mathematics

MDPI

Article

Capacity Costs of Probability Ambiguity and Support Rectangularization in Edge–Cloud Scheduling

Mikail Mohammed Salim [1,*] and Farheen Naaz [2]

[1] School of Computer Science and Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea
[2] Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea; farheenaaz@yu.ac.kr
* Correspondence: mikail@yu.ac.kr

Abstract

Edge–cloud task scheduling reserves uplink capacity before the wireless service state is known. The reserve depends on both the probability law over a prescribed service-state library and the geometry representing that library. This study separates and quantifies these two effects in a fixed-workload finite-state model. We eliminate post-state bandwidth–CPU recourse exactly, derive a quadratic capacity representation in inverse-square-root spectral efficiencies, and compare empirical (EMP), Wasserstein, hard-support (HARD), and rectangular (BOX) uncertainty models under schedule reoptimization. The analysis yields an exact support-rectangularization premium, a resource-weighted joint-extreme deficit with a relative-cost factor below two, a sandwich bound on the optimized premium, and a finite hardening radius for chance-constrained capacity. Across 60 instances, mean savings from retaining joint support are 6.45–9.32% relative to rectangular protection. Across 20 instances, mean relative increases are 7.39–10.82% for EMP-to-HARD, 8.80–9.26% for support rectangularization, and 16.98–20.89% for EMP-to-BOX; all are deterministic optimization bounds. Positive lower bounds persist for both changed-target workloads, all nine nested-support cases, and both probability-side variations, while intermediate Wasserstein capacity changes under the alternative transport metric. Probability protection and support geometry impose distinct, quantifiable capacity costs.

Keywords: edge–cloud computing; capacity planning; uncertainty geometry; support rectangularization; distributionally robust optimization; Wasserstein ambiguity; resource allocation

MSC: 90C17; 90C15; 68M20

Check for updates

Academic Editor: Ye Tian

Received: 28 September 2026
Revised: 18 October 2026
Accepted: 19 October 2026
Published: 20 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Edge–cloud computing couples task placement with communication and computation resources across end devices, edge servers, and remote cloud infrastructure. Computation offloading and radio–computation allocation are therefore naturally linked planning decisions [1–4]. Their interaction is explicit in formulations that jointly allocate wireless and computational resources under latency constraints [5,6]. When the wireless service state is revealed only after the task schedule is chosen, capacity planning must account for both the discrete placement decision and the continuous recourse available after state realization.

The required reserve also depends on how uncertainty is represented. With a prescribed set of service states, finite observations create *probability ambiguity* over that support;

Wasserstein distributionally robust optimization controls this uncertainty around the empirical law [7]. A separate modeling choice concerns *support geometry*. Coordinatewise robustification replaces jointly attainable service vectors with marginal extremes, thereby discarding dependence information. The resulting conservatism depends on the geometry of the uncertainty set and on the level of protection imposed [8,9]. Probability ambiguity changes how mass is assigned to the prescribed states, whereas support rectangularization changes the set of service vectors against which capacity is reserved.

This paper quantifies the two effects under a common fixed-workload scheduling problem. A task schedule is selected before the wireless state is known; uplink bandwidth and remote CPU rates adapt after the state is observed. The analysis determines the capacity added as probability protection is strengthened over the prescribed service-state support and the additional capacity added when that support is replaced by its coordinatewise bounding box. The discrete schedule reoptimizes under each uncertainty model, and the same framework characterizes the ambiguity radius at which Wasserstein reliability reaches hard protection over every state in the prescribed library.

For a fixed schedule x and service vector $\boldsymbol{\eta}$, let $u_i = 1/\sqrt{\eta_i}$. Eliminating the continuous bandwidth–CPU recourse gives

$$B_{\mathrm{req}}(x;\mathbf{u}) = \mathbf{u}^{\mathsf{T}} Q_x \mathbf{u}, \tag{1}$$

where Q_x is determined by the task–destination assignments, task sizes and deadlines, destination CPU capacities, and remaining CPU headroom. Equation (1) places the empirical (EMP), Wasserstein distributionally robust (DRO), hard prescribed-support (HARD), and rectangular (BOX) uncertainty models on the same schedule-dependent capacity map.

The contributions are threefold:

- *Exact recourse and capacity geometry:* The continuous bandwidth–CPU recourse problem is solved in closed form for any fixed schedule and realized state. The resulting quadratic representation yields exact state-sensitivity identities, a CPU–spectrum exchange relation, and a supermodular destination-load function with increasing marginal communication cost as CPU headroom is consumed.
- *Capacity cost of support rectangularization:* For a fixed schedule, an exact Q_x-weighted support-rectangularization premium is linked to a dimensionless resource-weighted joint-extreme deficit. The resulting relative-capacity bounds differ by an explicit factor $c_x < 2$. After schedule reoptimization, the optimized HARD-to-BOX premium is enclosed by the schedule-level premiums of the two model-specific optimizers, with equality characterized by cross-optimality.
- *Distributionally robust capacity and hardening:* The Wasserstein chance-constrained capacity is shown to be nondecreasing and finite-step in the ambiguity radius. A droppable-state representation separates admissible failure sets from their schedule-dependent resource cost, and an event-level transport threshold ϵ_{hard} characterizes the terminal HARD regime. This yields a common EMP–DRO–HARD–BOX hierarchy and a decomposition of the total capacity premium into probability-robustification and support-rectangularization components.

Numerically, the optimized support-rectangularization premium has a positive lower bound in all 60 instances of the three-regime study, and most numerically resolved HARD–BOX pairs retain the same selected assignment. In the separate 20-instance frontier, both probability-side robustification and support rectangularization contribute to the total reserve, while $\epsilon = 0.040$ lies beyond the hardening threshold for every instance. The numerical study also varies workload construction, transport metric, state law, failure tolerance, and support cardinality within the same capacity-planning formulation.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 defines the edge–cloud system and uncertainty models. Section 4 derives the exact recourse and quadratic capacity geometry. Section 5 develops the rectangularization and Wasserstein capacity results. Section 6 presents the numerical evaluation. Section 7 discusses the implications and scope, and Section 8 concludes the paper. Appendix A gives the numerical formulation and implementation details.

2. Related Work

Related work spans uncertainty-aware offloading, robust edge-resource planning, and distributionally robust network optimization. The comparison below holds workload and scheduling structure fixed while varying the representation of the same finite wireless uncertainty.

2.1. Uncertainty-Aware Task Offloading and Resource Allocation

Wireless uncertainty has been incorporated into offloading through channel-estimation error, failures, and fluctuating network conditions. Nguyen et al. [10] jointly optimize offloading and radio/computation resources under perfect and imperfect CSI. Qu et al. [11] minimize aggregate latency under bounded offloading failures, while Ji et al. [12] minimize worst-case expected device energy under channel and queue uncertainty. Processing uncertainty produces a complementary resource constraint: Li et al. [13] impose probabilistic deadline guarantees using moments of task-processing cycles, and Fan et al. [14] jointly allocate offloading and communication–computation resources under moment information and computation-outage constraints.

Several formulations combine uncertainty across system layers. Hao et al. [15] incorporate deviations between digital-twin estimates and physical states into robust offloading and resource allocation. Deng et al. [16] consider uncertain processing cycles with intermittent communication. Nan et al. [17] use min–max optimization under imperfect computing-capacity information, and Xu et al. [18] address multifactorial uncertainty in distributed offloading. These studies optimize under prescribed uncertainty models. Here, probability and support representations are compared within one fixed-workload capacity problem.

Bandyopadhyay et al. [19] study delay-sensitive edge–cloud offloading and bandwidth utilization with discretized differential evolution. Their experiments evaluate scheduling performance under a different objective. In the present formulation, an analytically verified feasible schedule gives a capacity upper bound, while the Wasserstein chance constraint limits worst-case failure probability over the prescribed ambiguity set. Certifying the optimized capacity and its premium additionally requires valid optimization lower bounds; empirical heuristic performance alone supplies neither guarantee.

2.2. Robust Edge Service Placement and Capacity Planning

At a longer planning timescale, reservation and placement decisions are made before demand or infrastructure conditions are realized. Li et al. [20] jointly reserve radio and virtual-machine resources under uncertain mobile-user demand. Nguyen et al. [21] formulate two-stage adaptive robust service placement, resource sizing, and workload allocation. Qu et al. [22] study edge-server placement under uncertain failures, and Cheng et al. [23] consider simultaneous demand and node-failure uncertainty in resilient edge provisioning.

The structure retained from the uncertainty process also affects planning decisions. Jeong et al. [24] use polyhedral uncertainty sets with configurable protection for application placement and traffic allocation. Cheng et al. [25] retain spatio-temporal demand correlation in a multi-period adaptive robust placement model. Bai et al. [26] combine strategic node

location and capacity configuration with a data-driven distributional model for long-term demand and a budget-based set for short-term fluctuations. These formulations establish the relevance of uncertainty structure to advance resource planning; here, the uncertain object is a finite, correlated wireless service-state support and the planning metric is minimum shared uplink capacity at fixed workload.

In a different planning setting, Yang et al. [27] use forecast-derived renewable-generation uncertainty sets for two-stage robust microgrid scheduling. Their objective is operational scheduling cost, whereas the present analysis quantifies the minimum shared uplink reserve induced by probability and support representations of wireless service states.

2.3. Distributionally Robust Edge and Network Resource Optimization

At the methodological level, robust optimization protects decisions against prescribed uncertainty sets and makes the resulting performance–conservatism tradeoff explicit [8,28], while correlated-set formulations show that retaining dependence can materially change robust solutions [9]. Wasserstein distributionally robust optimization supplies data-driven ambiguity sets around empirical distributions [7], and Wasserstein chance-constraint reformulations are established in general form [29,30]. Building on this uncertainty-set and duality machinery, the analysis derives capacity characterizations after exact edge–cloud resource recourse is eliminated.

Distributionally robust chance constraints also appear directly in network optimization. Ling et al. [31] apply them to computation offloading under imperfect CSI in a wireless body-area network, jointly controlling communication and computation variables to minimize latency. Tsai et al. [32] use a Wasserstein ambiguity set for routing in integrated satellite–terrestrial networks with uncertain latency distributions. The formulation here places EMP, positive-radius Wasserstein DRO, hard protection over the prescribed service-state support, and its coordinatewise BOX enlargement on one fixed-workload capacity map. The resulting structural analysis characterizes the support-rectangularization premium, its schedule-dependent geometry under reoptimization, and the finite-state hardening structure.

Table 1 summarizes representative studies by decision problem, uncertainty source, treatment, and planning criterion.

Table 1. Representative studies of uncertainty-aware offloading and robust edge–cloud resource planning.

Study	Decision Problem	Uncertainty Source	Uncertainty Treatment	Primary Criterion/ Planning Role
Nguyen et al. [10], TSC 2021	MIMO MEC offloading and radio/computation allocation	CSI estimation	Perfect/imperfect-CSI optimization	Offloading/resource efficiency under CSI error
Qu et al. [11], TMC 2022	Robust task-offloading scheduling	Offloading failures	Min–max robust optimization	Aggregate task latency
Ji et al. [12], TWC 2022	Time-critical computation offloading	Channel and queue uncertainty	Worst-case expected optimization with tail bounds	Device energy under intrinsic uncertainty
Ling et al. [31], TCOM 2021	Backscatter-assisted computation offloading	Imperfect CSI	Distributionally robust chance constraints	End-to-end latency

Table 1. *Cont.*

Study	Decision Problem	Uncertainty Source	Uncertainty Treatment	Primary Criterion/ Planning Role
Li et al. [20], TMM 2018	Radio/VM reservation and allocation	User resource demand	Robust resource reserva-tion/allocation	Advance provisioning cost and utilization
Nguyen et al. [21], IoTJ 2022	Service placement, resource sizing, workload allocation	Service demand	Two-stage adaptive robust optimization	Resource procurement and placement cost
Qu et al. [22], TMC 2023	Edge-server placement and workload service	Server failures	Robust max–min/submodular optimization	Serviceable workload under failures
Cheng et al. [23], TNSM 2024	Service placement, resource procurement, workload allocation	Demand and node failures	Two-stage adaptive robust optimization	Resilient edge provisioning
Jeong et al. [24], Omega 2024	Application placement and traffic allocation	Network latency	Polyhedral robust uncertainty sets	Placement cost and latency violations
Cheng et al. [25], TSC 2025	Dynamic service placement and workload allocation	Spatio-temporally correlated demand	Multi-period adaptive robust optimization	Resource reservation with correlated demand
Nan et al. [17], TMC 2025	Task offloading and local/edge resource allocation	Computing-capacity information	Min–max robust optimization	Robust latency/resource allocation
Bai et al. [26], EJOR 2026	Node location, capacity configuration, service allocation	Long- and short-term demand	Distributional plus budget-based robust optimization	Strategic capacity planning with recourse

The analysis uses established convex allocation, robust-set ordering, Wasserstein transport duality, and optimizer comparisons as mathematical foundations. Its capacity-specific contributions are the exact quadratic representation obtained after eliminating bandwidth–CPU recourse, the support-rectangularization premium and resource-weighted joint-extreme bounds derived from that representation, and their optimized-capacity sandwich with cross-optimality conditions. For probability ambiguity, the finite-state failure-set representation connects admissible state exclusions to schedule capacity and yields an explicit event-level hardening radius. Together, these results characterize the capacity costs of probability ambiguity and support enlargement under a common workload target and exact post-state recourse.

3. System Model and Problem Formulation

The model separates a discrete task schedule chosen before the wireless service state is observed from continuous bandwidth and destination-CPU recourse after state observation. The same task, destination, and recourse structure is used under each uncertainty description.

3.1. Edge–Cloud System and Task Model

Consider a set of indivisible tasks $\mathcal{I} = \{1, \ldots, K\}$. Task i is characterized by

$$(d_i, c_i, D_i, w_i), \tag{2}$$

where $d_i > 0$ is the input size in bits, $c_i > 0$ is the required number of CPU cycles, $D_i > 0$ is the completion deadline, and $w_i > 0$ is the admitted-workload weight. Each task is assigned to exactly one of local execution, one of the remote destinations, or rejection. The remote-destination set is

$$\mathcal{J} = \{1, 2, C\}, \tag{3}$$

where 1 and 2 denote the two edge servers and C denotes the cloud tier.

Let $x_i^{\mathrm{L}} \in \{0, 1\}$ indicate local execution, $x_{ij} \in \{0, 1\}$ indicate offloading to destination $j \in \mathcal{J}$, and $r_i \in \{0, 1\}$ indicate rejection. The assignment constraint is

$$x_i^{\mathrm{L}} + \sum_{j \in \mathcal{J}} x_{ij} + r_i = 1, \qquad i \in \mathcal{I}. \tag{4}$$

If task i is executed locally with device CPU rate $f_i^{\mathrm{L}} > 0$, local execution is admissible only when

$$\frac{c_i}{f_i^{\mathrm{L}}} \leq D_i. \tag{5}$$

Local execution consumes neither the shared uplink capacity nor remote CPU capacity.

Each remote destination j has CPU capacity $F_j > 0$ and a fixed non-radio latency τ_j. For task i, define the residual latency available to uplink transmission and remote execution as

$$L_{ij} = D_i - \tau_j. \tag{6}$$

An offloading choice with $L_{ij} \leq 0$ is excluded, and its coefficients are set to zero. The term τ_j represents the destination-specific non-radio latency used in the model.

The wireless service condition is represented by a finite state $s \in \mathcal{S} = \{1, \ldots, S\}$. In state s, user i has spectral efficiency $\eta_i^s > 0$. If b_{ij}^s hertz of uplink bandwidth is assigned to an offloaded task, its transmission rate is modeled as $b_{ij}^s \eta_i^s$. Thus, for $x_{ij} = 1$, successful completion in state s requires

$$\frac{d_i}{b_{ij}^s \eta_i^s} + \frac{c_i}{f_{ij}^s} \leq L_{ij}, \qquad i \in \mathcal{I},\ j \in \mathcal{J},\ s \in \mathcal{S}, \tag{7}$$

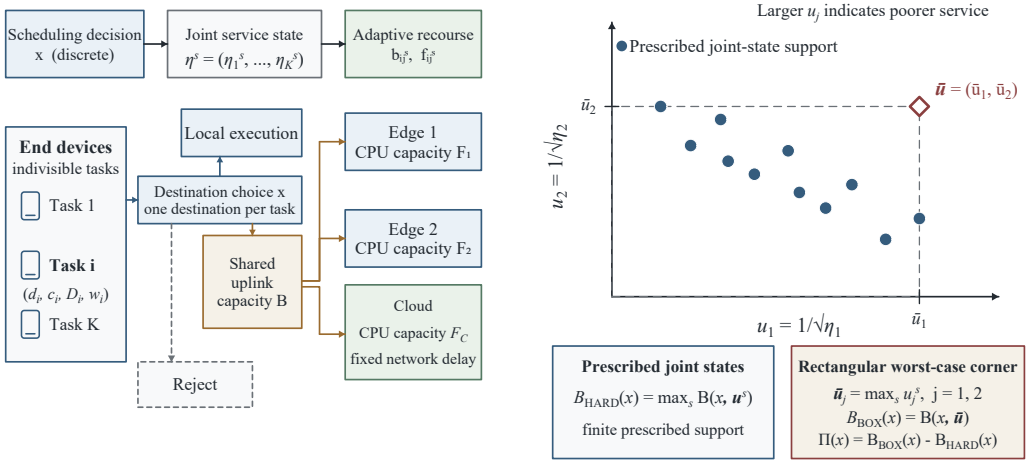
where f_{ij}^s is the destination CPU rate allocated to task i. In any state required to be successful, the resource allocations satisfy

$$\sum_{i \in \mathcal{I}} x_{ij} f_{ij}^s \leq F_j, \qquad j \in \mathcal{J},\ s \in \mathcal{S}, \tag{8}$$

$$\sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} x_{ij} b_{ij}^s \leq B, \qquad s \in \mathcal{S}, \tag{9}$$

with strictly positive b_{ij}^s and f_{ij}^s for active offloading assignments. These constraints apply in the states required by the selected reliability model. The schedule is fixed before state observation, whereas b_{ij}^s and f_{ij}^s adapt after s is observed; each state is treated as constant over the task-completion interval.

