Certified-Gap Dual-Price Policies for Real-Time Truckload Bid Acceptance with Relocating, Clock-Constrained Resources*

Aswin Chandrasekaran
Bubba AI
aswin@bubba.ai

July 2026

Abstract

A truckload carrier must accept or reject each load tender within seconds. The decision depends on fleet state, hours-of-service (HOS) clocks, and appointment windows. We model this as a weakly coupled dynamic program in which the resources relocate and carry clocks: serving a request moves the truck to a new market and depletes its clocks, and whether a truck can serve a request depends on its state. Occupancy-based reusable-resource models do not cover this setting.

We build a real-time dual-price policy from the same Lagrangian relaxation that gives the problem's upper bound. Policy and bound come from one object, so every run reports a *certified optimality gap*. We prove three things. First, the certificate is valid for any duals, any discretization, and any surrogate quality (Theorem 1). Second, the policy's same-time spatial-gradient rule is exactly fluid complementary slackness, and the policy is asymptotically optimal in the subcritical fluid regime (Theorem 2); the fitted prices are also portable across sample paths, by linear-programming basis stability. Third, certificates have limits: per-resource Lagrangian slack can stay bounded away from zero at every fleet size. We exhibit a three-truck kernel with an exact rational certificate and a replication lemma (Theorem 3).

On a public closed-loop benchmark with thirty paired seeds, the policy — which needs no rollout labels, only one offline dual solve — beats a rollout-trained surrogate on two of three scenarios (tight: +2.0 pp, 95% CI $[+0.5, +3.6]$, Wilcoxon $p = 0.023$; mild: +3.5 pp, CI $[+2.4, +4.5]$) and ties the third. It decides in 0.04–0.09 ms, three orders of magnitude faster than the Monte Carlo rollout teacher. Its certificates are stable across ten bounded instances per scenario, at 57–64% of optimal, within 3–6 points of what the 1000×-slower teacher certifies.

1 Introduction

Truckload carriers and brokers answer a stream of load tenders in real time. Each acceptance consumes a truck. The truck's future usefulness depends on where the load delivers it and how much drive and duty time its driver has left. Each rejection preserves options whose value depends on demand that has not arrived yet. The operational state of the art for future-aware acceptance is finite-lookahead Monte Carlo rollout [Bertsekas et al., 1997, Goodson et al., 2017]. Rollout is accurate, but it costs tens to hundreds of milliseconds per decision and scales poorly with fleet size. The benchmark release preceding this paper, FreightBidBench v0.3 [Chandrasekaran, 2026], made

*Evaluation artifact: FreightBidBench (`freightbidbench-v0.4-dev`); `https://github.com/aswincsekar/freightbidbench`. All values reproduce per Appendix A.

the problem reproducible: public calibration, explicit feasibility, versioned contracts, and hindsight upper bounds. Its best latency-aware method, a surrogate-to-rollout cascade, recovers about 98% of rollout profit. But the cascade inherits rollout's cost on escalated decisions, and it says nothing about distance from *optimal*.

This paper asks the harder question the bounds make precise: **how close to optimal can a real-time policy be, and can that closeness be certified per instance?** We use one object twice. The Lagrangian relaxation that dualizes the one-truck-per-load constraint gives two things: an upper bound on the optimum (the benchmark's ceiling), and per-load dual prices with per-market value potentials. From the prices and potentials we build an acceptance policy that runs in microseconds. Policy and bound come from the same relaxation, so every run ships with an observable certificate: realized policy value against realized bound.

Pairing a Lagrangian policy with its bound is the standard playbook of weakly coupled dynamic programs. The contribution here is the model in which we make it work, the form of the policy, and the two-sided theory of the certificate:

1. **Model (§2).** Resources whose *service capability is a controlled state variable*: state-dependent feasibility (location and HOS clocks gate eligibility), controlled relocation (service moves the resource), autonomous renewal (rests restore clocks). None of the adjacent literatures — classical weakly coupled DPs, online reusable resources, dual mirror descent for online allocation — covers this combination (§1.1).

2. **Policy and certificate (§3–4).** A dual-price policy whose relocation term is a *same-time spatial gradient* of an aggregated value surface, plus Theorem 1: the certificate is valid for any duals, any permissive discretization, and any surrogate quality. Lemma 1 isolates the design rule — *duals price time, gradients price place* — and we show the naive continuation rule collapses (14% of rollout) by double-charging the busy-time cost.

3. **Fluid consistency and portability (§5).** In the subcritical regime the gradient rule is exactly fluid complementary slackness (Lemma 2: idle potentials are flat), fitted prices converge to fluid rents (Lemma 3), and the policy is asymptotically optimal (Theorem 2). Basis stability further implies the fitted prices are *portable across days* — a prediction the benchmark confirms — and predicts where portability fails.

4. **Limits of the relaxation (§6).** Certificates cannot promise everything: we exhibit a three-truck kernel with a certified duality gap (exact rational arithmetic; a half-integral chain packing beats the integer optimum) and a replication lemma, so per-resource Lagrangian slack can stay bounded away from zero at every fleet size (Theorem 3). Random search shows such kernels are rare at small scale ($\sim$0.03%), while the benchmark's proportional-scaling experiments show per-resource bound slack *growing* with contention density — behavior outside the conditions under which classical $O(1)$-gap results hold.

5. **Evidence (§7).** Thirty paired seeds, paired-bootstrap confidence intervals: the training-free policy beats a rollout-trained linear surrogate on the tight (+2.0 pp, CI95 $[+0.5, +3.6]$) and mild (+3.5 pp) scenarios and ties it on scarce, at lower latency than the surrogate and $\sim 1000\times$ below rollout; certificates over ten independently bounded instances per scenario are stable (57–64% of optimal) and within 3–6 points of what rollout itself certifies — locating most of the certified gap in the relaxation, not the policy.

