distance of 1m and $\hat{\alpha} = 3.5$ denotes the path loss factor. We set the sub-carrier spacing (SCS) to 120kHz, and each slot contains 14 OFDM symbols and one cyclic prefix. Based on Q^{CQF} and SCS, $T_{\text{slot}} = 125\mu s$. In the absence of open-source 5G-TSN flow sets, TT flows are generated following [11], with f_i.size $\in [100\text{B}, 1000\text{B}]$. Three period-and-deadline (PD) configurations (Table D2) test robustness. Meanwhile, bursting flows are randomly generated per slot, with their source and destination devices randomly distributed among the industrial devices in the network.The size of a bursting flow is set as $\bar{f}_i$.size $\in$ [50B, 500B], and its deadline is set as $\bar{f}_i$.dd $\in$ {0.5ms, 1ms}.The bursting bandwidth (50–100 bits/μs) is defined as the total bandwidth demand of all bursting flows within the hyperperiod, which is used to limit the number and size of bursting flows. All experiments run on an i7-9750H CPU with 64 GB RAM.

Table D1 Network Topology

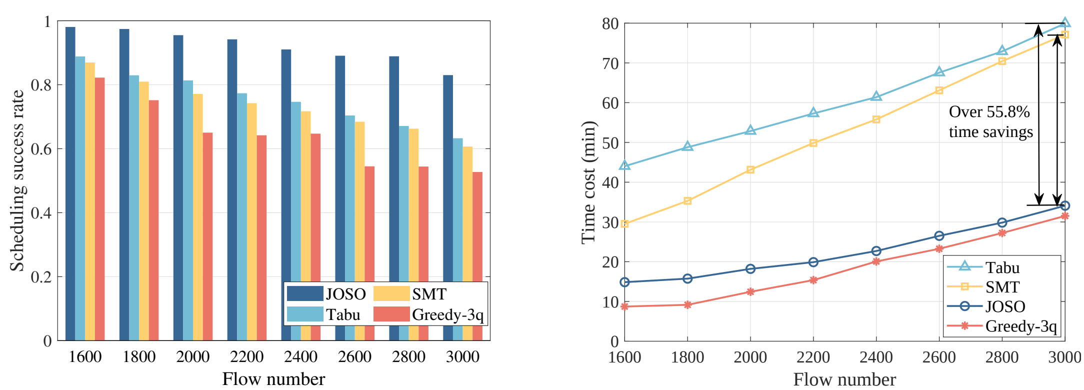
Topology	TSN bridge	5GS bridge	Devices per bridge
Ring	5	5	{1, 2, 3}
Linear	5	5	{1, 2, 3}
Mix	8	8	{1, 2, 3}

Table D2 PD Schemes of TT Flow

PD scheme	Period (ms)	Deadline (ms)
PD1	{2, 4, 8}	{1, 2, 4}
PD2	{1, 2, 4}	{1, 2, 4}
PD3	{2, 8, 10}	{2, 8, 10}

Appendix D.1 Performance Analysis of JOSO Algorithm

Figure D1 illustrates the performance comparison between the JOSO and the Tabu [10], SMT [9], and Greedy-3q [12] algorithms. Here, the SMT returns the satisfiable results of the problem computed by the Z3-solver tool for Python within an acceptable time set by 90 minutes. And Greedy-3q always chooses the sending time interval with the maximal capacity along the path. The exchanging mode is used in Tabu-ITP, where the size of tabu list is set to 500, and the iterations along with repetitions are 10,000 and 200, respectively. In this section, we adopt the ring topology with the PD1 scheme, the bursting bandwidth is set to 100 bits/us. Figure D1(a) investigates the TT flow scheduling success rate of each scheme. It can be observed that JOSO achieves a success rate exceeding 90% when scheduling fewer than 2400 TT flows. Even in a large-scale scenario with 3000 TT flows, it maintains an 83.05% success rate, outperforming Tabu, SMT, and Greedy-3q by 19.78%, 22.39%, and 30.34%, respectively. Figure D1(b) plots the time costs of different algorithms. Due to the large search space, Tabu and SMT incur the highest time overhead, reaching 79.92 min and 77.13 min respectively when scheduling 3000 flows, while JOSO significantly reduces time consumption to 34.09 min, saving over 55.8%. Compared to Greedy-3q, it was only 2.5 min slower but significantly improved the scheduling success rate[3)].