Figure 1a shows this two-stage structure: the assignment fixes each task destination, and the realized state determines the continuous resources required by the offloaded tasks.



(**a**) Scheduling and recourse. (**b**) Joint support and bounding box.

Figure 1. System model and uncertainty representation. In (**a**), the task assignment is selected before the joint service state is observed, after which uplink bandwidth and destination CPU resources are adapted. In (**b**), the transformed coordinates $u_i = 1/\sqrt{\eta_i}$ expose the distinction between the jointly specified service-state library and a rectangular corner formed from coordinatewise extrema. The schematic depicts the case in which that corner is not jointly attained by the prescribed finite support.

3.2. Service-State Uncertainty and Reliability Models

The prescribed service-state library is the finite collection of joint spectral-efficiency vectors

$$\boldsymbol{\eta}^s = (\eta_1^s, \ldots, \eta_K^s), \qquad s \in \mathcal{S}. \tag{10}$$

Each state is a jointly attainable multiuser service condition. For the resource geometry in Section 4, define the transformed coordinates

$$u_i^s = \frac{1}{\sqrt{\eta_i^s}}, \qquad \mathbf{u}^s = (u_1^s, \ldots, u_K^s)^{\mathsf{T}}. \tag{11}$$

Larger u_i^s corresponds to poorer spectral efficiency. The prescribed support in these coordinates is

$$\mathcal{U}_{\text{lib}} = \{\mathbf{u}^s : s \in \mathcal{S}\}. \tag{12}$$

The coordinatewise rectangular worst case is constructed from

$$\bar{u}_i = \max_{s \in \mathcal{S}} u_i^s, \qquad \bar{\mathbf{u}} = (\bar{u}_1, \ldots, \bar{u}_K)^{\mathsf{T}}. \tag{13}$$

The vector $\bar{\mathbf{u}}$ combines the marginally poorest service coordinate of every user. It coincides with a library state only when those coordinatewise extrema are attained jointly. Figure 1b illustrates the case in which the rectangular corner is not contained in the prescribed finite support. The finite library is the service-state support modeled by the optimization.

Let $B_{\text{req}}(x; \mathbf{u})$ denote the minimum shared uplink capacity required by schedule x under service vector $\mathbf{u}$ after optimal continuous resource recourse. Section 4 derives this quantity in closed form. For a provisioned capacity B, the statewise failure set of schedule x is

$$\mathcal{F}_B(x) = \left\{s \in \mathcal{S} : B_{\text{req}}(x; \mathbf{u}^s) > B\right\}. \tag{14}$$

Failure is defined by the strict exceedance in (14).

Suppose n observed service states yield empirical probabilities $\widehat{p}_s$, $s \in \mathcal{S}$, and let c_{ts} be a finite ground metric with $c_{tt} = 0$, $c_{ts} = c_{st} > 0$ for $t \neq s$, and the triangle inequality.

Positive off-diagonal distances make the radius-zero ambiguity set equal to the empirical law. The finite-support Wasserstein ambiguity set of radius $\epsilon \geq 0$ is

$$\mathcal{P}_\epsilon = \{\mathbf{p} \in \Delta_S : W_c(\mathbf{p}, \widehat{\mathbf{p}}) \leq \epsilon\}, \tag{15}$$

where Δ_S is the probability simplex. Explicitly,

$$W_c(\mathbf{p}, \widehat{\mathbf{p}}) = \min_{\substack{\pi_{ts} \geq 0 \\ \sum_s \pi_{ts} = \widehat{p}_t,\ \sum_t \pi_{ts} = p_s}} \sum_{t,s \in \mathcal{S}} c_{ts} \pi_{ts}. \tag{16}$$

For any failure set $\mathcal{F} \subseteq \mathcal{S}$, define its worst-case probability over the ambiguity set as

$$R_{\mathrm{wc}}(\mathcal{F}; \epsilon) = \sup_{\mathbf{p} \in \mathcal{P}_\epsilon} \sum_{s \in \mathcal{F}} p_s. \tag{17}$$

The uncertainty models share the same prescribed service-state library but impose different protection requirements. The empirical chance model requires

$$\sum_{s \in \mathcal{F}_B(x)} \widehat{p}_s \leq \alpha, \tag{18}$$

where $\alpha \in (0, 1)$ is the permitted failure probability. The distributionally robust model replaces (18) by

$$R_{\mathrm{wc}}(\mathcal{F}_B(x); \epsilon) \leq \alpha. \tag{19}$$

Hard prescribed-support robustness requires

$$\mathcal{F}_B(x) = \varnothing, \tag{20}$$

whereas rectangular robustness requires feasibility at the coordinatewise corner $\bar{\mathbf{u}}$. This is equivalent to protection over the box $\prod_i [\min_s u_i^s, \bar{u}_i]$ by the monotonicity established in Theorem 2. These models share the same task set, destinations, and resource-recourse mechanism; they differ in the uncertainty description used to determine the required shared capacity.

3.3. Scheduling and Capacity-Planning Problems

The admitted weighted workload of a schedule is

$$U(x) = \sum_{i \in \mathcal{I}} w_i (1 - r_i). \tag{21}$$

In the numerical study, the weights are half-integers, so the equivalent scaled workload

$$U_2(x) = 2U(x) \tag{22}$$

is used for exact integer bookkeeping. Let $\mathcal{X}$ denote the set of binary schedules satisfying (4) together with the local and offloading admissibility conditions above. For a prescribed scaled workload T, define

$$\mathcal{X}_T = \{x \in \mathcal{X} : U_2(x) = T\}. \tag{23}$$

A schedule with infeasible continuous recourse is assigned infinite required capacity. By Theorem 1, inadequate CPU headroom makes a schedule infeasible in every state. Assume that $\mathcal{X}_T$ contains at least one finite-recourse schedule, so the capacity optima are finite.

For an uncertainty model $\mathcal{M}$, let $B_{\mathcal{M}}(x)$ denote the minimum shared uplink capacity required by schedule x. The fixed-workload capacity-planning problem is

$$B^{\star}_{\mathcal{M}}(T) = \min_{x \in \mathcal{X}_T} B_{\mathcal{M}}(x). \tag{24}$$

All four models optimize over the same set $\mathcal{X}_T$; only the uncertainty requirement defining $B_{\mathcal{M}}(x)$ changes. Capacity differences therefore compare uncertainty representations at a common admitted workload.

4. Exact Resource Recourse and Capacity Geometry

Fixing the task schedule reduces the continuous CPU–bandwidth recourse to a closed-form statewise capacity requirement. In the transformed service coordinates of Section 3.2, this requirement becomes a quadratic form that defines the resource geometry used in Section 5.

4.1. Exact Communication–Computation Recourse

Fix a schedule x, a destination $j \in \mathcal{J}$, and a service state $s \in \mathcal{S}$. Let

$$\mathcal{I}_j(x) = \{i \in \mathcal{I} : x_{ij} = 1\} \tag{25}$$

denote the tasks offloaded to destination j. For compactness, define

$$A_{js}(x) = \sum_{i \in \mathcal{I}} x_{ij} \frac{d_i}{\eta_i^s L_{ij}}, \tag{26}$$

$$C_j(x) = \sum_{i \in \mathcal{I}} x_{ij} \frac{c_i}{L_{ij}}, \tag{27}$$

$$S_{js}(x) = \sum_{i \in \mathcal{I}} x_{ij} \frac{\sqrt{d_i c_i / \eta_i^s}}{L_{ij}}. \tag{28}$$

The quantity $C_j(x)$ is the CPU rate that would be consumed if every offloaded task were assigned exactly its deadline-normalized computational rate. Finite communication bandwidth requires strictly positive residual CPU headroom,

$$V_j(x) = F_j - C_j(x) > 0. \tag{29}$$

For fixed (x, j, s), the continuous recourse problem is

$$\begin{aligned} \min_{\{b_{ij}^s, f_{ij}^s\}} \quad & \sum_{i \in \mathcal{I}_j(x)} b_{ij}^s \\ \text{s.t.} \quad & \frac{d_i}{b_{ij}^s \eta_i^s} + \frac{c_i}{f_{ij}^s} \le L_{ij}, \quad i \in \mathcal{I}_j(x), \\ & \sum_{i \in \mathcal{I}_j(x)} f_{ij}^s \le F_j, \\ & b_{ij}^s > 0, \quad f_{ij}^s > 0. \end{aligned} \tag{30}$$

When $\mathcal{I}_j(x) = \varnothing$, its optimum is defined as zero. The optimum has the same square-root proportionality as classical capacity-allocation rules in communication and queueing networks [33,34]; here it follows from deterministic deadline-constrained bandwidth–CPU recourse.

Theorem 1 (Exact resource recourse). *Suppose $\mathcal{I}_j(x) \neq \varnothing$ and $V_j(x) > 0$. The minimum bandwidth required at destination j in state s is*

$$B^s_{j,\min}(x) = A_{js}(x) + \frac{S_{js}(x)^2}{F_j - C_j(x)}. \tag{31}$$

An optimal allocation is

$$f^{s\star}_{ij} = \frac{c_i}{L_{ij}} + \frac{F_j - C_j(x)}{S_{js}(x)} \frac{\sqrt{d_i c_i/\eta^s_i}}{L_{ij}}, \tag{32}$$

$$b^{s\star}_{ij} = \frac{d_i}{\eta^s_i L_{ij}} + \frac{S_{js}(x)}{F_j - C_j(x)} \frac{\sqrt{d_i c_i/\eta^s_i}}{L_{ij}}, \tag{33}$$

for every $i \in \mathcal{I}_j(x)$. If $C_j(x) \geq F_j$ for a nonempty $\mathcal{I}_j(x)$, no finite-bandwidth allocation satisfies the deadline constraints.

Proof of Theorem 1. At a bandwidth-minimizing solution, the deadline constraint of each active task is tight. Write

$$z^s_{ij} = f^s_{ij} - \frac{c_i}{L_{ij}} > 0. \tag{34}$$

Using $f^s_{ij} = c_i/L_{ij} + z^s_{ij}$ in the active deadline equality gives

$$b^s_{ij} = \frac{d_i}{\eta^s_i L_{ij}} + \frac{d_i c_i}{\eta^s_i L^2_{ij}} \frac{1}{z^s_{ij}}. \tag{35}$$

Let

$$q_{ijs} = \frac{\sqrt{d_i c_i/\eta^s_i}}{L_{ij}}. \tag{36}$$

The remaining problem is therefore to minimize $\sum_i q^2_{ijs}/z^s_{ij}$ subject to $\sum_i z^s_{ij} \leq V_j(x)$. Since increasing any z^s_{ij} decreases the objective, the CPU constraint is active whenever $\mathcal{I}_j(x)$ is nonempty. By the Cauchy–Schwarz inequality,

$$\left(\sum_{i \in \mathcal{I}_j(x)} q_{ijs} \right)^2 \leq \left(\sum_{i \in \mathcal{I}_j(x)} \frac{q^2_{ijs}}{z^s_{ij}} \right) \left(\sum_{i \in \mathcal{I}_j(x)} z^s_{ij} \right), \tag{37}$$

with equality when z^s_{ij} is proportional to q_{ijs}. Hence,

$$z^{s\star}_{ij} = \frac{V_j(x)}{S_{js}(x)} q_{ijs}, \tag{38}$$

and the minimum residual term is $S_{js}(x)^2/V_j(x)$. Substitution into (35) yields (31)–(33). If $C_j(x) \geq F_j$, the positive variables z^s_{ij} cannot have a positive aggregate budget, so finite bandwidth is impossible. □

The CPU constraints are destination-specific, while the shared uplink capacity is the sum of the bandwidth assigned across destinations. The exact statewise requirement of schedule x is therefore

$$B^s_{\text{req}}(x) = \sum_{j \in \mathcal{J}} B^s_{j,\min}(x), \tag{39}$$

with $B^s_{j,\min}(x) = 0$ for an inactive destination. A schedule with finite recourse is feasible under state s and provisioned capacity B if and only if

$$B^s_{\mathrm{req}}(x) \le B. \tag{40}$$

Thus, the continuous communication–computation recourse can be eliminated exactly once the discrete schedule and service state are fixed.

4.2. Quadratic Capacity Geometry and Destination Congestion

The exact recourse formula becomes particularly simple in the transformed coordinates $u_i = 1/\sqrt{\eta_i}$. For each destination, define the nonnegative vectors $\boldsymbol{a}_j(x), \boldsymbol{g}_j(x) \in \mathbb{R}^K$ by

$$[\boldsymbol{a}_j(x)]_i = x_{ij}\frac{d_i}{L_{ij}}, \tag{41}$$

$$[\boldsymbol{g}_j(x)]_i = x_{ij}\frac{\sqrt{d_i c_i}}{L_{ij}}. \tag{42}$$

For a generic positive service vector $\mathbf{u}$, (31) can be written as

$$B_{j,\min}(x;\mathbf{u}) = \mathbf{u}^{\mathsf{T}}\,\mathrm{Diag}(\boldsymbol{a}_j(x))\mathbf{u} + \frac{\left(\boldsymbol{g}_j(x)^{\mathsf{T}}\mathbf{u}\right)^2}{F_j - C_j(x)}. \tag{43}$$

Summing across destinations gives the following representation.

Theorem 2 (Quadratic capacity representation)**.** *For a fixed schedule x with positive CPU headroom at every active remote destination,*

$$B_{\mathrm{req}}(x;\mathbf{u}) = \mathbf{u}^{\mathsf{T}} Q_x \mathbf{u}, \tag{44}$$

where

$$Q_x = \sum_{j\in\mathcal{J}}\left[\mathrm{Diag}(\boldsymbol{a}_j(x)) + \frac{\boldsymbol{g}_j(x)\boldsymbol{g}_j(x)^{\mathsf{T}}}{F_j - C_j(x)}\right]. \tag{45}$$

The matrix Q_x is symmetric positive semidefinite and entrywise nonnegative. Its principal block on the offloaded-task coordinates is positive definite; coordinates corresponding to local or rejected tasks have zero rows and columns.

Proof of Theorem 2. Equation (44) follows by substituting $1/\eta_i = u_i^2$ and $1/\sqrt{\eta_i} = u_i$ into (31) and summing (43) over j. Each diagonal term in (45) is positive semidefinite, and each rank-one term is positive semidefinite because its denominator is positive. All entries are nonnegative. If task i is offloaded, exactly one destination contributes the strictly positive diagonal entry d_i/L_{ij} to the ith active coordinate. Hence the active principal block is positive definite. Local and rejected tasks have $x_{ij} = 0$ for every j and therefore contribute zero rows and columns. □

The quadratic representation gives several direct consequences. For any two nonnegative service vectors $\mathbf{u}$ and $\mathbf{v}$,

$$B_{\mathrm{req}}(x;\mathbf{u}) - B_{\mathrm{req}}(x;\mathbf{v}) = (\mathbf{u}-\mathbf{v})^{\mathsf{T}} Q_x (\mathbf{u}+\mathbf{v}), \tag{46}$$

and therefore

$$\left|B_{\mathrm{req}}(x;\mathbf{u}) - B_{\mathrm{req}}(x;\mathbf{v})\right| \le \|Q_x(\mathbf{u}+\mathbf{v})\|_2 \|\mathbf{u}-\mathbf{v}\|_2, \tag{47}$$

$$\le \|Q_x\|_2 \|\mathbf{u}-\mathbf{v}\|_2 \|\mathbf{u}+\mathbf{v}\|_2. \tag{48}$$

Because Q_x is entrywise nonnegative, $B_{\mathrm{req}}(x;\mathbf{u})$ is coordinatewise nondecreasing on $\mathbf{u} \ge \mathbf{0}$. This monotonicity is the link between the prescribed joint-state support and the rectangular worst-case corner developed in Section 5.1.

The recourse expression also gives the CPU–spectrum exchange at each destination. For fixed x and state s,

$$\frac{\partial B^s_{j,\min}}{\partial F_j} = -\frac{S_{js}(x)^2}{\left(F_j - C_j(x)\right)^2}, \tag{49}$$

$$\frac{\partial^2 B^s_{j,\min}}{\partial F_j^2} = \frac{2S_{js}(x)^2}{\left(F_j - C_j(x)\right)^3}. \tag{50}$$

Additional CPU capacity reduces the required spectrum with diminishing marginal benefit. For a fixed state, a continuous relaxation of the assignments makes A_{js}, C_j, and S_{js} affine, while $S_{js}^2/(F_j - C_j)$ is quadratic-over-linear and convex on the positive-headroom domain; its epigraph is rotated-second-order-cone representable. The scheduling problems themselves retain binary assignments.

The same headroom term creates an increasing marginal communication cost as tasks accumulate at one destination. Fix j and s, and for a task subset $\mathcal{A} \subseteq \mathcal{I}$ define

$$\Phi_{js}(\mathcal{A}) = A(\mathcal{A}) + \frac{G(\mathcal{A})^2}{F_j - H(\mathcal{A})}, \qquad H(\mathcal{A}) < F_j, \tag{51}$$

where

$$A(\mathcal{A}) = \sum_{i\in\mathcal{A}} a_i, \qquad a_i = \frac{d_i}{\eta_i^s L_{ij}}, \tag{52}$$

$$G(\mathcal{A}) = \sum_{i\in\mathcal{A}} g_i, \qquad g_i = \frac{\sqrt{d_i c_i/\eta_i^s}}{L_{ij}}, \tag{53}$$

$$H(\mathcal{A}) = \sum_{i\in\mathcal{A}} h_i, \qquad h_i = \frac{c_i}{L_{ij}}. \tag{54}$$

For a set of tasks assigned to destination j, Φ_{js} is exactly the destination bandwidth in (31).