1.1 Related work

Weakly coupled DPs and bid prices. Bid-price control goes back to Talluri and van Ryzin [1998], and fluid policies are asymptotically optimal in classical network revenue management [Gallego and van Ryzin, 1997]. Dynamic bid prices from value approximations originate with Adelman [2007]; Lagrangian decompositions of weakly coupled DPs with policy-plus-bound pairing are standard since Hawkins [2003], Adelman and Mersereau [2008], and Topaloglu [2009], who reports bounds and policy values on the same network revenue-management instances. Our Theorem 1 inherits this mechanism; the extension is the resource model (state-dependent participation, controlled relocation) and the explicit certificate framing at sub-millisecond decision latency. Brown and Zhang [2023] bound the ALP–Lagrangian gap by $O(1)$ constants under conditions on linking constraints that state-dependence violates; Theorem 3 and the scaling experiments exhibit exactly the behavior their conditions exclude — in the natural freight scaling, constraints number $\propto K$ and participation is state-dependent.

Online reusable resources. Zhang and Cheung [2022] and the nonstationary extension Zhang and Cheung [2025] allocate units that occupy and *return unchanged*; feasibility is availability-only. A truck returns elsewhere with depleted clocks. This controlled-relocation distinction is not cosmetic: it is what makes the chain-packing LP non-integral (§6) and what the same-time gradient prices (§5).

Online allocation with dual updates. Balseiro et al. [2020] attain optimal $O(\sqrt{T})$ regret with per-request dual mirror descent for consumable budgets, motivated by the same latency constraints. We deliberately make no regret-rate claim — rates are a settled literature — and instead fix prices offline (portable by Lemma 3), leaving the online variant to future work. A related alternative is re-solving a deterministic LP at each decision epoch, which achieves excellent regret in network revenue management; we do not pursue it because per-decision LP solves conflict with the benchmark's dependency-free runtime and with microsecond latency at growing instance sizes, but a re-solving baseline is a fair addition for future comparisons.

Fleet ADP. Industrial-scale approximate dynamic programming for truckload fleets [Simão et al., 2009, Powell et al., 2012] produced marginal values of driver states as model outputs, aimed at calibration rather than optimality certification; no per-instance gap reporting appears in this line.

Compute-aware deferral. Value-of-computation metareasoning for Monte Carlo search [Sezener and Dayan, 2020] has not been applied to operational dispatch; the benchmark's cascade and this paper's policy make the freight instantiation available, and we leave a learned deferral rule to future work.

2 Model

2.1 Resources, requests, and controlled relocation

Time is continuous on $[0, T]$. A fleet of K resources (trucks) has states $x_k = (m_k, \tau_k, h_k, d_k) \in \mathcal{X}$: market m in a finite set $\mathcal{M}$, next-available time τ, and remaining drive/duty clocks (h, d). Requests (load tenders) $\ell_1, \ldots, \ell_N$ arrive at times $t_1 < \cdots < t_N$; request ℓ carries origin $o(\ell)$, destination $\delta(\ell)$, and the price/cost/window attributes of the benchmark. Three primitives distinguish the model from occupancy-based reusable resources:

- **State-dependent feasibility.** A deterministic map $f(x, \ell) \in \{0, 1\}$ gates service: $f = 1$ only if the resource is in the origin market, can reach the pickup inside its appointment window, and admits an HOS-feasible schedule through delivery.
- **Controlled relocation.** Serving executes $x \leftarrow S(x, \ell)$: the market becomes $\delta(\ell)$, the next-available time advances by the service duration including inserted rests, and the clocks are updated by the drive/duty consumed. The resource does not return to its prior pool state.
- **Autonomous renewal.** Idle resources renew clocks by resting (11/14/10 in the benchmark), so service capability regenerates in a state-dependent way.

2.2 Joint control problem

At each arrival t the controller observes $(x_1, \dots, x_K, \ell_t)$ and chooses $a_t \in \{0, 1, \dots, K\}$ (reject, or assign to a feasible resource). Writing $a_t^{(k)} = \mathbf{1}\{a_t = k\}$, the joint constraint is $\sum_k a_t^{(k)} \leq 1$. Rewards follow the benchmark: realized profit $r(x_k, \ell_t)$ on feasible accepts, a service-failure penalty on infeasible attempts (excluded under f-gated policies), and terminal fleet value $\Phi = \omega \sum_k V(m_k(T))$. $V^\star(\xi)$ denotes the optimum on realized scenario ξ.

2.3 Lagrangian per-resource decomposition

Dualizing the joint constraint with prices $\lambda = (\lambda_t) \geq 0$,

$$L(\lambda; \xi) \;=\; \sum_t \lambda_t \;+\; \sum_k V_\lambda^k(x_k(0); \xi), \tag{1}$$

where V_λ^k is resource k's optimal value in its own sub-MDP with per-service reward $r(x, \ell_t) - \lambda_t$, retaining f, S, and renewal. Weak duality gives $V^\star(\xi) \leq L(\lambda; \xi)$ pointwise for every $\lambda \geq 0$ (see also the information-relaxation view of Brown et al., 2010). The benchmark computes L by bucketed forward enumeration with a permissive-corner guarantee, so reported bounds remain valid under discretization; per-resource solves parallelize (standard-library multiprocessing; bit-identical to serial).

3 The Dual-Price Policy

3.1 Price and value surrogates

Fix any duals $\lambda \geq 0$ (in practice ∼15 subgradient iterates on one training scenario; policy quality saturates well before bound-grade convergence). Two aggregated objects are fitted offline:

- **A price surface** $\hat{\lambda} : \mathcal{M} \times [0, 24) \to \mathbb{R}_{\geq 0}$: the mean dual of loads by origin market and hour-of-day — the opportunity cost of origin-market capacity-time (Lemma 3 identifies its limit as the fluid rent $\mathbb{E}[(r - \theta)_+]$).
- **An aggregated value-to-go** $\widehat{W} : \mathcal{M} \times [0, T] \to \mathbb{R}$: the backward recursion on the (market, hour) grid over the realized train stream with dual-netted rewards, $\widehat{W}(m, T) = \omega V(m)$ and

$$\widehat{W}(m, t) = \max\Big(\widehat{W}(m, t{+}1), \max_{\ell:\, t_\ell = t,\, o(\ell) = m} \big[\bar{r}(\ell) - \lambda_{t_\ell} + \widehat{W}(\delta(\ell), t + tt(\ell))\big]\Big). \tag{2}$$

3.2 Policy

The policy makes three moves at each arrival ℓ at time t.