(a) Scheduling success rate. (b) Time cost.

Figure D1 Performance comparison of JOSO with Tabu, SMT, and Greedy-3q.

Analysis reveals that while Tabu and SMT are widely adopted for 5G-TSN flow scheduling, they struggle to achieve desirable results despite substantial computational effort when confronting the large and complex solution space of large-scale mixed-flow scheduling. Greedy-3q consistently selects the current time slot with maximum capacity but struggles to predict all subsequent flow aggregation changes during forwarding. Conversely, by introducing convex optimization theories like SL0 and MM, JOSO ensures high success rates while reducing time consumption, demonstrating significant advantages in large-scale mixed-flow scheduling.

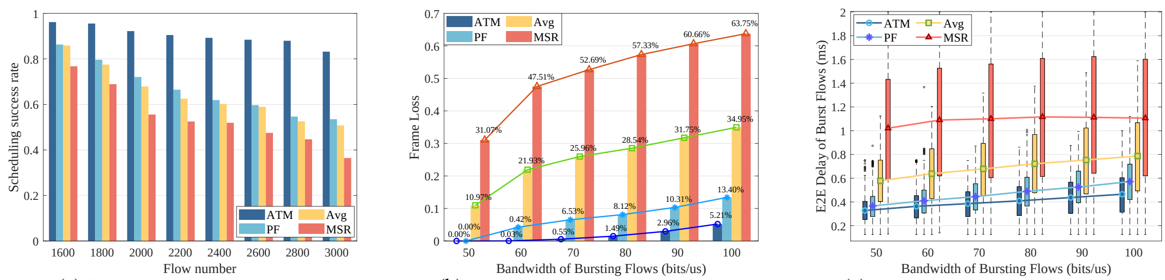
Appendix D.2 Performance Analysis of ATM Algorithm

Figure D2 plots the performance comparison of the ATM algorithm with three non-AoI benchmarks: proportional fairness (PF), maximize sum rate (MSR), and average (Avg). Here, the PF scheme allocates available bandwidth to each UE in proportion to its required bandwidth. The MSR scheme optimizes the bandwidth allocation by maximizing the sum rate. The Avg scheme allocates equal bandwidth to all UEs. In this section, we adopt the ring topology with the PD1 scheme. Figure D2(a) investigates the TT flow scheduling success rates across different schemes. It can be observed that ATM achieves an average scheduling success rate of 90.39%, which improves by 23.62%, 25.85%, and 36.1% compared to PF, MSR, and Avg, respectively. Additionally, Figure D2(b) plots the frame loss of bursting flows for different schemes. It can

3) Note that TT flows are periodic, offline hyper-period scheduling is performed once and reused [10].

be observed that ATM demonstrates excellent reliability across different bursting bandwidths. At a bursting bandwidth of 100 bits/us, its frame loss is only 5.21%, which reduces by 8.19%, 29.74%, and 58.54% compared to PF, MSR, and Avg, respectively. Finally, Figure D2(c) examines the delay of bursting flows for different schemes. It can be observed that ATM effectively controls the delay of bursting flows. At a bursting bandwidth of 100 bits/us, ATM maintains the delay boundary within 1 ms and achieves an average delay of 0.46 ms, which reduces by 19.30%, 41.78%, and 58.56% compared to PF, MSR, and Avg, respectively.