Proposition 1 (Supermodular destination congestion)**.** *The set function Φ_{js} is monotone nondecreasing and supermodular on its positive-headroom domain. In particular, for feasible $\mathcal{A} \subseteq \mathcal{B}$ and $k \notin \mathcal{B}$ such that both augmented sets remain feasible,*

$$\Phi_{js}(\mathcal{A}\cup\{k\}) - \Phi_{js}(\mathcal{A}) \le \Phi_{js}(\mathcal{B}\cup\{k\}) - \Phi_{js}(\mathcal{B}). \tag{55}$$

Proof of Proposition 1. Let $G = G(\mathcal{A})$, $H = H(\mathcal{A})$, and $v = F_j - H$. For task k, the exact marginal is

$$\Delta_k \Phi_{js}(\mathcal{A}) = a_k + \frac{2Gg_k + g_k^2}{v - h_k} + \frac{G^2 h_k}{v(v-h_k)}. \tag{56}$$

Every term is nonnegative, proving monotonicity. Moreover, the right-hand side is nondecreasing in G and in H over the positive-headroom domain, because increasing H

decreases the positive quantity v. If $\mathcal{A} \subseteq \mathcal{B}$, then $G(\mathcal{A}) \leq G(\mathcal{B})$ and $H(\mathcal{A}) \leq H(\mathcal{B})$, which gives (55). □

The exact marginal in (56) increases as CPU headroom is consumed and satisfies the immediate lower bound

$$\Delta_k \Phi_{js}(\mathcal{A}) \geq a_k + \frac{g_k^2}{F_j}. \tag{57}$$

Thus, the marginal spectrum cost of an additional offloaded task increases with destination load and depends jointly on task demand and remaining CPU headroom.

5. Capacity Cost of Uncertainty Representation

The quadratic resource geometry permits direct comparison of hard protection over the prescribed service-state support, coordinatewise support rectangularization, and Wasserstein probability ambiguity.

5.1. Support-Rectangularization Capacity Premium

For a fixed schedule x with finite recourse, define the hard prescribed-support capacity and rectangular capacity as

$$B_{\mathrm{HARD}}(x) = \max_{s \in \mathcal{S}} B_{\mathrm{req}}(x; \mathbf{u}^s), \tag{58}$$

$$B_{\mathrm{BOX}}(x) = B_{\mathrm{req}}(x; \bar{\mathbf{u}}). \tag{59}$$

Because $\mathbf{u}^s \leq \bar{\mathbf{u}}$ coordinatewise and Q_x is entrywise nonnegative, Theorem 2 gives

$$B_{\mathrm{HARD}}(x) \leq B_{\mathrm{BOX}}(x). \tag{60}$$

There is also a useful geometric interpretation in the transformed $\mathbf{u}$-space. Convexity of $\mathbf{u} \mapsto \mathbf{u}^{\mathsf{T}} Q_x \mathbf{u}$ gives

$$\max_{\mathbf{u} \in \mathrm{conv}\{\mathbf{u}^s : s \in \mathcal{S}\}} \mathbf{u}^{\mathsf{T}} Q_x \mathbf{u} = \max_{s \in \mathcal{S}} (\mathbf{u}^s)^{\mathsf{T}} Q_x \mathbf{u}^s = B_{\mathrm{HARD}}(x), \tag{61}$$

whereas coordinatewise monotonicity makes $B_{\mathrm{BOX}}(x)$ the maximum over the axis-aligned box $[\underline{\mathbf{u}}, \bar{\mathbf{u}}]$, with $\underline{u}_i = \min_s u_i^s$. Thus support rectangularization enlarges the convex hull of the prescribed transformed states to its coordinatewise bounding box.

The schedule-level capacity cost of replacing the prescribed joint support by the rectangular corner is therefore

$$\Pi(x) = B_{\mathrm{BOX}}(x) - B_{\mathrm{HARD}}(x) \geq 0. \tag{62}$$

For each state, let

$$\mathbf{d}_s = \bar{\mathbf{u}} - \mathbf{u}^s \geq \mathbf{0}, \tag{63}$$

and define the schedule-dependent Q_x-distance from the rectangular corner to the prescribed support by

$$r_Q(x)^2 = \min_{s \in \mathcal{S}} \mathbf{d}_s^{\mathsf{T}} Q_x \mathbf{d}_s, \tag{64}$$

$$R_Q(x)^2 = \bar{\mathbf{u}}^{\mathsf{T}} Q_x \bar{\mathbf{u}}. \tag{65}$$

Define also the first-order lower and upper envelopes

$$L_1(x) = \min_{s\in\mathcal{S}} \mathbf{d}_s^{\mathsf{T}} Q_x(\bar{\mathbf{u}} + \underline{\mathbf{u}}), \tag{66}$$

$$U_1(x) = \min_{s\in\mathcal{S}} 2\mathbf{d}_s^{\mathsf{T}} Q_x \bar{\mathbf{u}}. \tag{67}$$

If $A(x)$ is the set of active offloaded coordinates, let

$$c_x = \max_{i\in A(x)} \frac{2\bar{u}_i}{\bar{u}_i + \underline{u}_i}, \tag{68}$$

whenever $A(x) \neq \varnothing$. Since all service coordinates are positive, $1 \leq c_x < 2$.

Theorem 3 (Support-rectangularization capacity premium)**.** *For any fixed schedule x with finite resource recourse,*

$$\Pi(x) = \min_{s\in\mathcal{S}} \mathbf{d}_s^{\mathsf{T}} Q_x(\bar{\mathbf{u}} + \mathbf{u}^s), \tag{69}$$

and, if at least one task is offloaded,

$$r_Q(x)^2 \leq L_1(x) \leq \Pi(x) \leq \min\Big\{U_1(x),\, 2R_Q(x)r_Q(x) - r_Q(x)^2\Big\}, \tag{70}$$

with the additional enclosure

$$U_1(x) \leq c_x L_1(x), \qquad 1 \leq c_x < 2. \tag{71}$$

Moreover, $\Pi(x) = 0$ if and only if some prescribed state attains the rectangular corner on every active offloaded coordinate. If no task is offloaded, $Q_x = 0$ and all premium and bound quantities are zero.

Proof of Theorem 3. Using Theorem 2,

$$\begin{aligned}\Pi(x) &= \bar{\mathbf{u}}^{\mathsf{T}} Q_x \bar{\mathbf{u}} - \max_{s\in\mathcal{S}} (\mathbf{u}^s)^{\mathsf{T}} Q_x \mathbf{u}^s \\ &= \min_{s\in\mathcal{S}} \Big[\bar{\mathbf{u}}^{\mathsf{T}} Q_x \bar{\mathbf{u}} - (\mathbf{u}^s)^{\mathsf{T}} Q_x \mathbf{u}^s\Big],\end{aligned} \tag{72}$$

which factors as (69). Since $\bar{\mathbf{u}} = \mathbf{u}^s + \mathbf{d}_s$,

$$\mathbf{d}_s^{\mathsf{T}} Q_x(\bar{\mathbf{u}} + \mathbf{u}^s) = \mathbf{d}_s^{\mathsf{T}} Q_x \mathbf{d}_s + 2\mathbf{d}_s^{\mathsf{T}} Q_x \mathbf{u}^s. \tag{73}$$

The second term is nonnegative because $\mathbf{d}_s$, $\mathbf{u}^s$, and the entries of Q_x are nonnegative. For every state,

$$\mathbf{d}_s^{\mathsf{T}} Q_x(\bar{\mathbf{u}} + \underline{\mathbf{u}}) \leq \mathbf{d}_s^{\mathsf{T}} Q_x(\bar{\mathbf{u}} + \mathbf{u}^s) \leq 2\mathbf{d}_s^{\mathsf{T}} Q_x \bar{\mathbf{u}}, \tag{74}$$

because $\underline{\mathbf{u}} \leq \mathbf{u}^s \leq \bar{\mathbf{u}}$ coordinatewise. Minimizing the three expressions gives $L_1(x) \leq \Pi(x) \leq U_1(x)$. Also, $\bar{\mathbf{u}} + \underline{\mathbf{u}} - \mathbf{d}_s = \mathbf{u}^s + \underline{\mathbf{u}} \geq 0$, so $\mathbf{d}_s^{\mathsf{T}} Q_x \mathbf{d}_s \leq \mathbf{d}_s^{\mathsf{T}} Q_x(\bar{\mathbf{u}} + \underline{\mathbf{u}})$ and hence $r_Q(x)^2 \leq L_1(x)$.

On the active coordinates, the definition of c_x gives $2\bar{\mathbf{u}}_A \leq c_x(\bar{\mathbf{u}}_A + \underline{\mathbf{u}}_A)$ coordinatewise. Multiplication by the nonnegative row vector $\mathbf{d}_{s,A}^{\mathsf{T}} Q_{x,A}$ and minimization over s yield $U_1(x) \leq c_x L_1(x)$.

Let s_Q attain the minimum in (64). Since $\mathbf{u}^{s_Q} = \bar{\mathbf{u}} - \mathbf{d}_{s_Q}$,

$$\Pi(x) \leq 2\mathbf{d}_{s_Q}^{\mathsf{T}} Q_x \bar{\mathbf{u}} - \mathbf{d}_{s_Q}^{\mathsf{T}} Q_x \mathbf{d}_{s_Q}. \tag{75}$$

The Cauchy–Schwarz inequality for the seminorm induced by the positive-semidefinite matrix Q_x yields

$$\mathbf{d}_{s_Q}^{\mathsf{T}} Q_x \bar{\mathbf{u}} \le r_Q(x) R_Q(x), \tag{76}$$

which proves the upper bound. If $\Pi(x) = 0$, the lower bound forces $\mathbf{d}_s^{\mathsf{T}} Q_x \mathbf{d}_s = 0$ for a minimizing state. The active principal block of Q_x is positive definite by Theorem 2, so the active coordinates of that state equal the corresponding coordinates of $\bar{\mathbf{u}}$. The converse follows immediately from (69). The no-offloading case has $Q_x = 0$. □

Corollary 1 (Resource-weighted joint-extreme deficit)**.** *For any fixed schedule x with finite recourse and at least one offloaded task, define*

$$\omega_i(x) = \frac{\bar{u}_i \, [Q_x \bar{\mathbf{u}}]_i}{B_{\mathrm{BOX}}(x)}, \qquad i \in A(x), \tag{77}$$

$$\chi(x) = \min_{s \in \mathcal{S}} \sum_{i \in A(x)} \omega_i(x) \left(1 - \frac{u_i^s}{\bar{u}_i}\right). \tag{78}$$

Then $\omega_i(x) > 0$, $\sum_{i \in A(x)} \omega_i(x) = 1$, $0 \le \chi(x) < 1$, and

$$U_1(x) = 2 B_{\mathrm{BOX}}(x) \chi(x), \qquad \frac{2}{c_x} \chi(x) \le \frac{\Pi(x)}{B_{\mathrm{BOX}}(x)} \le 2\chi(x). \tag{79}$$

Moreover, $\chi(x) = 0$ if and only if $\Pi(x) = 0$.

Proof of Corollary 1. The positive active diagonal of Q_x gives $\omega_i(x) > 0$, and the numerators in (77) sum to $B_{\mathrm{BOX}}(x)$. Direct substitution gives $\chi(x) = U_1(x)/(2B_{\mathrm{BOX}}(x))$. Theorem 3 gives $\Pi(x) \le U_1(x) \le c_x L_1(x) \le c_x \Pi(x)$, proving the bounds. The zero condition follows from the same theorem. □

The quantity $\chi(x)$ separates support geometry from schedule sensitivity: the normalized shortfalls $1 - u_i^s / \bar{u}_i$ measure failure of the prescribed support to jointly attain its marginal worst conditions, while $\omega_i(x)$ supplies their contribution to rectangular capacity. Unlike the second-order distance bound r_Q^2, this representation retains the first-order support–resource interaction.

The capacity premium in (62) is schedule dependent. The planning problem permits the schedule to change under each uncertainty model. Let

$$x^\star_{\mathrm{HARD}} \in \arg\min_{x \in \mathcal{X}_T} B_{\mathrm{HARD}}(x), \qquad x^\star_{\mathrm{BOX}} \in \arg\min_{x \in \mathcal{X}_T} B_{\mathrm{BOX}}(x), \tag{80}$$

and define the optimized support-rectangularization premium

$$\Delta B^\star_{\mathrm{rect}} = B^\star_{\mathrm{BOX}}(T) - B^\star_{\mathrm{HARD}}(T). \tag{81}$$

Theorem 4 (Optimized support-rectangularization sandwich)**.** *For a common nonempty fixed-workload schedule set $\mathcal{X}_T$ with finite attained minima,*

$$\Pi(x^\star_{\mathrm{BOX}}) \le \Delta B^\star_{\mathrm{rect}} \le \Pi(x^\star_{\mathrm{HARD}}). \tag{82}$$

The lower equality holds if and only if the selected BOX-optimal schedule is also HARD-optimal, and the upper equality holds if and only if the selected HARD-optimal schedule is also BOX-optimal. No uniqueness assumption is required.

Proof of Theorem 4. Optimality of $x^\star_{\mathrm{HARD}}$ gives $B^\star_{\mathrm{HARD}}(T) \le B_{\mathrm{HARD}}(x^\star_{\mathrm{BOX}})$. Hence,

$$\begin{aligned}\Delta B^\star_{\mathrm{rect}} &= B_{\mathrm{BOX}}(x^\star_{\mathrm{BOX}}) - B^\star_{\mathrm{HARD}}(T)\\ &\ge B_{\mathrm{BOX}}(x^\star_{\mathrm{BOX}}) - B_{\mathrm{HARD}}(x^\star_{\mathrm{BOX}}) = \Pi(x^\star_{\mathrm{BOX}}).\end{aligned} \tag{83}$$

Similarly, $B^\star_{\mathrm{BOX}}(T) \le B_{\mathrm{BOX}}(x^\star_{\mathrm{HARD}})$, which gives the upper bound. The corresponding slack identities are

$$\Delta B^\star_{\mathrm{rect}} - \Pi(x^\star_{\mathrm{BOX}}) = B_{\mathrm{HARD}}(x^\star_{\mathrm{BOX}}) - B^\star_{\mathrm{HARD}}(T), \tag{84}$$

$$\Pi(x^\star_{\mathrm{HARD}}) - \Delta B^\star_{\mathrm{rect}} = B_{\mathrm{BOX}}(x^\star_{\mathrm{HARD}}) - B^\star_{\mathrm{BOX}}(T). \tag{85}$$

Each equality condition therefore follows from membership of the selected optimizer in the opposite model's optimal-schedule set. □

The optimized support-rectangularization cost is therefore enclosed by the schedule-level premiums evaluated at the HARD- and BOX-optimal schedules, with equality characterized by cross-optimality.

5.2. Distributionally Robust Capacity and Hardening

For a fixed schedule x with finite recourse, write its prescribed-state capacity requirements as

$$b_s(x) = B_{\mathrm{req}}(x; \mathbf{u}^s), \qquad s \in \mathcal{S}. \tag{86}$$

For any provisioned capacity B, the strict failure set is $\mathcal{F}_B(x)$ from (14). The minimum capacity that satisfies the Wasserstein reliability requirement at radius ϵ is

$$B_\epsilon(x) = \inf\{B \ge 0 : R_{\mathrm{wc}}(\mathcal{F}_B(x); \epsilon) \le \alpha\}. \tag{87}$$

At $\epsilon = 0$, the ambiguity set contains only the empirical distribution, so $B_0(x) = B_{\mathrm{EMP}}(x)$ is the empirical chance-constrained capacity. The HARD and BOX capacities remain those in (58) and (59).

For a nonempty failure set $\mathcal{F}$, define

$$d(t, \mathcal{F}) = \min_{s \in \mathcal{F}} c_{ts}. \tag{88}$$

Wasserstein chance-constraint reformulations are established in [29,30]. For the fixed finite support considered here, linear-programming duality applied to (16) gives the worst-case failure probability in the form

$$R_{\mathrm{wc}}(\mathcal{F}; \epsilon) = \min_{\lambda \ge 0} \left\{ \lambda\epsilon + \sum_{t \in \mathcal{S}} \widehat{p}_t (1 - \lambda d(t, \mathcal{F}))_+ \right\}, \tag{89}$$

where $(a)_+ = \max\{a, 0\}$ and $R_{\mathrm{wc}}(\varnothing; \epsilon) = 0$. For each source state t, the transport-dual contribution is $\max_s\{\mathbf{1}_{\mathcal{F}}(s) - \lambda c_{ts}\} = (1 - \lambda d(t, \mathcal{F}))_+$: a source outside $\mathcal{F}$ may remain at t at zero cost, while a source inside $\mathcal{F}$ has $d(t, \mathcal{F}) = 0$.

Define the family of state sets whose worst-case probability is admissible at radius ϵ by

$$\mathfrak{D}_\epsilon = \{\mathcal{F} \subseteq \mathcal{S} : R_{\mathrm{wc}}(\mathcal{F}; \epsilon) \le \alpha\}. \tag{90}$$

This family is downward closed: if $\mathcal{G} \subseteq \mathcal{F} \in \mathfrak{D}_\epsilon$, then $\mathcal{G} \in \mathfrak{D}_\epsilon$.