Step 1: pick the truck. Among feasible trucks, take the one with the best realized profit:

$$k^\star \in \arg \max_{k\colon f(x_k,\ell)=1} r(x_k,\ell). \tag{3}$$

Step 2: price the decision. Compute the relocation gain first. It is the value of standing at the destination instead of the origin, compared at the same time $t + tt(\ell)$:

$$g(\ell,t) \;=\; \widehat{W}\big(\delta(\ell), t+tt(\ell)\big) \;-\; \widehat{W}\big(o(\ell), t+tt(\ell)\big). \tag{4}$$

Then start from the profit, subtract the price of the slot, and add the relocation gain:

$$\text{score} \;=\; r(x_{k^\star},\ell) \;-\; \hat\lambda(o(\ell),t) \;+\; g(\ell,t). \tag{5}$$

Step 3: decide. Accept with $k^\star$ iff score ≥ 0.

We call g the **same-time spatial gradient**. The busy-time opportunity cost is priced once, by $\hat\lambda$; g prices only the change of place. Each decision costs a feasibility probe over origin-market resources plus two table lookups. Measured latency is 0.02–0.12 ms across fleet sizes 35–140, versus 32–159 ms for rollout. The advantage widens with fleet size.

4 Certified Gaps

Theorem 1 (Certified optimality gap)**.** *Let $\lambda \geq 0$ be arbitrary and π any feasible policy. On every realized scenario ξ, $V^\pi(\xi) \leq V^\star(\xi) \leq L(\lambda;\xi)$, hence the observable* $\text{gap}(\xi) := L(\lambda;\xi) - V^\pi(\xi)$ *upper-bounds the true suboptimality $V^\star(\xi) - V^\pi(\xi)$. The certificate remains valid for any λ, any per-resource discretization with the permissive-corner property, and any quality of $\hat\lambda, \widehat{W}$ — surrogate error degrades policy value (widening the reported gap) but never invalidates the certificate.*

Proof. We argue in four short steps.

Step 1. $V^\pi(\xi) \leq V^\star(\xi)$. The policy π is feasible, and $V^\star(\xi)$ is by definition the best value any feasible policy attains on ξ.

Step 2. $V^\star(\xi) \leq L(\lambda;\xi)$. This is weak duality for the dualized joint constraint (§2.3), and it holds pointwise in ξ for every $\lambda \geq 0$.

Step 3. Subtract $V^\pi(\xi)$ from both sides of Step 2:

$$V^\star(\xi) - V^\pi(\xi) \;\leq\; L(\lambda;\xi) - V^\pi(\xi) \;=\; \text{gap}(\xi).$$

Step 4. The invariances. The certificate uses two ingredients only: an upper bound computed from (λ,ξ), and the realized value of a feasible policy. Neither ingredient depends on how λ was chosen, on the discretization (which can only enlarge the bound, by the permissive-corner property), or on the quality of $\hat\lambda, \widehat{W}$ (which affect only which feasible policy we happen to run). $\square$

Proposition 1 (Exactness)**.** *If $\lambda^\star$ minimizes $L(\cdot\,;\xi)$ and the per-resource optimizers jointly satisfy primal feasibility and complementary slackness, the assembled plan is optimal and the certificate is tight.*

In practice a residual contested set (∼1–2% of loads) keeps the certificate strictly positive.

Lemma 1 (Busy-time double-counting)**.** *In recursion* (2)*,* $\widehat{W}(m,t)-\widehat{W}(m,t')$ *equals the value of the best dual-netted service chain departing* m *within* $[t,t')$*; hence the naive continuation* $\widehat{W}(\delta,t+tt)-\widehat{W}(o,t)$ *subtracts the busy-time opportunity value a second time on top of* $\hat{\lambda}$*, while the same-time gradient prices relocation alone.*

Proof. Define two quantities at the decision (ℓ,t) with $t'=t+tt(\ell)$:

$$D_{\text{time}} \;=\; \widehat{W}(o,t)-\widehat{W}(o,t'), \qquad D_{\text{place}} \;=\; \widehat{W}(\delta,t')-\widehat{W}(o,t').$$

Step 1. $D_{\text{time}} \geq 0$, and it equals the value of the best dual-netted chain departing o within $[t,t')$. The wait branch of recursion (2) makes $\widehat{W}(m,\cdot)$ non-increasing; unrolling the max branches over the window expresses the drop as exactly the best chain the truck gives up by being busy.

Step 2. The naive continuation splits into the two parts:

$$\widehat{W}(\delta,t')-\widehat{W}(o,t) \;=\; D_{\text{place}}-D_{\text{time}}.$$

This is a one-line check: add and subtract $\widehat{W}(o,t')$.

Step 3. The duals already price the busy time on the contested stream, so a rule that uses the naive continuation charges D_{time} a second time on top of $\hat{\lambda}$. The same-time gradient $g=D_{\text{place}}$ charges the relocation only. □

Empirically the naive rule collapses to 14–17% of rollout profit; the gradient rule attains 90–95% (§7). *Duals price time; gradients price place.* Lemma 2 upgrades this from a property of our construction to fluid optimality.