Analysis reveals that MSR and Avg fail to match actual conditions due to neglecting real-time service characteristics. While PF ensures fair resource allocation, it neglects strategies that could enhance network capacity. The ATM algorithm, on the other hand, significantly reduces frame loss and end-to-end delay for bursting flows while maintaining the schedulable number of TT flows. This is achieved through AoI-awareness of bursting flows and a reasonable problem transformation approach.



(a) Scheduling success rate. (b) Frame loss of bursting flows. (c) E2E delay of bursting flows.

Figure D2 Performance comparison of ATM with PF, MSR, and Avg.

Appendix E Conclusion

This paper proposes B-CUBE, a burst-aware 5G-TSN architecture for large-scale mixed-flow scheduling, which achieves wired, wireless, and dual-mode transmission by co-deploying TSN and 5GS bridges on the same node. Through the traffic splitting mechanism for TT flows and fixed wireless transmission for bursting flows, B-CUBE overcomes the flexibility limitations of planar architectures while ensuring real-time burst-awareness-and-transmission. Meanwhile, B-CUBE remains compliant with 3GPP standards, laying the foundation for its compatibility and scalability. To address cross-domain mixed-flow scheduling in B-CUBE, we propose the JOSO and ATM algorithms to solve wired and wireless scheduling, respectively. JOSO employs SL0 approximation and MM methods to transform the original problem into an approximate convex form, reducing complexity while ensuring high throughput. The ATM algorithm employs the AoI mechanism to guarantee high priority of bursting flows. It maps binary states to continuous ratios to derive an equivalent convex form, then obtains optimal resource allocation by solving KKT conditions. Experimental evaluations across ring, linear, and mix topologies confirmed B-CUBE's superiority: B-CUBE and the proposed algorithms outperform benchmarks in large-scale mixed-flow while reducing computational complexity.

References

1 Li W, Lv T, Cao Y, et al. Multi-carrier NOMA-empowered wireless federated learning with optimal power and bandwidth allocation. IEEE Trans. Wireless Commun., 2023, pages 1–1
2 Sinha A, Malo P, Deb K. A review on bilevel optimization: From classical to evolutionary approaches and applications. IEEE Trans. Evol. Comput., 2018, 22: 276–295
3 Kleinert T, Labbé M, Ljubić I, et al. A survey on mixed-integer programming techniques in bilevel optimization. EURO J. Comput. Optim., 2021, 9: 100007
4 Zhou R, Palomar D P. Solving high-order portfolios via successive convex approximation algorithms. IEEE Trans. Signal Process., 2021, 69: 892–904
5 Tao M, Chen E, Zhou H, et al. Content-centric sparse multicast beamforming for cache-enabled cloud RAN. IEEE Trans. Wireless Commun., 2016, 15: 6118–6131
6 Grant M, Boyd S P. CVX: MATLAB software for disciplined convex programming. URL http://cvxr.com/cvx
7 Chen J, Xie Y, Mu X, et al. Energy efficient resource allocation for IRS assisted CoMP systems. IEEE Trans. Wireless Commun., 2022, 21: 5688–5702
8 Boyd S, Boyd S P, Vandenberghe L. Convex optimization. Cambridge university press, 2004
9 Xu L, Xu Q, Zhang Y, et al. Co-design approach of scheduling and routing in time sensitive networking. In: Proceedings of Proc. IEEE Conf. Ind. Cyberphys. Syst. (ICPS), 2020. 111-116
10 Yan J, Quan W, Jiang X, et al. Injection time planning: Making CQF practical in time-sensitive networking. In: Proceedings of Proc. IEEE INFOCOM Conf. Comput. Commun, 2020. 616-625
11 IEC/IEEE 60802 TSN Profile for Industrial Automation. Standard IEEE 802.1, URL https://1.ieee%20802.org/tsn/iec-ieee-60802
12 Krolikowski J, Martin S, Medagliani P, et al. Joint routing and scheduling for large-scale deterministic IP networks. Comput. Commun., 2021, 165: 33–42