Proposition 2 (Droppable-state representation)**.** *For every fixed schedule x with finite recourse,*

$$B_\epsilon(x) = \min_{\mathcal{F}\in\mathfrak{D}_\epsilon} \max_{s\in\mathcal{S}\setminus\mathcal{F}} b_s(x), \tag{91}$$

where the maximum over an empty retained set is irrelevant because $\mathcal{S} \notin \mathfrak{D}_\epsilon$ for $\alpha < 1$. For a retained state set $\mathcal{A} \subseteq \mathcal{S}$, define

$$B^\star_{\mathrm{HARD}}(T;\mathcal{A}) = \min_{x\in\mathcal{X}_T} \max_{s\in\mathcal{A}} b_s(x). \tag{92}$$

Then

$$B^\star_\epsilon(T) = \min_{\mathcal{F}\in\mathfrak{D}_\epsilon} B^\star_{\mathrm{HARD}}(T;\mathcal{S}\setminus\mathcal{F}). \tag{93}$$

If $W(x) = \arg\max_s b_s(x)$ denotes the complete set of worst-capacity states, then

$$B_\epsilon(x) < B_{\mathrm{HARD}}(x) \quad\Longleftrightarrow\quad W(x) \in \mathfrak{D}_\epsilon. \tag{94}$$

Proof. For any $\mathcal{F} \in \mathfrak{D}_\epsilon$, set $B = \max_{s\notin\mathcal{F}} b_s(x)$. The resulting strict failure set is contained in $\mathcal{F}$, so downward closure gives a worst-case failure probability at most α; hence $B_\epsilon(x)$ is no larger than the right-hand side of (91). Conversely, for any feasible B, its actual failure set $\mathcal{F}_B(x)$ belongs to $\mathfrak{D}_\epsilon$ and $\max_{s\notin\mathcal{F}_B(x)} b_s(x) \le B$. Taking the infimum over feasible B proves (91). Minimizing over the finite schedule set and interchanging the two finite minima gives (93). Finally, capacity can fall strictly below the hard requirement exactly when every state tied at the hard maximum can be included in one admissible droppable set, which is equivalent by downward closure to $W(x) \in \mathfrak{D}_\epsilon$. □

The family $\mathfrak{D}_\epsilon$ depends on the empirical law, transport metric, α, and ϵ, whereas the retained-state cost remains schedule dependent through $b_s(x)$. Equation (93) therefore separates admissible failure-set selection from schedule-dependent resource cost, with the prescribed service geometry retained in the transport metric.

Proposition 3 (Distributionally robust capacity hierarchy)**.** *Let $0 \le \epsilon_1 \le \epsilon_2$ and $0 < \alpha < 1$. For any fixed schedule x with finite statewise requirements,*

$$B_0(x) \le B_{\epsilon_1}(x) \le B_{\epsilon_2}(x) \le B_{\mathrm{HARD}}(x) \le B_{\mathrm{BOX}}(x). \tag{95}$$

Moreover, $B_\epsilon(x)$ is attained at one of the distinct values in $\{b_s(x) : s \in \mathcal{S}\}$. For the common fixed-workload set $\mathcal{X}_T$,

$$B^\star_0(T) \le B^\star_{\epsilon_1}(T) \le B^\star_{\epsilon_2}(T) \le B^\star_{\mathrm{HARD}}(T) \le B^\star_{\mathrm{BOX}}(T). \tag{96}$$

Proof of Proposition 3. Enlarging the Wasserstein radius enlarges the ambiguity set and therefore cannot decrease the worst-case probability of any fixed failure set. Hence, a capacity feasible at ϵ_2 is feasible at every smaller radius, which gives the first three terms of (95). The capacity $B_{\mathrm{HARD}}(x) = \max_s b_s(x)$ makes the failure set empty and is feasible for every radius. The final inequality is (60).

Let $\beta_1 < \cdots < \beta_m$ be the distinct statewise requirements. For $B < \beta_1$, every state fails, and the risk is one. Between consecutive levels, the failure set is unchanged, while at $B = \beta_k$ all states tied at that level become nonfailing because failure is defined by strict exceedance. The first feasible value can therefore be selected from $\{\beta_1, \ldots, \beta_m\}$. Finally, every inequality in (95) holds pointwise for the same schedule set $\mathcal{X}_T$; minimizing each side over that common set gives (96). □

The radius at which this ordered frontier reaches the terminal HARD regime is characterized by event transport thresholds. For a probability vector $\mathbf{p}$, write $\mathbf{p}(\mathcal{F}) = \sum_{s\in\mathcal{F}} p_s$. For every nonempty set $\mathcal{F}$, define

$$\kappa_{\mathcal{F}}(\alpha) = \inf\{W_c(\mathbf{p}, \widehat{\mathbf{p}}) : \mathbf{p} \in \Delta_S,\ \mathbf{p}(\mathcal{F}) \geq \alpha\}. \tag{97}$$

Thus $\kappa_{\mathcal{F}}(\alpha) = 0$ when $\widehat{\mathbf{p}}(\mathcal{F}) \geq \alpha$; otherwise it is the minimum transport budget required to raise that probability to α. Events with empirical probability strictly above α are already infeasible at radius zero. In the finite metric space, it is obtained by moving the required probability mass from complement states into $\mathcal{F}$ in increasing order of $d(t, \mathcal{F})$, with a fractional final transfer when necessary.

Define the event-level hardening radius for the prescribed support, empirical law, and metric by

$$\epsilon_{\text{hard}} = \max_{s\in\mathcal{S}} \kappa_{\{s\}}(\alpha). \tag{98}$$

For any nonempty $\mathcal{F}$ and any $s \in \mathcal{F}$,

$$\kappa_{\mathcal{F}}(\alpha) \leq \kappa_{\{s\}}(\alpha), \tag{99}$$

because moving probability into the singleton $\{s\}$ is also sufficient to raise the probability of $\mathcal{F}$.

Theorem 5 (Finite-step frontier and hardening)**.** *For a fixed schedule x with finite recourse, the mapping $\epsilon \mapsto B_\epsilon(x)$ is a nondecreasing finite-step function whose values are drawn from the distinct statewise requirements $\{b_s(x)\}$. Its terminal plateau is $B_{\text{HARD}}(x)$. More precisely,*

$$\epsilon > \epsilon_{\text{hard}} \quad \Longrightarrow \quad B_\epsilon(x) = B_{\text{HARD}}(x) \tag{100}$$

for every schedule x with finite recourse. For $0 \leq \epsilon < \epsilon_{\text{hard}}$, at least one nonempty event in 2^S remains admissible. This is an event-level statement; schedule-generated failure sets, and hence the optimized capacity, may reach HARD at a smaller radius. At equality, a nonempty event may remain feasible under the non-strict reliability condition.

Because $\mathcal{X}_T$ is finite, the optimized frontier $\epsilon \mapsto B^\star_\epsilon(T)$ is also nondecreasing and finite-step, and

$$\epsilon > \epsilon_{\text{hard}} \quad \Longrightarrow \quad B^\star_\epsilon(T) = B^\star_{\text{HARD}}(T). \tag{101}$$

Proof of Theorem 5. For a fixed schedule, let $\beta_1 < \cdots < \beta_m$ denote the distinct statewise requirements and set $\mathcal{F}_k = \{s : b_s(x) > \beta_k\}$. A level whose empirical failure probability exceeds α is infeasible already at radius zero. Otherwise, if $\mathcal{F}_k$ is nonempty, transporting complementary mass in increasing order of its strictly positive distance to $\mathcal{F}_k$ gives

$$R_{\text{wc}}(\mathcal{F}_k; \epsilon) \leq \alpha \quad \Longleftrightarrow \quad \epsilon \leq \kappa_{\mathcal{F}_k}(\alpha). \tag{102}$$

The empty failure set is feasible for every radius. Hence

$$B_\epsilon(x) = \min\left\{\beta_k : \mathcal{F}_k = \varnothing \text{ or } \left[\sum_{s\in\mathcal{F}_k} \widehat{p}_s \leq \alpha,\ \epsilon \leq \kappa_{\mathcal{F}_k}(\alpha)\right]\right\}, \tag{103}$$

which is nondecreasing and changes only at finitely many transport thresholds.

If $\epsilon > \epsilon_{\text{hard}}$, then (99) implies $\epsilon > \kappa_{\mathcal{F}}(\alpha)$ for every nonempty failure set. Consequently, every nonempty failure set has worst-case probability strictly greater than α, and feasibility requires $\mathcal{F}_B(x) = \varnothing$. This is equivalent to $B \geq \max_s b_s(x)$ and proves (100). Conversely, if

$0 \le \epsilon < \epsilon_{\text{hard}}$, a singleton attaining the maximum in (98) has empirical probability below α and remains feasible at that radius. This gives event-level sharpness; schedule-generated failure sets can be more restrictive. At equality, such an event can attain worst-case probability exactly α and remain feasible. Equation (103) gives the schedule-level breakpoints.

Finally, the discrete assignment model makes $\mathcal{X}_T$ finite. Infeasible schedules have infinite capacity at every radius and can be discarded. Each remaining schedule-level frontier has finitely many values and breakpoints, so their union is finite. Monotonicity of the optimized lower envelope follows from Proposition 3, and (101) follows from the fixed-schedule equality for every $x \in \mathcal{X}_T$. □

Let

$$D_c = \max_{s,t \in \mathcal{S}} c_{st} \tag{104}$$

denote the diameter of the finite ground metric. Since no more than α probability mass must be transported to raise a singleton event to probability α, and every transport distance is at most D_c,

$$\epsilon_{\text{hard}} \le \alpha D_c. \tag{105}$$

This provides a uniform support-metric upper bound on the hardening radius.

For independent and identically distributed (IID) observations from a law $\mathbf{p}^\star$ on the fixed support of cardinality $S \ge 2$, the finite-alphabet deviation inequality [35] gives, for $r > 0$,

$$\Pr\{\|\widehat{\mathbf{p}} - \mathbf{p}^\star\|_1 > r\} \le (2^S - 2)e^{-nr^2/2}. \tag{106}$$

Since $W_c(\mathbf{p}^\star, \widehat{\mathbf{p}}) \le (D_c/2)\|\widehat{\mathbf{p}} - \mathbf{p}^\star\|_1$, an analytical $(1-\delta)$ containment radius, for $0 < \delta < 1$, is

$$\epsilon_{\text{conf}}(n, \delta) = D_c \sqrt{\frac{\log((2^S - 2)/\delta)}{2n}}. \tag{107}$$

Combining (107) with Theorem 5 gives the direct sample-size implication

$$\epsilon_{\text{conf}}(n, \delta) > \epsilon_{\text{hard}} \quad \Longrightarrow \quad B^\star_{\epsilon_{\text{conf}}}(T) = B^\star_{\text{HARD}}(T), \tag{108}$$

where the comparison uses the same empirical law on both sides. For $\epsilon_{\text{hard}} > 0$, placing this radius strictly below the event-level threshold is equivalent to

$$n > \frac{D_c^2}{2\epsilon_{\text{hard}}^2} \log\left(\frac{2^S - 2}{\delta}\right). \tag{109}$$

For a fixed empirical law and metric, (109) gives the sample sizes for which $\epsilon_{\text{conf}}(n, \delta) < \epsilon_{\text{hard}}$. In that regime, the event-level hardening theorem no longer forces the confidence-calibrated capacity to equal HARD; schedule-realizable failure sets determine whether the optimized frontier has already hardened. Equations (107)–(109) apply to IID sampling on the prescribed finite support with a metric fixed independently of the observations.

6. Numerical Results

Two fixed-workload populations are used. The first contains 60 instances, with 20 each in clustered, moderate, and diverse user-angle regimes, and compares hard protection over the prescribed service-state support with its coordinatewise BOX model. The second contains 20 independent diverse-regime instances and evaluates the EMP–Wasserstein–HARD–BOX capacity hierarchy. Each uncertainty model reoptimizes the discrete schedule at the same admitted workload. Every minimum-capacity result is reported as a deterministic lower–upper optimization interval; intervals of width at most 1 Hz are termed

point-resolved. For ordered models with $\ell_a \le B_a^\star \le u_a$ and $\ell_b \le B_b^\star \le u_b$, the corresponding premium satisfies

$$[\max\{0, \ell_b - u_a\},\ u_b - \ell_a]. \tag{110}$$

Population summaries average the instance-wise endpoints of these intervals.

6.1. Numerical Setup and Service-State Construction

Each numerical instance has $K = 20$ tasks and $S = 64$ jointly generated service states. A service-state generator based on a reconfigurable intelligent surface (RIS) produces the correlated multiuser spectral-efficiency library, which is fixed before scheduling. It generates correlated state variation for the optimization study and follows the standard programmable-surface channel abstraction [36] with a 3.5 GHz carrier, an 8×8 RIS, half-wavelength spacing, and a 2-bit phase alphabet. Table 2 gives the remaining system and sampling parameters.

Let $\mathbf{p}_i$ be the position of user i, $\mathbf{p}_n$ the position of RIS element n, and $\mathbf{p}_A$ the access-point position. With wavelength λ, wave number $k_0 = 2\pi/\lambda$, user–AP distance $r_{iA} = \|\mathbf{p}_i - \mathbf{p}_A\|_2$, user–element distance $r_{in} = \|\mathbf{p}_i - \mathbf{p}_n\|_2$, and element–AP distance $r_{nA} = \|\mathbf{p}_n - \mathbf{p}_A\|_2$, the generator uses

$$h_{i,d} = \frac{e^{-\mathrm{j}k_0 r_{iA}}}{r_{iA}}, \qquad c_{i,n} = \frac{e^{-\mathrm{j}k_0(r_{in}+r_{nA})}}{r_{in} r_{nA}}, \qquad h_i(\mathbf{q}) = h_{i,d} + \sum_{n=1}^{64} c_{i,n} e^{\mathrm{j}\phi_n(q_n)}. \tag{111}$$

Equation (111) is an element-separable spherical-distance line-of-sight construction used to generate correlated state-to-state service variation.

For element n, let $g_n = \sum_i c_{i,n}$. Quantizing $-\arg(g_n)$ modulo 2π to the nearest 2-bit phase defines the nominal profile, with exact half-step ties assigned to the upper phase index. Each of the remaining 63 states changes seven distinct elements by one randomly signed phase step modulo four; duplicate phase vectors are regenerated. The resulting 64 vectors form the prescribed joint service-state library, from which the rectangular corner is constructed separately.

For each user, a nominal signal-to-noise ratio (SNR) $\gamma_{i,0}^{\mathrm{dB}} \sim \mathcal{U}(5, 15)$ dB supplies the link-scale anchor. Let h_i^{nom} and h_i^s denote the nominal and state-s channels generated by (111). The state-dependent SNR and spectral efficiency are

$$\gamma_i^s = 10^{\gamma_{i,0}^{\mathrm{dB}}/10} \frac{|h_i^s|^2}{|h_i^{\mathrm{nom}}|^2}, \qquad \eta_i^s = \log_2(1 + \gamma_i^s). \tag{112}$$

The normalization in (112) fixes each user's sampled nominal SNR while retaining the relative service variation across RIS states. The resulting η_i^s values are the scheduling inputs.

The Wasserstein ground metric is constructed in the same reciprocal-square-root coordinates that arise in the capacity geometry. For $u_i^s = 1/\sqrt{\eta_i^s}$, define the userwise range normalization

$$\widetilde{u}_i^s = \begin{cases} \dfrac{u_i^s - \min_t u_i^t}{\max_t u_i^t - \min_t u_i^t}, & \max_t u_i^t - \min_t u_i^t > 10^{-12}, \\ 0, & \text{otherwise}, \end{cases} \qquad c_{st} = \frac{1}{\sqrt{K}} \left\| \widetilde{\mathbf{u}}^s - \widetilde{\mathbf{u}}^t \right\|_2. \tag{113}$$

Equation (113) uses the reciprocal-square-root service coordinates of the exact capacity representation, so transport is measured in coordinates directly tied to the quadratic capacity map. Userwise range normalization prevents a coordinate from dominating solely through its numerical scale, and the resulting metric is schedule-independent. Section 6.5 compares this choice with a diameter-matched RIS phase-index Hamming metric.

Table 2. System and numerical settings used in the numerical experiments.

Category	Quantity	Setting
System scale	Tasks/joint service states	$K = 20/S = 64$
	Support-rectangularization workload baseline/capacity ceiling	$B_{\text{ref}} = 15$ MHz
	Destination CPU capacities	$F_1 = 45$, $F_2 = 55$, $F_C = 150$ GHz
	Fixed non-radio latencies	$\tau_1 = 15$, $\tau_2 = 25$, $\tau_C = 80$ ms
Task generation	Input size d_i	$\mathcal{U}(0.4, 1.2)$ Mbit
	Required cycles c_i	$\mathcal{U}(0.3, 1.0)$ Gcycles
	Deadline D_i	$\mathcal{U}(0.25, 0.55)$ s
	Local CPU f_i^{L}	$\mathcal{U}(1.2, 2.4)$ GHz
	Utility weight w_i	$\{1.0, 1.5, 2.0\}$
Service-state generator	Carrier/RIS	3.5 GHz/8×8, 2-bit, $\lambda/2$ spacing
	Access point	$(20, 0, 5)$ m
	Prescribed state library	1 nominal + 63 unique perturbed states
	Perturbation per nonnominal state	7 distinct elements, each shifted by one $\pm$ phase step
	User distance/elevation	$\mathcal{U}(15, 35)$ m/$\mathcal{U}(-10^\circ, 10^\circ)$
	Nominal SNR	$\mathcal{U}(5, 15)$ dB
	Azimuth half-spread	10° clustered; 35° moderate; 60° diverse
	Transport cost	Userwise range-normalized $u_i = 1/\sqrt{\eta_i}$; Euclidean distance/$\sqrt{K}$
Reliability model	Failure tolerance	$\alpha = 0.10$
	Training observations	$n = 1000$ IID state draws
	Wasserstein radii	$\epsilon \in \{0, 0.005, 0.010, 0.020, 0.040\}$
Optimization	Support-rectangularization/frontier populations	60 instances (20 per regime)/20 instances
	Per-solve time limit	300 s
	Solver	PySCIPOpt 6.2.1/SCIP 10.0.2/SoPlex 8.0.2

The ranges in Table 2 vary task size, CPU demand, deadlines, destination resources, and user geometry while holding the capacity objective and recourse model fixed. The three spatial regimes vary the joint service-state geometry, and the 64-state libraries permit full-state checking of each selected schedule. The parameter ranges define controlled heterogeneous test instances rather than estimates of a particular deployment population. In the separate 20-instance capacity-frontier population, the mean selected task counts across the 19 point-resolved BOX cases are positive for local execution, both edge servers, and the cloud (Section 6.4), confirming that these settings exercise multiple execution tiers.