5 Fluid Consistency in the Subcritical Regime

5.1 Fluid model and its LP

Scale fleet and demand together: K resources, lane-ℓ arrivals a Poisson process of rate $K\Lambda_\ell(t)$ with mean reward $\bar{r}_\ell$ and travel time τ_ℓ, lane geometry fixed. Normalize per resource. The fluid control problem routes resource mass over the market–time network: dispatch rates $u_\ell(t)\in[0,\Lambda_\ell(t)]$ and idle mass $z_m(t)\geq 0$, with flow balance

$$\dot{z}_m(t)=\sum_{\ell:\,\delta_\ell=m} u_\ell(t-\tau_\ell)-\sum_{\ell:\,o_\ell=m} u_\ell(t), \qquad \max\int_0^T \sum_\ell \bar{r}_\ell u_\ell(t)\,dt+\omega\sum_m V(m)z_m(T). \tag{6}$$

This is a continuous transportation LP, in the fluid tradition of revenue management [Gallego and van Ryzin, 1997]; write V_f for its value and $(\lambda_\ell(t),w_m(t))$ for optimal duals on the arrival-capacity and flow-balance constraints. Dual feasibility on dispatch and waiting arcs reads

(D1) $\quad w_{o_\ell}(t)\geq \bar{r}_\ell-\lambda_\ell(t)+w_{\delta_\ell}(t+\tau_\ell),\quad$ with equality on used arcs,

(D2) $\quad w_m(t)\geq w_m(t^+),\quad$ with equality where $z_m(t)>0$.

Assumption 1. *(A1) Subcriticality: the fluid optimum keeps* $z_m(t)\geq z_{\min}>0$ *for all* m,t*. (A2) Clock and window slack: per-resource service intensity at the fluid optimum is below the HOS renewal rate, and idle in-market resources are feasible for in-market arrivals (appointment windows and pickup reach are non-binding for idle local resources), so feasibility reduces to the presence of idle mass in the limit. (A3) Nondegeneracy: the fluid LP has a unique dual optimum with finitely many basis-switch times. (A4) Bounded rewards and travel times;* $\Lambda_\ell(\cdot)$ *piecewise continuous (the benchmark's periodic schedule qualifies); and rewards within a lane–hour cell are homogeneous at scale* ($r\equiv\bar{r}_\ell(h)$ *up to* $o(1)$ *dispersion). Remark 1 discusses heterogeneous rewards.*

Lemma 2 (Idle potentials are flat; the gradient rule is fluid CS)**.** *Under (A1), (D2) holds with equality for all m, t: $w_m(\cdot)$ is constant between basis switches, and in particular $w_o(t) = w_o(t+\tau)$ whenever idle mass persists at o on $[t, t+\tau]$. Consequently the fluid-optimal acceptance rule — serve lane ℓ at t iff $\bar{r}_\ell - \lambda_\ell(t) + w_{\delta_\ell}(t+\tau_\ell) - w_{o_\ell}(t) \geq 0$ — coincides with the same-time spatial-gradient rule*

$$\bar{r}_\ell - \lambda_\ell(t) + \big[w_{\delta_\ell}(t+\tau_\ell) - w_{o_\ell}(t+\tau_\ell)\big] \geq 0, \tag{7}$$

and the naive-continuation penalty $w_o(t) - w_o(t+\tau)$ is exactly the binding-capacity correction: zero under (A1), strictly positive where capacity binds.

Proof. Step 1. Complementary slackness on a waiting arc says: if idle mass sits at (m, t), the arc is used, so its dual inequality (D2) is tight. Under (A1) idle mass is positive everywhere, so (D2) holds with equality everywhere.

Step 2. Equality in (D2) at every t means $w_m(\cdot)$ cannot drop between basis switches: w_m is flat. In particular $w_o(t) = w_o(t+\tau_\ell)$.

Step 3. Take the fluid-optimal rule from (D1), $\bar{r}_\ell - \lambda_\ell(t) + w_{\delta_\ell}(t+\tau_\ell) - w_{o_\ell}(t) \geq 0$, and replace $w_{o_\ell}(t)$ by $w_{o_\ell}(t+\tau_\ell)$ using Step 2. This gives rule (7) exactly.

Step 4. Where capacity binds ($z_o = 0$) Step 1 fails, (D2) can be strict, and the replaced term $w_o(t) - w_o(t+\tau)$ becomes a positive correction. That correction is what the naive rule charges even when capacity is slack. □

Lemma 2 upgrades Lemma 1 to the statement that *the same-time gradient is the fluid-optimal rule in the subcritical regime*, and explains the naive rule's empirical collapse as a misapplied binding-capacity correction.

5.2 Price consistency and the theorem

Lemma 3 (Dual stability)**.** *Work with the hourly fluid LP (cells (ℓ, h); the fitting granularity of $\hat{\lambda}$ and $\widehat{W}$). Under (A1)–(A4), on a sample path at scale K: (a) the realized arc-flow LP's node potentials $\hat{w}_m(h)$ equal the fluid potentials $w_m(h)$ with probability $1 - o(1)$, and its cell thresholds converge at $O_p(K^{-1/2})$; (b) the bucket-averaged load duals satisfy $\hat{\lambda}(\ell, h) \to \lambda_\ell(h) := \mathbb{E}[(r - \theta_{\ell h})_+]$ at $O_p(K^{-1/2})$, where $\theta_{\ell h} = w_{o_\ell}(h) - w_{\delta_\ell}(h + \tau_\ell)$ is the fluid acceptance threshold; (c) the chain-packing duals of Lemma 4 coincide with the arc-flow duals on an event of probability $1 - o(1)$.*

Proof. (Step 1: cell LP and basis stability.) Aggregate the realized instance into cells: $N_{\ell h}$ loads arrive in cell (ℓ, h) with i.i.d. rewards; the arc-flow relaxation over the market–hour network is a finite LP with data $b_K = N/K$ (arrival masses) and c_K (mean rewards). By Poisson concentration and the CLT over finitely many cells, $b_K = b + O_p(K^{-1/2})$ and $c_K = c + O_p(K^{-1/2})$. Let $B^\star$ be the fluid LP's optimal basis, unique and nondegenerate by (A3). Its optimality conditions — reduced-cost signs and basic-variable positivity [e.g., Bertsimas and Tsitsiklis, 1997] — hold with strict inequality, hence remain valid on an open neighborhood of (b, c). On the event $E_K = \{(b_K, c_K) \text{ in that neighborhood}\}$ — probability $1 - o(1)$ — the realized LP has the same optimal basis. Dual solutions are functions of the basis and objective only, $y = c_B (B^\star)^{-1}$: they do not depend on b, so $\hat{w}_m(h) = w_m(h)$ exactly on E_K up to the $O_p(K^{-1/2})$ objective perturbation. This proves (a).

(Step 2: disaggregation.) In the load-level LP each load i in cell (ℓ, h) has its own capacity constraint with dual λ_i. With the basis fixed, complementary slackness gives $\lambda_i = (r_i - \hat{\theta}_{\ell h})_+$: loads strictly above threshold are served and earn their surplus as rent; loads below have slack capacity and zero dual. Averaging over the $N_{\ell h} = \Theta(K)$ i.i.d. rewards, $\hat{\lambda}(\ell, h) \to \mathbb{E}[(r - \theta_{\ell h})_+]$ by the law of large numbers at rate $O_p(K^{-1/2})$, using $\hat{\theta} \to \theta$ from (a). This proves (b).

(Step 3: chains vs flows.) $\mathrm{conv}(\mathrm{chains}_k) \subseteq \{\text{unit flows from } k\text{'s start}\}$, with equality of optimal values whenever the resource-budget (clock) side constraints are slack at the flow optimum — which holds under (A2) on an event of probability $1 - o(1)$ (realized service intensities track fluid intensities by Step 1 plus Azuma's inequality). On that event the additional chain constraints are non-binding and both LPs share optimal duals. This proves (c). ☐

Remark 1 (Heterogeneous rewards)**.** *Under within-cell homogeneity, the mean rent equals the classical bid-price form:* $\mathbb{E}[(r - \theta)_+] = (\bar{r} - \theta)_+$*, and subtracting it in rule* (5) *reproduces the fluid rule exactly. With genuinely heterogeneous within-cell rewards, the fluid optimum is a* threshold *rule (accept iff* $r \geq \theta$*), while the mean rent is strictly positive on contested cells; a policy that subtracts the mean rent over-shades and accepts conservatively. The consistent extension is to fit quantile (marginal) prices rather than mean rents. We leave that refinement to future work and note two things: the certificate of Theorem 1 is unaffected (it holds for any policy), and empirically the conservative shading is partially absorbed by the single-stream bias of* $\widehat{W}$*'s gradient, which the strong held-out results reflect.*

Remark 2 (Portability, explained)**.** *Step 1 says the price surface is a* basis object*: two independent sample paths yield the same potentials whenever both land in the stability neighborhood — probability* $(1 - o(1))^2$*. The empirical portability finding (identical policy decisions under frozen vs fresh duals) is LP basis stability, not an accident of the benchmark; and portability should fail near fluid basis switches (arrival-regime changes), a testable prediction.*

Theorem 2 (Subcritical fluid consistency)**.** *Under (A1)–(A4), the dual-price policy* π_λ *built from* $\hat{\lambda}, \widehat{W}$ *fitted from an optimal (or* $o(1)$*-suboptimal) dual solution of an independent sample path's relaxation satisfies*

$$V^{\pi_\lambda}(K) \geq V_f \cdot K - o(K) \quad \textit{and} \quad V^\star(K) \leq V_f \cdot K + o(K), \qquad \textit{hence} \quad \frac{V^{\pi_\lambda}(K)}{V^\star(K)} \to 1. \tag{8}$$

Proof sketch. We prove the two inequalities separately. Steps 1–3 and 5 are complete arguments; Step 4 states the coupling estimate whose full interacting-fleet argument (a propagation-of-chaos coupling of the policy-driven fleet process to the fluid trajectory) is the one step we present at sketch level.

Step 1 (upper bound on $V^\star$*).* The realized instance is the fluid LP with perturbed arrival counts. Poisson concentration (Bernstein, applied per lane–hour cell and summed over the finitely many cells) bounds the perturbation's effect on the LP value by $o(K)$. Hence $V^\star(K) \leq V_f K + o(K)$.

Step 2 (the fitted surfaces are nearly fluid). By Lemma 3, $\hat{\lambda}$ and $\widehat{W}$ are within $o_p(1)$ of the fluid duals, except near the finitely many basis-switch times, whose neighborhoods have total measure $o(T)$.

Step 3 (the policy plays the fluid rule). By Lemma 2, the fluid-optimal rule is the same-time gradient rule. Combining with Step 2, the policy's decisions agree with the fluid-optimal decisions except on three exception sets: switch-time neighborhoods; decisions whose score is within $o_p(1)$ of zero (their fluid mass vanishes by (A3)); and sample-path fluctuation events.

Step 4 (accepted arrivals find trucks). Under (A1), idle mass at every origin market stays above $z_{\min} K - O_p(\sqrt{K}) > 0$ (Azuma over arrival blocks), and (A2) keeps clocks and windows non-binding. So every fluid-accepted arrival finds a feasible resource, and realized service rates track fluid rates to $O_p(\sqrt{K})$.

Step 5 (sum up). The reward lost on the three exception sets is $o(K)$ in total, so $V^{\pi_\lambda}(K) \geq V_f K - o(K)$. Combine with Step 1 to get the ratio. ☐

Remark 3. *(a) The critical regime $z_{\min} = 0$ (tight/scarce economics) is genuinely open: (D2) equality fails, the gradient rule loses its CS status, and §6 shows per-resource duality slack can persist; we conjecture $V^{\pi_\lambda}/V^\star \to 1 - \varepsilon(\rho)$ with ε growing in contention. (b) Assumption (A2) is where HOS clocks live; relaxing it requires a regeneration argument at rest events and is this section's main technical debt. (c) In practice the fitted duals are $\sim$15 subgradient iterates rather than exact optima. Under (A3) the dual function has a unique minimizer and is polyhedral, hence sharp: $o(1)$ suboptimality in value forces $o(1)$ distance to the optimal dual vector, so value convergence of the subgradient method suffices. Empirically the policy's decisions are already invariant at that budget (§7), consistent with the basis-stability mechanism of Lemma 3.*

6 Limits of the Relaxation

Theorem 3 (Per-resource slack bounded away from zero)**.** *There exist instance families of the model with $[\min_\lambda L(\lambda) - V^\star]/K \ge \gamma > 0$ for all fleet sizes K.*