For the 60-instance support-rectangularization population, $B_{\text{ref}} = 15$ MHz defines both the workload-construction bandwidth and the capacity ceiling. The fixed target T_i is the maximum BOX workload U_2 attainable at this bandwidth; HARD and BOX then minimize capacity independently subject to $U_2(x) = T_i$. The targets range from 50 to 69. Section 6.5 repeats this construction at 12 and 18 MHz on a nine-instance subset.

The capacity variable is bounded directly by 15 MHz in both models, including every retained-state HARD master. Each BOX workload-construction schedule meets the same target within that ceiling; the fixed-schedule inclusion $B_{\text{HARD}}(x) \le B_{\text{BOX}}(x)$ therefore gives $B^\star_{\text{HARD}}(T_i) \le B^\star_{\text{BOX}}(T_i) \le 15$ MHz. All 60 stored construction schedules are feasible for both models. No final feasible capacity upper endpoint lies within 1 Hz of the ceiling (HARD: 0/60; BOX: 0/60); the largest endpoints are 13.620509 and 14.588789 MHz, respectively.

The 20-instance capacity-frontier population uses new diverse-regime instances and an uncapped minimum-capacity variable. For each instance, define

$$\mathcal{C}_i = \{x_i^{\text{NOM}}, x_i^{\text{EMP}}, x_i^{0.005}, x_i^{0.010}, x_i^{0.020}, x_i^{0.040}, x_i^{\text{HARD}}\},$$

where each element is the selected workload-maximizing schedule under the indicated uncertainty model. The fixed target is $T_i = \max_{x \in \mathcal{C}_i} U_2(x)$; subsequent capacity minimization ranges over the full set $\{x : U_2(x) = T_i\}$ rather than over $\mathcal{C}_i$. The targets range from 55 to

65 (median 60) and exceed the corresponding local-only maxima by 17–45 scaled workload units. The reference state law is

$$p_1^\star = 0.50, \qquad p_s^\star = \frac{0.50}{63}, \quad s = 2, \ldots, 64, \tag{114}$$

and $n = 1000$ IID state indices are sampled from (114) after complete generation of the service-state instance. The empirical distribution uses these counts without smoothing. The reference law assigns one half of the probability mass to the nominal state; Section 6.5 also evaluates a uniform state law. The reliability level is $\alpha = 0.10$, with Wasserstein radii $\epsilon \in \{0, 0.005, 0.010, 0.020, 0.040\}$. The prescribed radius grid is a capacity-sensitivity study and is distinct from the confidence-calibrated radius in (107).

Three $K = 8$ instances were exhaustively enumerated over all $5^8 = 390{,}625$ assignments. Their optima agree with the mixed-integer formulation, and 12 independent minimum-capacity comparisons across EMP, positive-radius DRO, HARD, and BOX agree to within a maximum absolute difference of 2.33×10^{-10} Hz. Reported intervals combine valid solver lower bounds with analytically evaluated feasible upper bounds; every upper endpoint satisfies the 64-state resource and target-risk conditions. Point resolution denotes a final gap of at most 1 Hz. The $\epsilon = 0.040$ and confidence-calibrated capacities follow from HARD through Theorem 5; Appendix A.3 gives the numerical bound construction.

Relative quantities use conservative interval arithmetic. For a nonnegative premium $P \in [\underline{P}, \overline{P}]$ and a reference capacity $C \in [\underline{C}, \overline{C}]$ with $\underline{C} > 0$,

$$\frac{P}{C} \in \left[\frac{\underline{P}}{\overline{C}}, \frac{\overline{P}}{\underline{C}}\right]. \tag{115}$$

Relative summaries average the instance-wise lower and upper ratio bounds separately.

6.2. Support Rectangularization: Capacity Cost, Geometry, and Schedule Stability

Of the 60 HARD–BOX instance pairs, 31 have both capacities point-resolved within 1 Hz. Individually, 31 HARD and 55 BOX capacities are point-resolved. The other 29 pairs have at least one unresolved optimality interval; feasible upper endpoints are available for both models in every instance.

For each instance, the optimized support-rectangularization premium is

$$\Delta B^\star = B^\star_{\mathrm{BOX}} - B^\star_{\mathrm{HARD}}, \tag{116}$$

where HARD and BOX independently minimize capacity at the same workload. Figure 2 reports all 60 deterministic intervals. Every lower endpoint is positive, with a minimum of 0.026254 MHz. The population mean satisfies

$$0.594122 \text{ MHz} \le \overline{\Delta B^\star} \le 0.926170 \text{ MHz}, \tag{117}$$

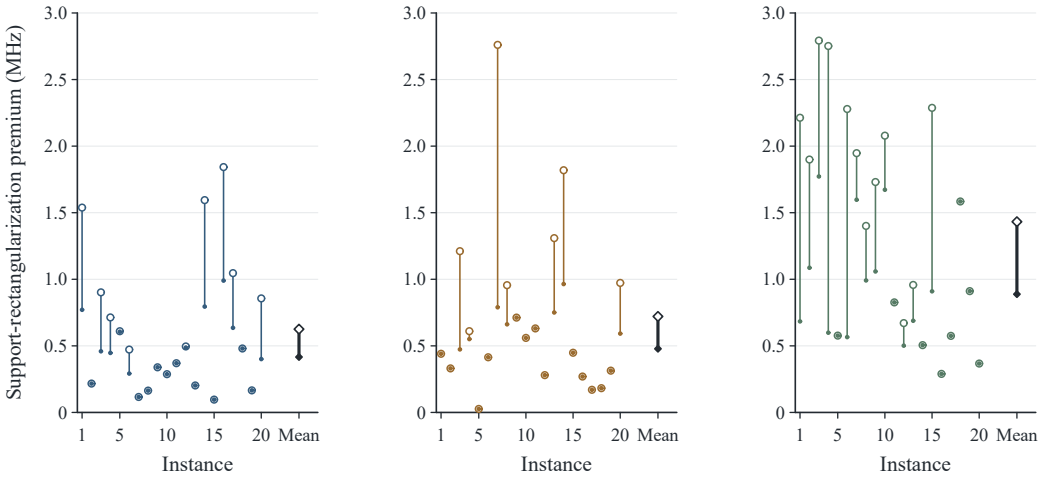
For this population, the relative quantity is the capacity saved by retaining the prescribed support instead of replacing it with the BOX model, normalized by the BOX capacity:

$$\rho_i^{\mathrm{save}} = \frac{B^\star_{\mathrm{BOX},i} - B^\star_{\mathrm{HARD},i}}{B^\star_{\mathrm{BOX},i}}. \tag{118}$$

Applying (115) instance-wise gives a mean savings interval of $[6.45\%, 9.32\%]$ relative to BOX.

The regime means are $[0.416110, 0.625446]$ MHz (clustered), $[0.478111, 0.720855]$ MHz (moderate), and $[0.888144, 1.432208]$ MHz (diverse). The difference between the diverse and

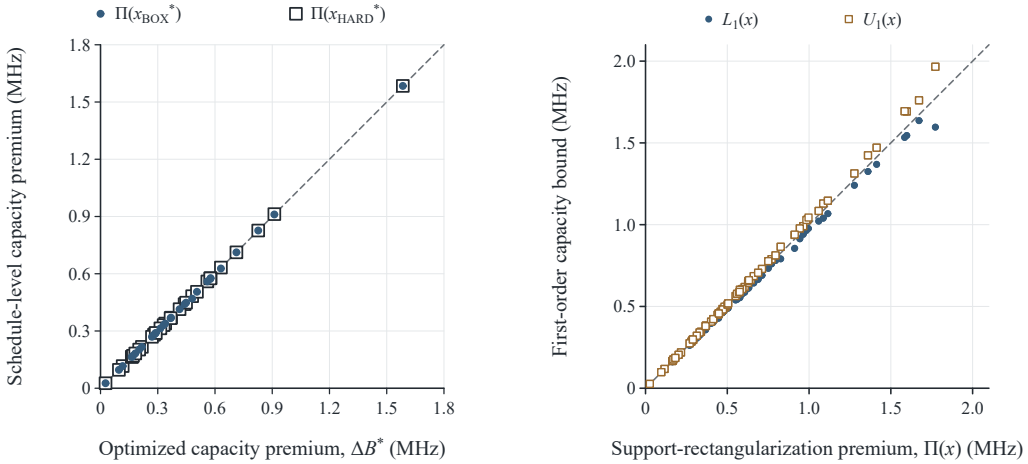
moderate means is $[0.167289, 0.954097]$ MHz, and that between the diverse and clustered means is $[0.262698, 1.016098]$ MHz. The moderate–clustered difference, $[-0.147335, 0.304745]$ MHz, overlaps zero. The premium varies across the three spatial populations and depends on the complete state geometry and Q_x.



(**a**) Clustered regime. (**b**) Moderate regime. (**c**) Diverse regime.

Figure 2. Optimized capacity cost of replacing the prescribed joint service-state support by its coordinatewise BOX model. Each panel contains 20 fixed-workload instances in original seed order. Vertical segments are deterministic lower–upper optimization intervals, and the separated diamond interval is the regime mean. Filled markers denote lower endpoints and open markers denote upper endpoints. Both uncertainty models may reoptimize the task schedule.

For the 31 jointly point-resolved pairs, Theorem 4 holds within the numerical tolerance in every case (Figure 3a). Twenty-seven optimizer pairs are cross-optimal and attain equality on both sides of the sandwich; the other four are strict. The selected HARD and BOX assignments are also identical in 27/31 pairs, with median taskwise Hamming distance zero. The capacity premium therefore often arises without a placement change.



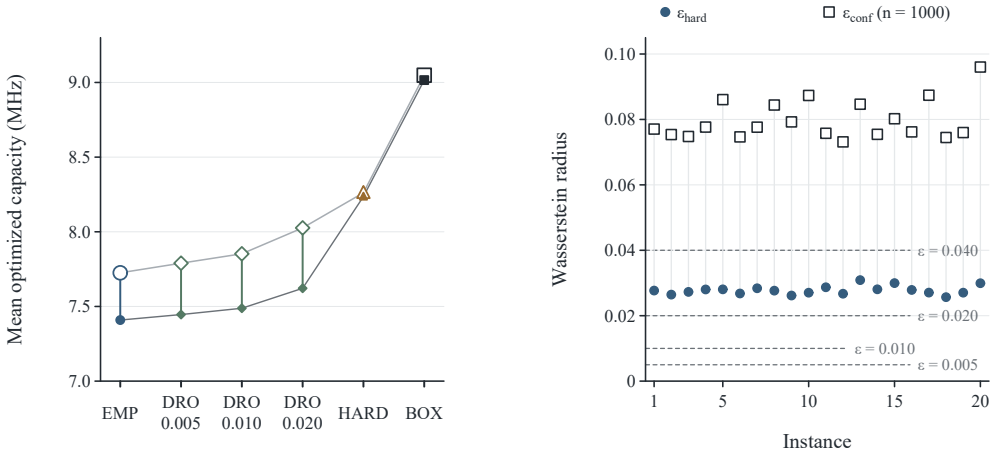
(**a**) Optimized premium sandwich. (**b**) First-order premium bounds.

Figure 3. Relation between the optimized support-rectangularization premium and the schedule-dependent capacity geometry. In (**a**), the two schedule-level premiums in Theorem 4 are plotted against the optimized premium resolved within 1 Hz for the 31 pairs; the 27 cross-optimal pairs necessarily attain equality on both sides of the sandwich. In (**b**), the exact fixed-schedule premium $\Pi(x)$ is compared with the first-order lower and upper bounds $L_1(x)$ and $U_1(x)$ for all 120 schedules evaluated under the two models.

All 120 schedules satisfy the first-order enclosure and the coarser Q_x bounds. In Figure 3b, L_1/Π has median 0.974255 and minimum 0.900803. By Corollary 1, $U_1/B_{\mathrm{BOX}} = 2\chi$, so $\Pi/U_1 = [\Pi/B_{\mathrm{BOX}}]/[2\chi]$ measures the empirical tightness of the joint-extreme upper predictor for the relative premium. Across the 120 schedules, Π/U_1 ranges from 0.901544 to 0.990979, with median 0.976240; equivalently, the exact relative premium is 90.15–99.10% of $2\chi(x)$, with median 97.62%. The corresponding $\chi(x)$ values range from 0.00776 to 0.09951 (median 0.03176), while c_x has median 1.09135 and maximum 1.35837. The distance-only lower bound has median $r_Q^2/\Pi = 0.023430$, while the geometric upper characterization has median $\Pi/[2R_Q r_Q - r_Q^2] = 0.833353$. The first-order and joint-extreme bounds retain both support-to-corner displacement and its schedule-dependent resource weighting.

6.3. Wasserstein Capacity Frontier and Hardening

Figure 4a gives the all-instance mean capacity frontier at a common workload. The EMP interval is $[7.408972, 7.725290]$ MHz; DRO at $\epsilon = 0.005$, 0.010, and 0.020 gives $[7.445766, 7.789725]$, $[7.487997, 7.852681]$, and $[7.621607, 8.026130]$ MHz. HARD requires $[8.235795, 8.261338]$ MHz and BOX $[9.015908, 9.048421]$ MHz, consistent with Proposition 3.



(**a**) Optimized capacity frontier. (**b**) Event-level hardening thresholds.

Figure 4. Distributionally robust capacity and hardening. In (**a**), filled and open markers trace the lower and upper endpoints of the all-instance mean capacity intervals from the empirical model through the rectangular worst case. Circles, diamonds, triangles, and squares represent EMP, DRO, HARD, and BOX, respectively. In (**b**), each instance's event-level hardening threshold is shown together with its own $S = 64$, $n = 1000$, $\delta = 0.05$ analytical Wasserstein confidence radius and the radii 0.005, 0.010, 0.020, and 0.040.

Table 3 summarizes the corresponding capacity increments.

Across the 20 instances, ϵ_{hard} ranges from 0.025669 to 0.030879, with median 0.027671. The radii 0.005, 0.010, and 0.020 are below every event-level threshold, while 0.040 is above every threshold. Theorem 5 therefore gives

$$B^\star_{0.040} = B^\star_{\mathrm{HARD}} \qquad \text{for all 20 instances,} \tag{119}$$

The event-level threshold is uniform over schedules; individual optimized frontiers may reach HARD at smaller radii.

For the 20 instances, the ground-metric diameter ranges from 0.475322 to 0.623998. Equation (107) with $S = 64$, $n = 1000$, and $\delta = 0.05$ gives instance-specific ϵ_{conf} values from 0.073142 to 0.096020 (median 0.077345), all above the corresponding ϵ_{hard} values in Figure 4b. With the empirical law and metric held fixed, the sample size for which the analytical radius first satisfies $\epsilon_{\mathrm{conf}} < \epsilon_{\mathrm{hard}}$ ranges from 6973 to 10436 (median 7758).

For $\alpha = 0.01$, 0.05, 0.10, and 0.20, the median hardening radii are 0.001504, 0.012213, 0.027671, and 0.061744, with corresponding median fixed-center crossover sizes of approximately 2.62×10^6, 3.98×10^4, 7758, and 1634. Because the finite-alphabet bound in (107) is distribution-free, these crossover sizes are sufficient calibration thresholds implied by the analytical containment bound. Empirical scheduling sample complexity is a separate quantity.

The adjacent probability-robustification increments have positive lower bounds in 8/20, 9/20, and 11/20 instances for EMP → 0.005, 0.005 → 0.010, and 0.010 → 0.020, respectively. The 0.020 →HARD increment is positive in 18/20 instances, while HARD → BOX is positive in all 20.

Table 3. Optimized capacity hierarchy and capacity premiums for the 20-instance fixed-workload population. Intervals are deterministic optimization bounds.

(a) Optimized capacity requirements.

Uncertainty Model	Radius ϵ	Resolved/20	Mean Capacity Interval (MHz)	
			Lower	Upper
Empirical chance	0	12	7.408972	7.725290
Wasserstein DRO	0.005	8	7.445766	7.789725
Wasserstein DRO	0.010	9	7.487997	7.852681
Wasserstein DRO	0.020	10	7.621607	8.026130
Hard prescribed support	–	19	8.235795	8.261338
Rectangular worst case	–	19	9.015908	9.048421

(b) Capacity premiums.

Transition	Mean Premium Interval (MHz)	Positive Lower Bound	Mean Relative-Premium Interval
EMP → DRO 0.005	[0.034284, 0.380753]	8/20	–
DRO 0.005 → DRO 0.010	[0.033987, 0.406915]	9/20	–
DRO 0.010 → DRO 0.020	[0.061221, 0.538133]	11/20	–
DRO 0.020 → HARD	[0.234326, 0.639732]	18/20	–
HARD → BOX	[**0.754570**, **0.812626**]	**20/20**	[**8.80**%, **9.26**%]
EMP → HARD	[0.522298, 0.852366]	19/20	[7.39%, 10.82%]
EMP → BOX	[**1.290618**, **1.639449**]	**20/20**	[**16.98**%, **20.89**%]

Resolved counts use the 1 Hz final-gap criterion. Relative premiums are instance-wise interval ratios normalized by the transition's starting model. The $\epsilon = 0.040$ and confidence-calibrated endpoints equal HARD by Theorem 5.