Lemma 4 (Dantzig–Wolfe equivalence)**.** *For a realized instance, let $\mathcal{C}_k$ denote resource k's finitely many feasible service chains and $\mathrm{val}(S)$ a chain's total reward. Then $\min_{\lambda \ge 0} L(\lambda)$ equals the chain-packing LP*

$$\max\Big\{\textstyle\sum_k \sum_{S\in\mathcal{C}_k} \mathrm{val}(S)\, x_{k,S} \;:\; \sum_k \sum_{S\ni t} x_{k,S} \le 1\ \forall t,\ \ \sum_S x_{k,S} \le 1\ \forall k,\ \ x \ge 0\Big\}. \tag{9}$$

Proof. $L(\lambda) = \sum_t \lambda_t + \sum_k \max_{S\in\mathcal{C}_k\cup\{\emptyset\}}[\mathrm{val}(S) - \lambda(S)]$ is the Lagrangian dual function of the integer chain-packing program; each inner maximum is over a finite set, so $\min_\lambda L$ equals the LP relaxation over the convex hulls of the per-resource feasible sets [Dantzig–Wolfe; Geoffrion, 1974], which is the displayed column LP. □

Lemma 5 (Replication)**.** *Let I_0 have k_0 resources and duality gap $g_0 = \min_\lambda L(\lambda; I_0) - V^\star(I_0) > 0$. Then n disjoint copies satisfy $V^\star(I_0^n) = nV^\star(I_0)$ and $\min_\lambda L(\cdot\,; I_0^n) = n \min_\lambda L(\cdot\,; I_0)$, hence per-resource slack $g_0/k_0 > 0$ for every $K = nk_0$.*

Proof. No chain, constraint, or transition couples distinct copies, so both the joint DP and the chain-packing LP separate and optima add; apply Lemma 4 per copy. □

Kernel K_1 (exhibited; exact rational certificates). Three resources with staggered availabilities and service budget 2 (a clock proxy), six loads in cyclic geography over three markets (`scripts/find_gap_kernel.py`; `reports/gap_kernel_search.md`). The certificate is checkable by hand in four steps.

Check 1 (integer optimum). Exhaustive enumeration of the joint decision tree gives $V^\star(K_1) = 41$.

Check 2 (a fractional packing). Put weight $\frac{1}{2}$ on five chains:

$$T_0{:}\{L_4\},\quad T_1{:}\{L_2, L_3\},\quad T_1{:}\{L_5\},\quad T_2{:}\{L_3, L_5\},\quad T_2{:}\{L_1, L_4\}.$$

Feasibility: each shared load appears in exactly two chains (L_3, L_4, L_5), so its total weight is $\frac{1}{2} + \frac{1}{2} = 1$; every other load appears once at weight $\frac{1}{2}$; trucks T_1 and T_2 each run two chains at total weight 1, and T_0 runs one at $\frac{1}{2}$.

Check 3 (its value). The chain values are $7, 20, 14, 28, 20$, so the packing is worth

$$\tfrac{1}{2}\,(7 + 20 + 14 + 28 + 20) \;=\; \tfrac{89}{2} \;=\; 44.5.$$

Table 1: Retention vs the rollout teacher over thirty held-out seed pairs. The last rows give the paired difference of the zero-label dual policy against the rollout-trained surrogate, and mean per-decision latency.

policy	tight	scarce	mild
`bid_price`	91.1%	85.9%	99.1%
surrogate (200 labels/pair)	93.4%	91.1%	98.3%
`dual_price` (flat ablation)	91.2%	86.6%	99.1%
`dual_price_vf` (zero labels)	**95.4%**	**90.0%**	**101.8%**
paired vf − surrogate	+2.0 pp [+0.5, +3.6]	−1.0 pp [−3.2, +1.3]	+3.5 pp [+2.4, +4.5]
Wilcoxon signed-rank p	0.023	0.299	< 0.001
latency vf / surrogate / rollout (ms)	0.059 / 0.081 / 74.5	0.040 / 0.075 / 48.0	0.035 / 0.106 / 111.7

Check 4 (conclude). By Lemma 4, $\min_\lambda L$ equals the chain-packing LP value, which is at least the value of any feasible packing. So

$$\min_{\lambda} L \;\geq\; \tfrac{89}{2} \;>\; 41 \;=\; V^\star(K_1),$$

a certified duality gap of at least $\frac{7}{2}$ (8.5% of $V^\star$). The binding structure — one resource holding two valuable chains it can execute only once, with odd-cyclic overlap through shared loads — is exactly the structure whose absence provably collapses the gap (identical loads, one-shot assignment, and exchangeable resources each yield zero gap). Service budgets are the special case of duty clocks whose renewal is slower than the residual horizon, so K_1 lies inside the model of §2; and K_1 is not subcritical — its contested loads bind capacity — so there is no tension with Theorem 2. Theorem 3 follows from K_1 with $\gamma = 7/6$.

Remark 4 (Rarity and density)**.** *Random micro-instances carry certified gaps in only* $\sim$0.03% *of draws — at small scale the relaxation is almost always tight — while the benchmark's proportional-scaling experiments show per-resource bound slack growing with density* (+17% *per resource from* $K = 70$ *to* 140*, robust to two step-size schedules). The density-growth characterization is open. There is no contradiction with the* $O(1)$ *total-gap results of Brown and Zhang [2023]: those hold for a fixed number of state-independent linking constraints, whereas here constraints number* $\propto K$ *with state-dependent participation — the natural scaling for online freight.*

7 Experiments

All experiments use the public benchmark's frozen `scenario-v0.3.2` contract [Chandrasekaran, 2026], thirty held-out paired seeds (pairs 1–30; the dual tables are fitted on pair 0's stream and never refitted), paired-bootstrap 95% confidence intervals (20,000 resamples), and two-sided sign tests. Policies: `bid_price` (future-value proxy), `surrogate_linear` (ridge regression on 200 rollout labels per pair, retrained per pair), `dual_price` (flat-price ablation), `dual_price_vf` (this paper), and `rollout_teacher`. Latencies in Table 1 were measured single-threaded on a 4-core cloud VM; absolute values are Python reference latencies, and the cross-policy ratios are the meaningful quantity.