6.4. Decomposition of the Total Robustness Premium

The EMP-to-BOX premium separates into probability-robustification and support-rectangularization components. For instance i, define

$$A_i = B^\star_{\mathrm{HARD},i} - B^\star_{\mathrm{EMP},i}, \qquad R_i = B^\star_{\mathrm{BOX},i} - B^\star_{\mathrm{HARD},i}, \tag{120}$$

so that

$$B^\star_{\mathrm{BOX},i} - B^\star_{\mathrm{EMP},i} = A_i + R_i. \tag{121}$$

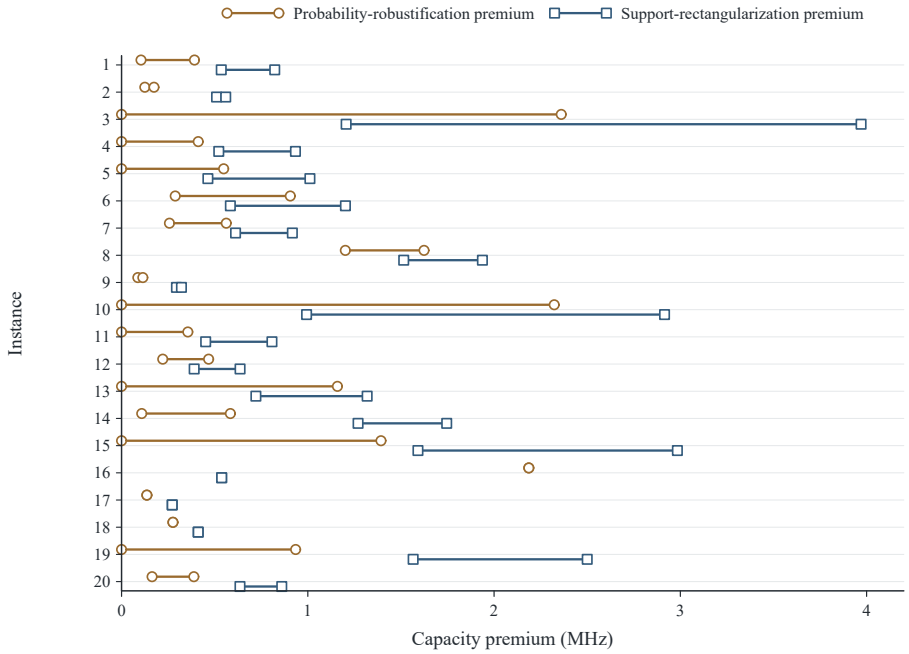
Here A_i is the EMP–HARD probability-robustification premium and R_i is the HARD–BOX support-rectangularization premium. Figure 5 reports both components and their relative share.

At the population level, the mean EMP → HARD premium is $[0.522298, 0.852366]$ MHz, and the mean HARD → BOX premium is $[0.754570, 0.812626]$ MHz. The total EMP

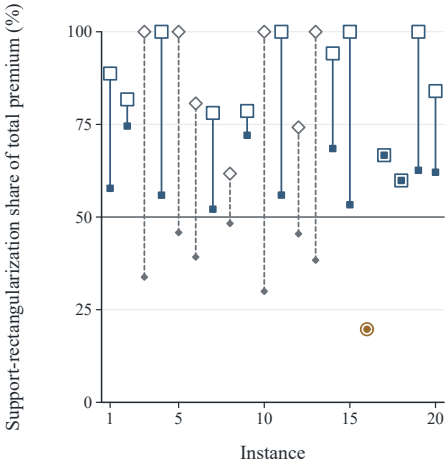
→ BOX premium is $[1.290618, 1.639449]$ MHz. The corresponding relative increases are defined by

$$\rho_i^{\mathrm{A}} = \frac{A_i}{B^{\star}_{\mathrm{EMP},i}}, \qquad \rho_i^{\mathrm{R}} = \frac{R_i}{B^{\star}_{\mathrm{HARD},i}}, \qquad \rho_i^{\mathrm{total}} = \frac{A_i + R_i}{B^{\star}_{\mathrm{EMP},i}}. \tag{122}$$

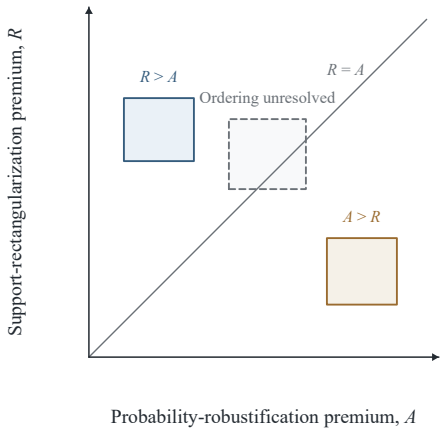
Their mean interval bounds are $[7.39\%, 10.82\%]$, $[8.80\%, 9.26\%]$, and $[16.98\%, 20.89\%]$, respectively. Each ratio is normalized by the starting capacity of its transition and averaged instance-wise.



(**a**) Instance-wise probability-robustification and support-rectangularization premiums.



(**b**) Deterministic bounds on the support-rectangularization share.



(**c**) Geometric interpretation of interval ordering.

Figure 5. Decomposition of the optimized empirical-to-BOX capacity premium. Panel (**a**) shows deterministic intervals for the probability-robustification and rectangularization premiums, A_i and R_i. Panel (**b**) shows deterministic lower and upper bounds on the rectangularization share $100R_i/(A_i + R_i)$ in percent; the horizontal 50% reference indicates equal premiums. In (**b**), solid square intervals lie above the 50% reference, dashed diamond intervals cross it, and the circled point lies below it. Panel (**c**) uses the interval Cartesian product to determine whether $R_i > A_i$, $A_i > R_i$, or neither ordering is certified.

For Figure 5b, write $P_i = A_i + R_i$ and denote the available component and total-premium intervals by $[\underline{A}_i, \overline{A}_i]$, $[\underline{R}_i, \overline{R}_i]$, and $[\underline{P}_i, \overline{P}_i]$. All 20 total-premium lower bounds are positive. The rectangularization share $\sigma_i = R_i / P_i$ is enclosed by

$$\begin{aligned} \underline{\sigma}_i &= \max\left\{ \frac{\underline{R}_i}{\overline{P}_i}, \frac{\underline{R}_i}{\underline{R}_i + \overline{A}_i} \right\}, \\ \overline{\sigma}_i &= \min\left\{ \frac{\overline{R}_i}{\underline{P}_i}, \frac{\overline{R}_i}{\overline{R}_i + \underline{A}_i} \right\}. \end{aligned} \tag{123}$$

Equation (123) combines the direct total-premium ratio bounds with the identity $P_i = A_i + R_i$; Figure 5b reports the resulting percentages.

Using the interval-ordering interpretation in Figure 5c, the component ordering is resolved in 16 of 20 instances: support rectangularization is larger in 12 and probability robustification in 4; the remaining four intervals overlap. Both components contribute to the total EMP–BOX premium in the 20-instance population.

Assignment stability is also common across the uncertainty hierarchy. Besides the 27/31 identical HARD–BOX pairs, 10 of the 12 point-resolved EMP–BOX pairs select the same assignment. The uncertainty model can therefore change the required reserve while leaving the discrete placement unchanged.

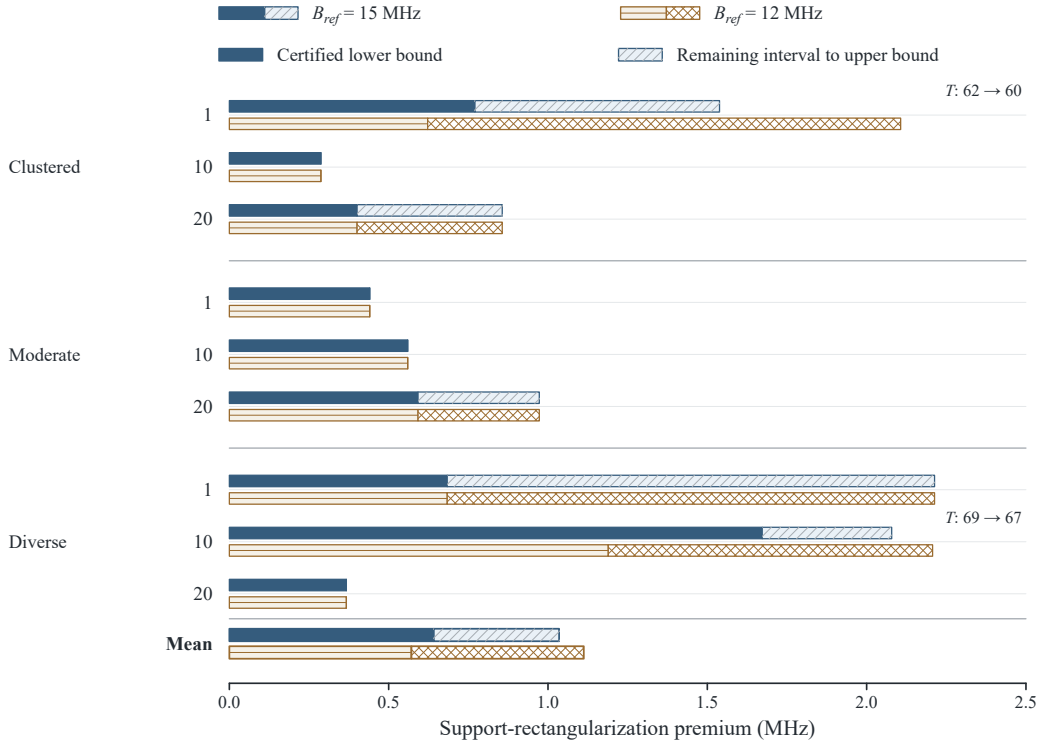
Across the 19 point-resolved BOX instances, the mean task counts are 11.84 local, 2.16 at Edge 1, 2.68 at Edge 2, and 3.32 in the cloud; all 19 use the cloud, and none rejects a task. The mix reflects the joint CPU–latency–spectrum tradeoff described by Proposition 1.

6.5. Sensitivity to Modeling Choices

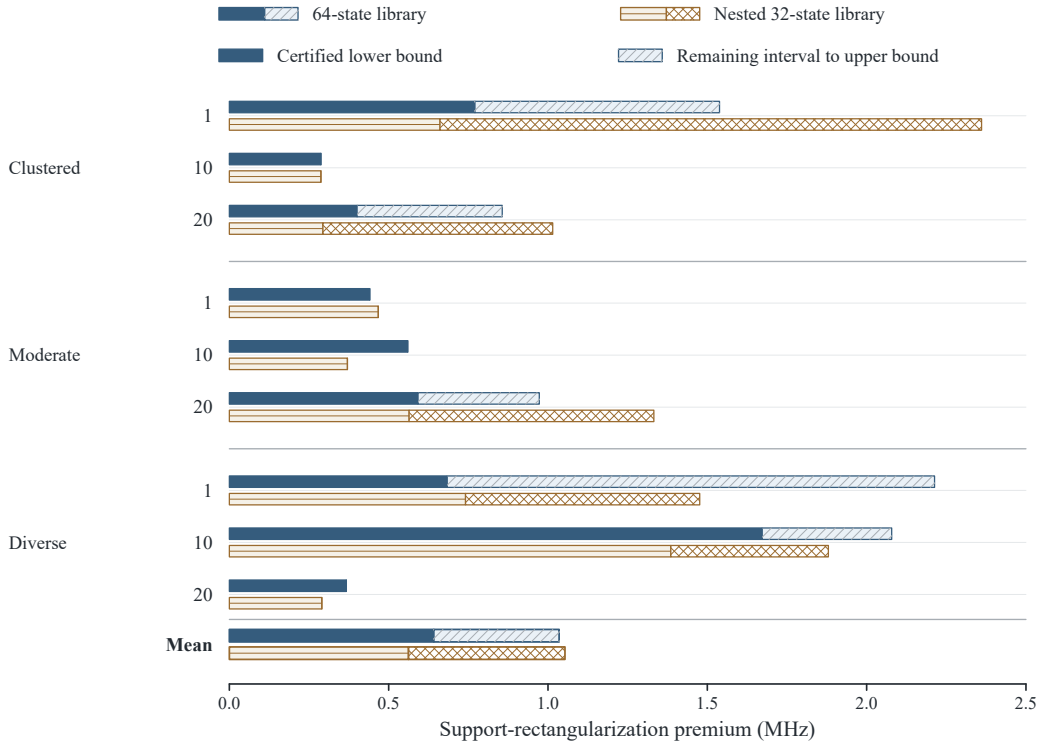
The sensitivity analysis varies one modeling choice at a time while retaining the remaining scheduling and resource parameters. The workload and support-cardinality subset uses replications 1, 10, and 20 from each spatial regime; the transport-metric, probability-law, and failure-tolerance subset uses capacity-frontier replications 1, 4, 7, 10, 13, 16, 19, and 20. Sensitivity cases requiring new optimization use a single 300 s solve per model–instance pair.

Figure 6 examines the support-rectangularization premium. At $B_{\mathrm{ref}} = 12$ MHz, all nine HARD $\to$ BOX premium intervals retain positive lower endpoints and the mean is $[0.571461, 1.112325]$ MHz, compared with $[0.641644, 1.034899]$ MHz at 15 MHz. The fixed target changes in two instances, from 62 to 60 and from 69 to 67, and the HARD $\to$ BOX lower bound remains positive in both. At $B_{\mathrm{ref}} = 18$ MHz, the fixed targets equal their 15 MHz values in all nine instances, so the downstream capacity problems are unchanged. Under the fixed nested 32-state library, the mean HARD $\to$ BOX premium is $[0.562376, 1.053269]$ MHz and all nine lower endpoints remain positive, compared with $[0.641644, 1.034899]$ MHz for the corresponding 64-state library. The 32-state comparison changes support cardinality while keeping the seven-element perturbation depth and 2-bit phase resolution fixed.

The transport metric changes the numerical calibration of Wasserstein protection. Figure 7a compares the corresponding event-level hardening thresholds. The diameter-matched phase-Hamming metric gives $\epsilon_{\mathrm{hard}} \in [0.024015, 0.030264]$ across the eight instances, compared with $[0.027034, 0.030879]$ under the normalized service metric. Six thresholds move downward and two upward, while every threshold under both metrics remains above 0.020 and below 0.040. Figure 7b represents each deterministic capacity-displacement interval $[L_\Delta, U_\Delta]$ as the point (L_Δ, U_Δ) for phase-Hamming minus normalized-service capacity. At $\epsilon = 0.005$, 2/8 intervals lie strictly below zero, and 6/8 contain zero; the corresponding counts are 4/8 and 4/8 at both $\epsilon = 0.010$ and 0.020. No interval is strictly positive. The direction remains unresolved wherever the displacement interval contains zero.

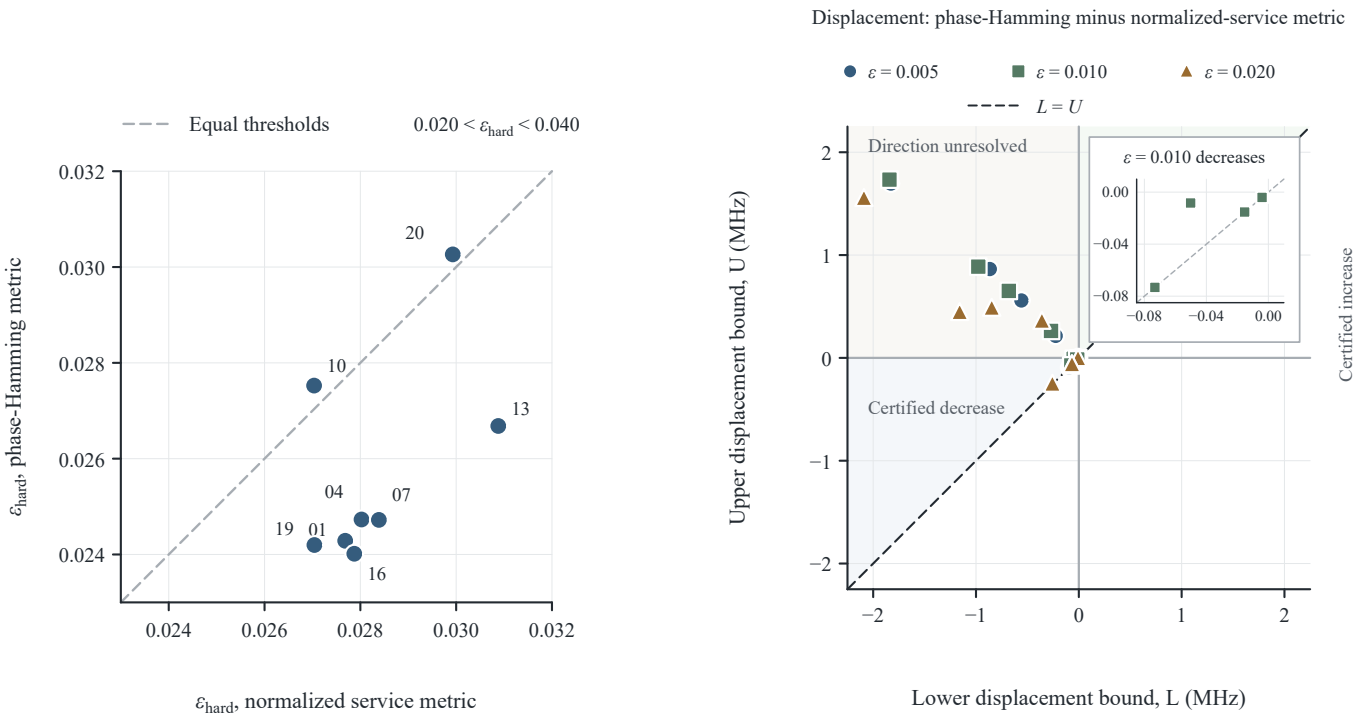


(**a**) Workload construction.



(**b**) Support cardinality.

Figure 6. Support-rectangularization premium under selected modeling variations. Each horizontal bar represents a deterministic interval $[L, U]$: the solid segment from zero to L is the certified lower bound, and the hatched continuation has length $U - L$. Panel (**a**) compares the 15 and 12 MHz workload constructions; the annotations identify the two instances whose fixed target changes. The 18 MHz construction reproduces the 15 MHz target and capacity problem in all nine cases and is omitted. Panel (**b**) compares the prescribed 64-state library with its fixed nested 32-state subset. The separated bottom bars show population means.



(**a**) Hardening-threshold comparison.

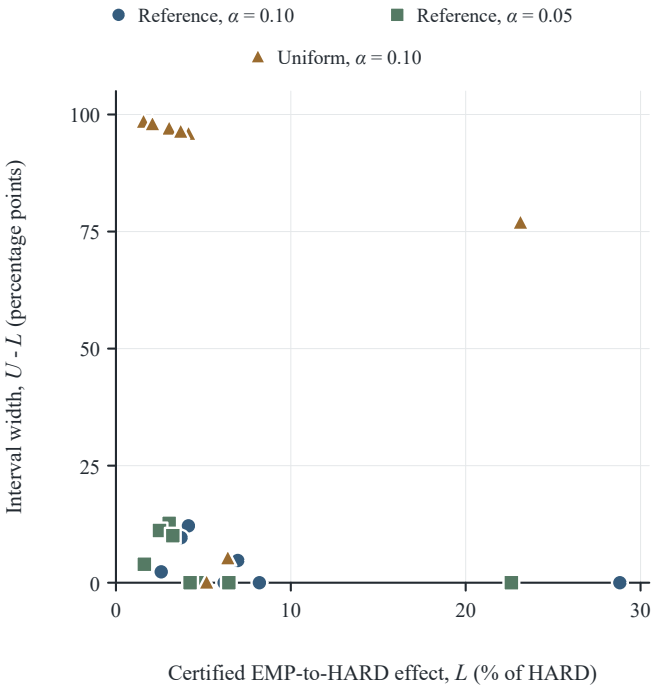
(**b**) Intermediate capacity displacement.

Figure 7. Sensitivity to the Wasserstein ground metric. Panel (**a**) compares the event-level hardening radius under the normalized service metric and the diameter-matched phase-Hamming metric; the dashed diagonal denotes equal thresholds. Panel (**b**) maps each deterministic capacity-displacement interval $[L_\Delta, U_\Delta]$ to the point (L_Δ, U_Δ). Points with $U_\Delta < 0$ establish a decrease under the phase-Hamming metric, points with $L_\Delta \leq 0 \leq U_\Delta$ leave the direction unresolved, and points with $L_\Delta > 0$ establish an increase.

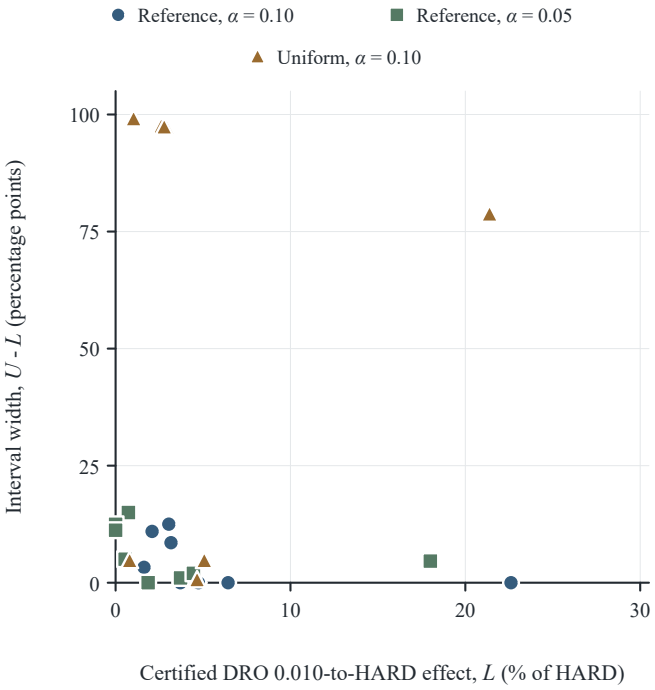
Probability-side capacity also changes with the state law and failure tolerance. Under a uniform state law with $\alpha = 0.10$, the mean EMP $\rightarrow$ HARD premium is $[0.487015, 6.987898]$ MHz, with positive lower endpoints in all eight instances; the EMP and DRO intervals are wider than under the reference law. Under the reference law with $\alpha = 0.05$, the mean EMP $\rightarrow$ HARD premium is $[0.480324, 0.967830]$ MHz and is again positive in all eight instances. Figure 8 plots the certified relative effect size on the horizontal axis and the deterministic interval width on the vertical axis. HARD capacity is used as the common denominator because six of the eight uniform-law EMP lower bounds are zero, so an EMP-normalized deterministic upper ratio is not finite in those cases. Panel (b) applies the same effect-versus-resolution display to the DRO 0.010 $\rightarrow$ HARD premium. The uniform-law cases consequently occupy larger interval-width values.

Table 4 collects the population-level sensitivity results. The two workload cases with changed fixed targets retain positive support-rectangularization lower bounds, as do all nine nested-support comparisons. The EMP $\rightarrow$ HARD lower bound is also positive under both probability-side variations, while intermediate Wasserstein capacity varies with the transport metric.

A separate pattern concerns the discrete schedule itself. Table 5 reports comparisons for which the selected assignments associated with the capacity or support models can be compared at the stated resolution. Capacity effects frequently remain positive while the selected assignment is unchanged.



(**a**) EMP-to-HARD premium.



(**b**) DRO 0.010-to-HARD premium.

Figure 8. Probability-side sensitivity. For each condition, the horizontal coordinate is the certified lower bound on the indicated premium as a percentage of HARD capacity, and the vertical coordinate is the width of the corresponding deterministic interval in percentage points. Panel (**a**) uses the EMP $\rightarrow$ HARD premium; panel (**b**) uses DRO 0.010 $\rightarrow$ HARD. The three marker types denote the reference state law at $\alpha = 0.10$, the same law at $\alpha = 0.05$, and the uniform state law at $\alpha = 0.10$.

Table 4. Sensitivity of the optimized capacity results to selected modeling choices. All ranges are deterministic optimization bounds for the displayed populations.

Choice and Variation	Cases	Deterministic Capacity Result	Additional Observation
Workload construction $B_{\mathrm{ref}} = 12$ MHz	9	HARD $\rightarrow$ BOX mean $[0.571461, 1.112325]$ MHz; positive lower bound 9/9	Fixed target changes in 2/9; the premium lower bound remains positive in both changed-target cases.
Workload construction $B_{\mathrm{ref}} = 15$ and 18 MHz	9	HARD $\rightarrow$ BOX mean $[0.641644, 1.034899]$ MHz; positive lower bound 9/9	The 18 MHz construction reproduces the 15 MHz fixed target in 9/9, so the downstream capacity problem is unchanged.
Transport metric Diameter-matched phase-Hamming	8	$\epsilon_{\mathrm{hard}} \in [0.024015, 0.030264]$; 0.040 remains beyond hardening in 8/8	Mean phase-Hamming minus normalized-service capacity at $\epsilon = 0.010$ is $[-0.489182, 0.429189]$ MHz.
State law Uniform, $\alpha = 0.10$	8	EMP $\rightarrow$ HARD mean $[0.487015, 6.987898]$ MHz; positive lower bound 8/8	Several EMP/DRO endpoints remain wide, limiting fine-grained magnitude comparison.
Failure tolerance Reference law, $\alpha = 0.05$	8	EMP $\rightarrow$ HARD mean $[0.480324, 0.967830]$ MHz; positive lower bound 8/8	The comparison changes only α from the reference value 0.10.
Support cardinality Fixed nested 32-state library	9	HARD $\rightarrow$ BOX mean $[0.562376, 1.053269]$ MHz; positive lower bound 9/9	The corresponding 64-state mean is $[0.641644, 1.034899]$ MHz.

Table 5. Capacity effects and selected-assignment stability across the comparisons. Assignment counts use selected witnesses only when the corresponding endpoints are jointly resolved; multiple optimal assignments may exist.

Comparison	Resolved Pairs	Capacity Result	Identical Selected Assignments	Placement Evidence
Main-study HARD → BOX	31 point-resolved	Positive lower bound 31/31; all-60 mean $[0.594122, 0.926170]$ MHz	27/31	Hamming median 0 (max 2); 27/31 pairs cross-optimal in both directions.
Frontier HARD → BOX	19	All-20 mean $[0.754570, 0.812626]$ MHz; positive lower bound 20/20	15/19	Hamming median 0 (max 3).
Frontier EMP → BOX	12	All-20 mean $[1.290618, 1.639449]$ MHz; positive lower bound 20/20	10/12	Hamming median 0 (max 3).
HARD, 64 vs. 32 states	4	32-state HARD → BOX lower bound positive 9/9	4/4	Hamming distance 0 in every reportable comparison.
BOX, 64 vs. 32 states	8	32-state HARD → BOX lower bound positive 9/9	8/8	Hamming distance 0 in every reportable comparison.

7. Discussion

Uncertainty representation is a capacity-planning choice in the two-stage edge–cloud model. HARD protects the jointly specified service states in the prescribed library, whereas BOX protects their coordinatewise envelope. The optimized sandwich in Theorem 4 carries this distinction through schedule reoptimization. The HARD–BOX difference therefore measures the capacity effect of support enlargement after schedule reoptimization.

The quadratic map $B_{\mathrm{req}}(x;\mathbf{u}) = \mathbf{u}^{\mathsf{T}} Q_x \mathbf{u}$ identifies how that effect depends on the schedule and resource state. Corollary 1 separates normalized failure to jointly attain marginal worst conditions from the schedule-dependent weighting $\omega_i(x)$. Across the 120 HARD/BOX schedules, the exact relative premium is 90.15–99.10% of the joint-extreme predictor $2\chi(x)$, with median 97.62%. Corollary 1 provides the corresponding general two-sided bound.

The capacity effect is also distinct from placement sensitivity. Most resolved HARD–BOX and EMP–BOX comparisons retain the same selected assignment, and the 64-versus-32-state comparisons retain identical selected assignments wherever both endpoints are resolved. Thus, uncertainty representation can change the required reserve without changing the selected assignment.

The EMP–HARD–BOX decomposition separates probability-side robustification from support enlargement. Proposition 2 gives the structural distinction: $\mathfrak{D}_\epsilon$ determines which failure sets are admissible under the probability model, while $b_s(x)$ determines the capacity required by the retained states. A schedule remains below HARD exactly when its complete set of worst-capacity states belongs to $\mathfrak{D}_\epsilon$; hardening occurs once that set is no longer admissible. With the diameter-matched phase-Hamming metric, ϵ_{hard} shifts but remains between 0.020 and 0.040 in all eight instances, as it does under the normalized service metric. The intermediate capacity intervals also change with the metric.

The EMP → HARD lower bound remains positive in all eight instances under both the uniform state law at $\alpha = 0.10$ and the reference law at $\alpha = 0.05$. The uniform-law EMP and DRO intervals are wider, so the corresponding premium magnitudes are less

tightly localized. Probability-side capacity therefore varies with both the state law and failure tolerance.

Computational effort depends on the assignment and state dimensions. With K tasks and J remote destinations, the assignment layer contains $K(J+2)$ binary variables and admits up to $(J+2)^K$ assignments before feasibility restrictions. The statewise conic recourse contains terms indexed by destination and protected state, while the positive-radius Wasserstein formulation adds S failure indicators and S^2 transport-dual inequalities. For $K = 20$, $J = 3$, and $S = 64$, the pre-presolve BOX, HARD, and positive-radius DRO formulations contain, respectively, 107 variables and 28 constraints, 296 and 280, and 489 and 4441 (Table A1). Retained-state refinement increased HARD point resolution from 3/20 to 19/20 in the frontier study; positive-radius DRO was point-resolved in 8–10 of 20 instances under the reported computational limits. The recourse formula enables direct capacity evaluation for a fixed assignment, while global capacity minimization also requires resolving the combinatorial assignment decisions and statewise constraints.

The normalized service metric measures transport in the inverse-square-root efficiency coordinates of the quadratic capacity map; the diameter-matched phase-Hamming metric instead measures RIS phase-index changes. The metric variation changes intermediate DRO capacity intervals and hardening thresholds in the evaluated cases. For IID observations on the prescribed finite support and a metric fixed independently of the observations, (107) supplies a sufficient $(1-\delta)$ Wasserstein containment radius. At $n = 1000$ and $\delta = 0.05$, the calculated radii 0.073142–0.096020 exceed the corresponding event-level hardening thresholds for all 20 frontier instances, so the confidence-calibrated capacities equal HARD. The smaller radii used for the plotted frontiers quantify sensitivity rather than the same finite-sample confidence guarantee. Held-out service observations can guide operational radius selection, with statistical coverage assessed under the assumptions of the chosen calibration procedure.

The RIS generator supplies correlated state variation. The adopted rate is proportional to bandwidth under fixed transmit power spectral density. With fixed total transmit power, the rate $b\log_2(1+Pg/(N_0 b))$ defines a different recourse model. The containment result concerns IID observations and the prescribed support; dependence, nonstationarity, or unrepresented states require corresponding changes to the uncertainty model.

8. Conclusions

Exact post-state bandwidth–CPU recourse gives a quadratic capacity map for each fixed schedule. This map yields the support-rectangularization premium, resource-weighted joint-extreme bounds, and an optimized HARD–BOX sandwich under schedule reoptimization. The finite-state Wasserstein analysis characterizes admissible failure sets, the EMP–DRO–HARD–BOX capacity hierarchy, and the event-level radius at which chance-constrained capacity reaches HARD.

All 60 support-rectangularization instances retain positive certified premium lower bounds, with mean BOX-normalized capacity savings enclosed by 6.45–9.32%. In the 20-instance frontier, probability protection and support enlargement contribute separately to minimum capacity, while the hardening threshold and intermediate DRO capacities depend on the chosen transport metric. The workload, support, probability-law, and tolerance sensitivity cases preserve the stated positive-premium findings within their evaluated populations. These results quantify uncertainty-representation costs for offline capacity planning over prescribed finite service-state libraries. The principal numerical studies use 20 tasks and 64 prescribed service states; the finite-sample Wasserstein containment guarantee applies to IID observations on that support with a ground metric fixed independently of the observations.

Author Contributions: Conceptualization, M.M.S.; methodology, M.M.S.; software, M.M.S.; validation, M.M.S. and F.N.; formal analysis, M.M.S.; investigation, M.M.S.; data curation, M.M.S.; visualization, M.M.S.; writing—original draft preparation, M.M.S. and F.N.; writing—review and editing, M.M.S. and F.N.; project administration, M.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and code supporting the findings of this study are publicly available through Zenodo, https://doi.org/10.5281/zenodo.22996277.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Mixed-Integer Capacity Formulation and Numerical Solution

This appendix specifies the scaled mixed-integer formulations and the deterministic-bound construction used in the minimum-capacity calculations. The task, state, and workload definitions are those of Section 3; Theorem 1 eliminates the continuous per-task bandwidth–CPU allocation before optimization.

Appendix A.1. Scaled Recourse and HARD Models

Let $B_{\mathrm{u}} = 10^6$ Hz and $F_{\mathrm{u}} = 10^9$ cycles/s be the bandwidth and CPU units, and write $\beta = B/B_{\mathrm{u}}$. For a modeled service vector $\boldsymbol{\eta}^s$, define

$$a_{js}(x) = \sum_{i:L_{ij}>0} \frac{d_i}{B_{\mathrm{u}}\eta_i^s L_{ij}} x_{ij}, \tag{A1}$$

$$g_{js}(x) = \sum_{i:L_{ij}>0} \frac{\sqrt{d_i c_i/\eta_i^s}}{\sqrt{B_{\mathrm{u}}F_{\mathrm{u}}}L_{ij}} x_{ij}. \tag{A2}$$

The destination-headroom variable v_j satisfies

$$v_j = \frac{F_j}{F_{\mathrm{u}}} - \sum_{i:L_{ij}>0} \frac{c_i}{F_{\mathrm{u}}L_{ij}} x_{ij}, \qquad 0 \le v_j \le \frac{F_j}{F_{\mathrm{u}}}. \tag{A3}$$

Introduce $t_{js} \ge 0$ and impose

$$g_{js}(x)^2 \le t_{js} v_j. \tag{A4}$$

For $v_j > 0$, the minimum feasible t_{js} is $g_{js}(x)^2/v_j$, so $a_{js}(x) + t_{js}$ equals the destination bandwidth in MHz. Equation (A4) is imposed as a quadratic constraint; equivalently,

$$\left\| \begin{pmatrix} 2g_{js}(x) \\ t_{js} - v_j \end{pmatrix} \right\|_2 \le t_{js} + v_j. \tag{A5}$$

For an active destination, positivity of d_i, c_i, and η_i^s implies $g_{js}(x) > 0$ and hence $v_j > 0$. For an inactive destination, $g_{js} = a_{js} = 0$ and $t_{js} = 0$ is feasible.