The training-free policy beats the rollout-trained surrogate on tight (bootstrap CI excludes zero; Wilcoxon $p = 0.023$; the pure sign test is weaker at 19/30 wins, $p = 0.20$ — the edge is in magnitudes) and on mild (all three tests agree; 26/30 wins), and ties on scarce. On mild the policy also exceeds the rollout teacher itself (101.8%): rollout is a finite-lookahead stochastic heuristic

Table 2: Per-instance certificates: realized policy value over the instance's Lagrangian bound, across ten independently bound-solved instances per scenario (45-iteration solves).

	`dual_price_vf`	`rollout_teacher`
tight	**60.0%** of optimal (56.9–63.6)	62.8% (60.5–66.8)
scarce	**57.5%** (53.4–61.5)	63.7% (61.2–65.3)

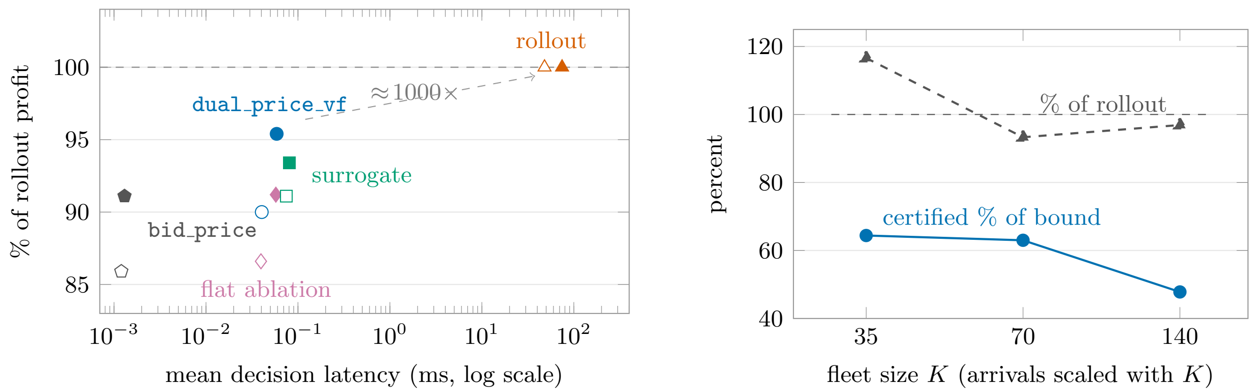

Figure 1: Left: the latency–quality frontier over thirty held-out seed pairs (solid marks: tight; open marks: scarce). The log axis shows what the tables cannot: `dual_price_vf` gives up 4.6–10 points of rollout profit to move three orders of magnitude left, and it sits above the rollout-trained surrogate at lower latency. Right: proportional fleet scaling. The policy tracks rollout (dashed, top) while the certified fraction of the bound falls at $K = 140$ — the bound loosens with density (Theorem 3); the policy does not degrade.

whose Monte Carlo noise mis-rejects at low contention, while the global price signal does not — the same mechanism as the K=35 scaling cell below. The flat ablation sits at bid-price level: the same-time gradient is the source of the edge, as Lemmas 1 and 2 predict.

Certificates are narrow, stable distributions; the 1000×-slower rollout teacher certifies only 2.8–6.2 pp higher, locating most of the certified gap in the relaxation — consistent with §6 and with the scaling diagnosis below. Figure 1 shows the latency–quality frontier (left) and the scaling decomposition (right).

The policy is scale-robust against rollout while per-resource bound slack inflates at 2× density (Figure 1, right); the level is robust to a 2× step-size schedule reset (13 aggressive iterations never improve on the fine-schedule bound). Rollout's latency grows 32 → 159 ms over this range; the dual policy stays ∼0.1 ms.

Portability and solver budgets. On the tested pairs, decisions are identical under frozen (pair-0) vs freshly solved (pair-1) duals on both stress scenarios — the finite-sample shadow of Lemma 3's basis stability. Policy quality saturates by ∼15 subgradient iterations; bound-grade certificates require ∼45; per-resource solves parallelize at 3.05× on four cores.

8 Discussion

A certificate culture for freight decisioning. Every policy run on the benchmark can now report "provably ≥ x% of optimal on this instance." The certificates are conservative exactly

Table 3: Proportional fleet scaling (tight economics; plateau-converged solves, trajectory-verified). Unlike Table 1, policy values here are in-sample: tables are fitted from each cell's own solved stream, since certificates require the bound and the policy on the same instance.

K	bound/K	policy/K	policy vs rollout	certified gap
35	$26,780	$17,242	116.6%	35.6%
70	$26,929	$16,970	93.3%	37.0%
140	$31,585	$15,104	96.9%	52.2%

where §6 says they must be. The decomposition into policy loss and relaxation slack tells methods researchers where improvement is possible (tighter bounds) and where it is not (the policy is already near rollout).

Relation to the benchmark's cascade. The v0.3 surrogate-to-rollout cascade [Chandrasekaran, 2026] attains about 98% of rollout on these scenarios. But it needs 200 rollout labels per stream to train, it invokes rollout on escalated decisions (12–17 ms mean latency), and it carries no optimality statement. The dual policy trades 3–8 retention points for three orders of magnitude in latency, zero training data, and a certificate. The two are complementary. Replacing the cascade's cheap stage with the dual policy is a natural follow-up, and we leave it to future work.

Prices are infrastructure. Basis stability makes the fitted price surface a reusable asset. One offline solve prices weeks of operations. There is a testable warning condition — arrival-regime switches — for when to refit. Per-request-learning schemes do not offer this property.

Where the money is. The remaining certified gap at scale is dominated by cross-resource coupling slack. Neither faster policies nor more solver iterations recover it. Tightening the bound itself — chain-aware duals, restricted exchange constraints — is the natural next methodological target, and Theorem 3's kernel families are the test cases.