For a retained state index set $\mathcal{A}$, the HARD capacity model is

$$\begin{aligned}
\min_{x,\beta,v,t} \quad & \beta \\
\text{s.t.} \quad & x \in \mathcal{X}_T, \qquad \beta \geq 0, \\
& \text{(A3)}, \qquad\qquad\qquad\qquad j \in \mathcal{J}, \\
& g_{js}(x)^2 \leq t_{js} v_j, \quad t_{js} \geq 0, \quad j \in \mathcal{J},\ s \in \mathcal{A}, \\
& \sum_{j \in \mathcal{J}} \left(a_{js}(x) + t_{js}\right) \leq \beta, \qquad s \in \mathcal{A}.
\end{aligned} \tag{A6}$$

Here $x \in \mathcal{X}_T$ contains the binary assignment variables, the assignment equality (4), local and offloading admissibility, and the exact workload equality $\sum_i (2w_i)(1 - r_i) = T$. The five choices are local, edge 1, edge 2, cloud, and reject; inadmissible choices have variable upper bound zero.

For HARD, $\mathcal{A} = \mathcal{S}$. The nominal model retains only the commanded state, while BOX uses the single service vector $\eta_i^{\mathrm{BOX}} = (\bar{u}_i)^{-2}$. In the 20-instance capacity study, β is unbounded above. The 60-instance support-rectangularization study uses $0 \leq \beta \leq 15$ because its fixed workload is constructed from a BOX-feasible schedule at the 15 MHz reference bandwidth. The cap is the capacity-variable upper bound in both BOX and every retained-state HARD master, not a post-processing truncation.

Appendix A.2. Empirical and Wasserstein Reliability

For empirical and distributionally robust capacity, introduce $y_s \in \{0, 1\}$ for each prescribed library state. Setting $y_s = 1$ relaxes the statewise bandwidth requirement, while $y_s = 0$ activates

$$y_s = 0 \quad \Longrightarrow \quad \sum_{j \in \mathcal{J}} \left(a_{js}(x) + t_{js}\right) \leq \beta, \qquad s \in \mathcal{S}. \tag{A7}$$

Equation (A7) is implemented with native indicator constraints.

For the selected state set, the finite transport problem is

$$\max_{\pi \geq 0} \left\{ \sum_{t,s \in \mathcal{S}} \pi_{ts} y_s : \sum_s \pi_{ts} = \widehat{p}_t,\ \sum_{t,s} c_{ts} \pi_{ts} \leq \epsilon \right\}. \tag{A8}$$

Its dual introduces $\lambda \geq 0$ and $\mu_t \geq 0$:

$$\mu_t \geq y_s - \lambda c_{ts}, \qquad t, s \in \mathcal{S}, \tag{A9}$$

$$\epsilon\lambda + \sum_{t \in \mathcal{S}} \widehat{p}_t \mu_t \leq \alpha. \tag{A10}$$

Because $c_{tt} = 0$, the diagonal inequalities imply $\mu_t \geq y_t \geq 0$. The positive-radius DRO model minimizes β over $x \in \mathcal{X}_T$, v_j, t_{js}, y_s, λ, and μ_t subject to (A3) and (A4), and (A7)–(A10); β, t_{js}, λ, and μ_t are unbounded above.

At $\epsilon = 0$, the empirical model is equivalently represented without the transport dual by the exact constraint

$$\sum_{s \in \mathcal{S}} \widehat{p}_s y_s \leq \alpha. \tag{A11}$$

Both forms represent the same empirical chance constraint.

The formulation is equivalent to the reliability condition on the actual failure set. The indicators give $\mathcal{F}_B(x) \subseteq \{s : y_s = 1\}$, and (A9) and (A10) bound the worst-case probability of that set. Conversely, any reliability-feasible (x, B) admits $y_s = \mathbf{1}_{\{s \in \mathcal{F}_B(x)\}}$ together with feasible recourse epigraphs and transport-dual variables. Since $\alpha < 1$, at least one state remains protected, so positive CPU headroom is required at every active remote destination.

For $K = 20$, $|\mathcal{J}| = 3$, and $S = 64$, the model sizes before presolve are summarized in Table A1. The positive-radius DRO count includes the native indicator representation.

Table A1. Formulation size and computational summary for the 20-instance capacity study. Model sizes are before presolve; mean times refer to the initial 300 s solves.

Model	Variables	Constraints	Mean Time (s)	Resolved/20, Initial → Final
Nominal	107	28	24.50	19 → 19
EMP (transport dual)	489	4441	194.25	12 → 12
DRO 0.005	489	4441	246.05	8 → 8
DRO 0.010	489	4441	230.55	9 → 9
DRO 0.020	489	4441	233.95	10 → 10
HARD (all states)	296	280	285.00	3 → 19
BOX	107	28	24.65	19 → 19

The initial EMP solves used the transport-dual formulation in (A9) and (A10). The direct empirical form in (A11), used for the additional EMP solves, has 424 variables and 345 constraints before presolve.

Appendix A.3. Numerical Solution and Deterministic Bounds

Minimum-capacity solves use a 300 s time limit. Feasible schedules available for the same instance are evaluated analytically under the target uncertainty model and may initialize the solver. Their continuous variables are reconstructed from (A3) and (A4). The final uncapped EMP/DRO refinement adds 0.01 Hz only to the submitted bandwidth warm start, not to the reported capacity bounds. The capped 60-instance HARD/BOX implementation submits the reconstructed capacity without this increment. For a nonempty excluded set $\mathcal{F}$, a feasible transport-dual initialization is obtained from

$$\lambda \in \{0\} \cup \{d(t,\mathcal{F})^{-1} : d(t,\mathcal{F}) > 0\}, \qquad \mu_t = \big(1 - \lambda d(t,\mathcal{F})\big)_+. \tag{A12}$$

The empty set uses $\lambda = 0$ and $\mu_t = 0$. Default solver settings are retained apart from the time limit; minimum bandwidth is the sole objective.

For every integer schedule used to form an upper bound, capacity is reconstructed from the original resource expressions. In EMP and DRO, the distinct statewise capacity levels are examined in increasing order under the strict failure set $\{s : b_s(x) > B\}$; the first level with risk at most α is the schedule capacity. Probability comparisons use tolerances of 10^{-12} in level selection and 10^{-9} in the final feasibility check. Local-deadline admissibility uses 10^{-10} s and failure-index reconstruction uses a 10^{-7} Hz comparison guard. The mathematical failure condition remains the strict inequality in (14).

Let $L_{i,m}^{\mathrm{own}}$ be the nonnegative solver lower bound in Hz for instance i and model m, and let $\mathcal{W}_i$ denote the available feasible schedules. Analytic evaluation gives

$$U_{i,m} = \min_{x \in \mathcal{W}_i} B_m(x). \tag{A13}$$

For the ordered chain EMP, DRO 0.005, DRO 0.010, DRO 0.020, HARD, BOX, valid lower bounds propagate as

$$L_{i,m}^{\mathrm{prop}} = \max_{k \preceq m} L_{i,k}^{\mathrm{own}}. \tag{A14}$$

The nominal model remains outside this chain. The reported lower endpoint is $L_{i,m} = \min\{L_{i,m}^{\mathrm{prop}}, U_{i,m}\}$ when a numerical reversal is at most 1 Hz; larger reversals are treated as invalid. An interval is point-resolved when $U_{i,m} - L_{i,m} \le 1$ Hz. For two ordered models $a \preceq b$, the capacity-premium interval is

$$[\max\{0, L_{i,b} - U_{i,a}\},\ \max\{0, U_{i,b} - L_{i,a}\}]. \tag{A15}$$

Population summaries average the instance-wise endpoints. By Theorem 5, the $\epsilon = 0.040$ and confidence-calibrated capacities equal HARD and therefore inherit its deterministic interval directly.

For HARD cases with a gap above 1 Hz, a retained-state master supplies a valid lower bound by imposing capacity constraints on a subset of prescribed states. Each candidate schedule is evaluated over all 64 prescribed states, and any omitted state whose requirement exceeds the master capacity is added. The same finite-state procedure is used in the 60-instance support-rectangularization study and in the 20-instance frontier.

For EMP and positive-radius DRO cases with a gap above 1 Hz, one further 300 s solve is performed with the exact formulation. Positive-radius DRO additionally uses the valid inequalities

$$\sum_s \widehat{p}_s y_s \le \alpha, \qquad y_s = 0 \text{ if } R_{\mathrm{wc}}(\{s\};\epsilon) > \alpha, \qquad y_s + y_t \le 1 \text{ if } R_{\mathrm{wc}}(\{s,t\};\epsilon) > \alpha. \tag{A16}$$

Each reported upper endpoint is verified by full-state resource evaluation and exact target-risk evaluation.

The HARD retained-state procedure required 17 master solves and 2223 s of SCIP time, closing 16 of the 17 cases that remained above the 1 Hz criterion after the all-state solves. For EMP and DRO, the 41 model–instance pairs with gaps above 1 Hz each underwent one further 300 s solve under the formulations above; the deterministic intervals tightened while the point-resolved counts remained those reported in Table A1.

References

1. Mach, P.; Becvar, Z. Mobile Edge Computing: A Survey on Architecture and Computation Offloading. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1628–1656. [CrossRef]
2. Mao, Y.; You, C.; Zhang, J.; Huang, K.; Letaief, K.B. A Survey on Mobile Edge Computing: The Communication Perspective. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 2322–2358. [CrossRef]
3. Hong, C.H.; Varghese, B. Resource Management in Fog/Edge Computing: A Survey on Architectures, Infrastructure, and Algorithms. *ACM Comput. Surv.* **2019**, *52*, 97. [CrossRef]
4. Luo, Q.; Hu, S.; Li, C.; Li, G.; Shi, W. Resource Scheduling in Edge Computing: A Survey. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 2131–2165. [CrossRef]
5. Sardellitti, S.; Scutari, G.; Barbarossa, S. Joint Optimization of Radio and Computational Resources for Multicell Mobile-Edge Computing. *IEEE Trans. Signal Inf. Process. Netw.* **2015**, *1*, 89–103. [CrossRef]
6. You, C.; Huang, K.; Chae, H.; Kim, B.H. Energy-Efficient Resource Allocation for Mobile-Edge Computation Offloading. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 1397–1411. [CrossRef]
7. Mohajerin Esfahani, P.; Kuhn, D. Data-Driven Distributionally Robust Optimization Using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations. *Math. Program.* **2018**, *171*, 115–166. [CrossRef]
8. Bertsimas, D.; Sim, M. The Price of Robustness. *Oper. Res.* **2004**, *52*, 35–53. [CrossRef]
9. Yuan, Y.; Li, Z.; Huang, B. Robust Optimization under Correlated Uncertainty: Formulations and Computational Study. *Comput. Chem. Eng.* **2016**, *85*, 58–71. [CrossRef]
10. Nguyen, T.T.; Le, L.B.; Le-Trung, Q. Computation Offloading in MIMO Based Mobile Edge Computing Systems Under Perfect and Imperfect CSI Estimation. *IEEE Trans. Serv. Comput.* **2021**, *14*, 2011–2025. [CrossRef]
11. Qu, Y.; Dai, H.; Wu, F.; Lu, D.; Dong, C.; Tang, S.; Chen, G. Robust Offloading Scheduling for Mobile Edge Computing. *IEEE Trans. Mob. Comput.* **2022**, *21*, 2581–2595. [CrossRef]
12. Ji, T.; Luo, C.; Yu, L.; Wang, Q.; Chen, S.; Thapa, A.; Li, P. Energy-Efficient Computation Offloading in Mobile Edge Computing Systems With Uncertainties. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 5717–5729. [CrossRef]
13. Li, S.; Li, C.; Huang, Y.; Jalaian, B.A.; Hou, Y.T.; Lou, W. Enhancing Resilience in Mobile Edge Computing Under Processing Uncertainty. *IEEE J. Sel. Areas Commun.* **2023**, *41*, 659–674. [CrossRef]
14. Fan, R.; Liang, B.; Zuo, S.; Hu, H.; Jiang, H.; Tao, X. Robust Task Offloading and Resource Allocation in Mobile Edge Computing With Uncertain Distribution of Computation Burden. *IEEE Trans. Commun.* **2023**, *71*, 4283–4299. [CrossRef]

15. Hao, Y.; Wang, J.; Huo, D.; Guizani, N.; Hu, L.; Chen, M. Digital Twin-Assisted URLLC-Enabled Task Offloading in Mobile Edge Network via Robust Combinatorial Optimization. *IEEE J. Sel. Areas Commun.* **2023**, *41*, 3022–3033. [CrossRef]
16. Deng, T.; Yu, Z.; Yuan, D. Task Offloading Optimization in Mobile Edge Computing under Uncertain Processing Cycles and Intermittent Communications. *Comput. Netw.* **2024**, *245*, 110359. [CrossRef]
17. Nan, Z.; Han, Y.; Yan, J.; Zhou, S.; Niu, Z. Robust Task Offloading and Resource Allocation Under Imperfect Computing Capacity Information in Edge Intelligence Systems. *IEEE Trans. Mob. Comput.* **2025**, *24*, 6154–6167. [CrossRef]
18. Xu, B.; Bian, H.; Cui, Q.; Yu, X.; Qi, J.; Ji, Y. Distributed Optimization of Task Offloading and Resource Allocation for Mobile Edge Computing With Multifactorial Uncertainty. *IEEE Trans. Mob. Comput.* **2026**, *25*, 1644–1659. [CrossRef]
19. Bandyopadhyay, B.; Kuila, P.; Govil, M.C.; Bey, M. Delay-Sensitive Task Offloading and Efficient Resource Allocation in Intelligent Edge–Cloud Environments: A Discretized Differential Evolution-Based Approach. *Appl. Soft Comput.* **2024**, *159*, 111637. [CrossRef]
20. Li, Y.; Liu, J.; Cao, B.; Wang, C. Joint Optimization of Radio and Virtual Machine Resources With Uncertain User Demands in Mobile Cloud Computing. *IEEE Trans. Multimed.* **2018**, *20*, 2427–2438. [CrossRef]
21. Nguyen, D.T.; Nguyen, H.T.; Trieu, N.; Bhargava, V.K. Two-Stage Robust Edge Service Placement and Sizing under Demand Uncertainty. *IEEE Internet Things J.* **2022**, *9*, 1560–1574. [CrossRef]
22. Qu, Y.; Wang, L.; Dai, H.; Wang, W.; Dong, C.; Wu, F.; Guo, S. Server Placement for Edge Computing: A Robust Submodular Maximization Approach. *IEEE Trans. Mob. Comput.* **2023**, *22*, 3634–3649. [CrossRef]
23. Cheng, J.; Nguyen, D.T.; Bhargava, V.K. Resilient Edge Service Placement Under Demand and Node Failure Uncertainties. *IEEE Trans. Netw. Serv. Manag.* **2024**, *21*, 558–573. [CrossRef]
24. Jeong, J.; Premsankar, G.; Ghaddar, B.; Tarkoma, S. A Robust Optimization Approach for Placement of Applications in Edge Computing Considering Latency Uncertainty. *Omega* **2024**, *126*, 103064. [CrossRef]
25. Cheng, J.; Nguyen, D.T.A.; Nguyen, D.T. Robust Dynamic Edge Service Placement Under Spatio-Temporal Correlated Demand Uncertainty. *IEEE Trans. Serv. Comput.* **2025**, *18*, 2372–2387. [CrossRef]
26. Bai, Z.; Li, J.; Shen, H. A Robust Two-Stage Edge Cloud Resource Allocation under Edge–Edge Collaboration. *Eur. J. Oper. Res.* **2026**, *335*, 947–965. [CrossRef]
27. Yang, M.; Wang, J.; Chen, Y.; Zeng, Y.; Su, X. Data-Driven Robust Optimization Scheduling for Microgrid Day-Ahead to Intra-Day Operations Based on Renewable Energy Interval Prediction. *Energy* **2024**, *313*, 134058. [CrossRef]
28. Bertsimas, D.; Brown, D.B.; Caramanis, C. Theory and Applications of Robust Optimization. *SIAM Rev.* **2011**, *53*, 464–501. [CrossRef]
29. Xie, W. On Distributionally Robust Chance Constrained Programs with Wasserstein Distance. *Math. Program.* **2021**, *186*, 115–155. [CrossRef]
30. Chen, Z.; Kuhn, D.; Wiesemann, W. Technical Note—Data-Driven Chance Constrained Programs over Wasserstein Balls. *Oper. Res.* **2024**, *72*, 410–424. [CrossRef]
31. Ling, Z.; Hu, F.; Zhang, Y.; Fan, L.; Gao, F.; Han, Z. Distributionally Robust Chance-Constrained Backscatter Communication-Assisted Computation Offloading in WBANs. *IEEE Trans. Commun.* **2021**, *69*, 3395–3408. [CrossRef]
32. Tsai, K.C.; Fan, L.; Lent, R.; Wang, L.C.; Han, Z. Distributionally Robust Optimal Routing for Integrated Satellite–Terrestrial Networks Under Uncertainty. *IEEE Trans. Commun.* **2024**, *72*, 6401–6415. [CrossRef]
33. Kleinrock, L. *Communication Nets: Stochastic Message Flow and Delay*; McGraw-Hill: New York, NY, USA, 1964.
34. Wein, L.M. Capacity Allocation in Generalized Jackson Networks. *Oper. Res. Lett.* **1989**, *8*, 143–146. [CrossRef]
35. Weissman, T.; Ordentlich, E.; Seroussi, G.; Verdú, S.; Weinberger, M.J. *Inequalities for the L_1 Deviation of the Empirical Distribution*; Technical Report HPL-2003-97 (R.1); Hewlett-Packard Laboratories: Palo Alto, CA, USA, 2003.
36. Basar, E.; Di Renzo, M.; de Rosny, J.; Debbah, M.; Alouini, M.S.; Zhang, R. Wireless Communications Through Reconfigurable Intelligent Surfaces. *IEEE Access* **2019**, *7*, 116753–116773. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.