9 Limitations

Certificates inherit the bound's looseness. In the critical regime they understate policy quality (Table 2's rollout row), and Theorem 2 does not apply there; the critical case is open. Assumption (A2) folds HOS-clock renewal into the regime definition, and relaxing it needs a regeneration argument. The surrogate baseline is the benchmark's dependency-free linear model, not a tuned learned policy; rollout is the strong baseline. Benchmark caveats — public calibration, lane concentration — carry over from the v0.3 release [Chandrasekaran, 2026]. We defer the online-updating variant and a learned deferral rule to future work, by scope discipline.

10 Conclusion

One Lagrangian object gives two things: a microsecond, training-free, portable acceptance policy, and a per-instance certificate of its optimality gap. The theory says when the certificate is tight (subcritical fluid consistency, Theorem 2) and shows, with exact certificates, why it cannot always be (kernels with irreducible per-resource slack, Theorem 3). On a public benchmark the policy beats a rollout-trained surrogate on two of three scenarios and ties the third, at three orders of magnitude lower latency than the teacher. We release the benchmark, the solver, the kernel search, and every table for independent verification.

Declaration of Competing Interest

The author is employed by Bubba AI, which develops AI-based load-planning and carrier-operations products in the freight domain addressed by this paper. This study uses only public data and the open-source benchmark; no proprietary data or systems were used, and Bubba AI had no role in the study design, the analysis, or the decision to publish.

Acknowledgements

The author used Anthropic's Claude, a large language model, to assist with software implementation and experiment scaffolding, drafting and editing of the manuscript, and literature and formatting support. All model designs, proofs, experiments, results, and claims were verified and approved by the author, who takes full responsibility for the content.

A Reproducibility

Python 3.10 or newer, standard library only. Headline commands: the 30-seed program (`scripts/run_30seed_program.sh`: Phase A policy comparisons via `scripts/run_dual_price_experiment.py`, Phase B certificates via `scripts/run_lagrangian_bound.py` with `--workers 4`; the corrected `mild` rows with fitted tables are under `benchmark_runs/v04_dev/seed30_mild_fitted/`), price/value fitting (`scripts/fit_dual_prices.py`, `scripts/fit_value_togo.py`), scaling cells (`configs/freightbidbench_v04_dev_scaling.json`), kernel search with exact certificates (`scripts/find_gap_kernel.py`). The public v0.4 contract (`configs/freightbidbench_v04_scenarios.json`, `policy-set-v0.4.0`) includes both dual policies in the default policy set; artifacts under `benchmark_runs/v04_dev/`; manifests and per-iteration dual checkpoints throughout.

References

Daniel Adelman. Dynamic bid prices in revenue management. *Operations Research*, 55(4):647–661, 2007. doi: 10.1287/opre.1060.0368.

Daniel Adelman and Adam J. Mersereau. Relaxations of weakly coupled stochastic dynamic programs. *Operations Research*, 56(3):712–727, 2008. doi: 10.1287/opre.1070.0445.

Santiago R. Balseiro, Haihao Lu, and Vahab Mirrokni. Dual mirror descent for online allocation problems. In *Proceedings of the 37th International Conference on Machine Learning (ICML), PMLR 119*, 2020.

Dimitri P. Bertsekas, John N. Tsitsiklis, and Cynara Wu. Rollout algorithms for combinatorial optimization. *Journal of Heuristics*, 3(3):245–262, 1997. doi: 10.1023/A:1009635226865.

Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, 1997.

David B. Brown and Jingwei Zhang. On the strength of relaxations of weakly coupled stochastic dynamic programs. *Operations Research*, 2023. doi: 10.1287/opre.2022.2287.

David B. Brown, James E. Smith, and Peng Sun. Information relaxations and duality in stochastic dynamic programs. *Operations Research*, 58(4):785–801, 2010. doi: 10.1287/opre.1090.0796.

Aswin Chandrasekaran. Latency-aware bid acceptance under operational feasibility: A public benchmark with hindsight ceilings. arXiv preprint arXiv:2607.07343, 2026.

Guillermo Gallego and Garrett van Ryzin. A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research*, 45(1):24–41, 1997.

Arthur M. Geoffrion. Lagrangean relaxation for integer programming. *Mathematical Programming Study*, 2:82–114, 1974.

Justin C. Goodson, Barrett W. Thomas, and Jeffrey W. Ohlmann. A rollout algorithm framework for heuristic solutions to finite-horizon stochastic dynamic programs. *European Journal of Operational Research*, 258(1):216–229, 2017. doi: 10.1016/j.ejor.2016.09.040.

Jeffrey Thomas Hawkins. *A Lagrangian Decomposition Approach to Weakly Coupled Dynamic Optimization Problems and its Applications*. PhD thesis, Massachusetts Institute of Technology, 2003.

Warren B. Powell, Hugo P. Simão, and Belgacem Bouzaiene-Ayari. Approximate dynamic programming in transportation and logistics: A unified framework. *EURO Journal on Transportation and Logistics*, 1(3):237–284, 2012. doi: 10.1007/s13676-012-0015-8.

Eren Sezener and Peter Dayan. Static and dynamic values of computation in mcts. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), PMLR 124*, 2020.

Hugo P. Simão, Joseph Day, Abraham P. George, Ted Gifford, John Nienow, and Warren B. Powell. An approximate dynamic programming algorithm for large-scale fleet management: A case application. *Transportation Science*, 43(2):178–197, 2009. doi: 10.1287/trsc.1080.0238.

Kalyan T. Talluri and Garrett van Ryzin. An analysis of bid-price controls for network revenue management. *Management Science*, 44(11):1577–1593, 1998.

Huseyin Topaloglu. Using lagrangian relaxation to compute capacity-dependent bid prices in network revenue management. *Operations Research*, 57(3):637–649, 2009. doi: 10.1287/opre.1080.0641.

Xilin Zhang and Wang Chi Cheung. Online resource allocation for reusable resources. arXiv:2212.02855, 2022.

Xilin Zhang and Wang Chi Cheung. Online allocation of reusable resources in nonstationary environments. *Mathematics of Operations Research*, 2025. doi: 10.1287/moor.2023.0250.