Online Appendix of:
Near-Optimal Dispatch Policies for Emergency Medical Services

Cheng Hua, Tong Wang, Jingwei Zhang, Ziyan Zhou

A Miscellaneous Proofs

A.1 Proofs in Section 3.1

Proof of Theorem 3.1.

We first show that the Lagrangian relaxation provides a lower bound for any Lagrange multiplier $\boldsymbol{\gamma} \geq 0$. Let π^* denote an optimal policy for the original Problem (4). The actions $a_{ik}^{\pi^*}$ induced by policy π^* are feasible for the relaxed Lagrangian DP (5), which implies that $c^{\boldsymbol{\gamma}} \leq c^*$, as

$$\gamma_j \left(1 - \sum_{i=1}^{N} \sum_{k=1}^{B_i} a_{ik}^{\pi^*}\right) \geq 0.$$

To show that (6) holds, it is sufficient to show that it satisfies the relaxed Lagrangian DP (5) with the optimal unit-specific Problem (7). For any state $\boldsymbol{s}$ and arriving call j, we have

$$\begin{aligned}
&V^{\boldsymbol{\gamma}}(\boldsymbol{s}, j) + c^{\boldsymbol{\gamma}} \\
=&V_0^{\boldsymbol{\gamma}}(j) + \sum_{i=1}^{N} \sum_{k=1}^{B_i} V_i^{\boldsymbol{\gamma}}(s_{ik}, j) + c_0 + \sum_{i=1}^{N} B_i c_i^{\boldsymbol{\gamma}} = V_0^{\boldsymbol{\gamma}}(j) + c_0 + \sum_{i=1}^{N} \sum_{k=1}^{B_i} \left(V_i^{\boldsymbol{\gamma}}(s_{ik}, j) + c_i^{\boldsymbol{\gamma}}\right) \\
=&C_j - \gamma_j + \mathbb{E}\left[V_0^{\boldsymbol{\gamma}}(\tilde{j}) \middle| j\right] + \sum_{i=1}^{N} \sum_{k=1}^{B_i} \min_{a_{ik} \in A_i(s_{ik})} \left\{(c_{ij} - C_j + \gamma_j) a_{ik} + \mathbb{E}\left[V_i^{\boldsymbol{\gamma}}(\tilde{s}_{ik}, \tilde{j}) \middle| (s_{ik}, j, a_{ik})\right]\right\} \\
=& \min_{\boldsymbol{a} \in \boldsymbol{A}(\boldsymbol{s})} \left\{C_j - \gamma_j + \sum_{i=1}^{N} \sum_{k=1}^{B_i} (c_{ij} - C_j + \gamma_j) a_{ik} + \mathbb{E}\left[V_0^{\boldsymbol{\gamma}}(\tilde{j}) + \sum_{i=1}^{N} \sum_{k=1}^{B_i} V_i^{\boldsymbol{\gamma}}(\tilde{s}_{ik}, \tilde{j}) \middle| (\boldsymbol{s}, j, \boldsymbol{a})\right]\right\} \\
=& \min_{\boldsymbol{a} \in \boldsymbol{A}(\boldsymbol{s})} \left\{C_j + \sum_{i=1}^{N} \sum_{k=1}^{B_i} (c_{ij} - C_j) a_{ik} - \gamma_j \left(1 - \sum_{i=1}^{N} \sum_{k=1}^{B_i} a_{ik}\right) + \mathbb{E}\left[V^{\boldsymbol{\gamma}}(\tilde{\boldsymbol{s}}, \tilde{j}) \middle| (\boldsymbol{s}, j, \boldsymbol{a})\right]\right\}.
\end{aligned}$$

Therefore, the decomposition (6) satisfies the relaxed Lagrangian DP (5). □

A.2 Proofs in Section 3.2

Proof of Theorem 3.2.

Let $(u_i^*(s, j, a))$ denote an optimal solution to the Lagrangian relaxation dual LP (9). Consider a solution (x_{ij}) where

$$x_{ij} = \Lambda u_i^*(0, j, 1),\ x_{0j} = \lambda_j - \Lambda \sum_{i=1}^{N} u_i^*(0, j, 1).$$

We first note that $x_{ij} \geq 0$ and $x_{0j} \geq 0$, due to the non-negativity of $u_i^*(s, j, a)$ and constraint (9c) in the Lagrangian dual LP $\lambda_j \geq \Lambda \sum_{i=1}^{N} u_i(0, j, 1)$. Inserting the above solution into the objective value of (11), we have

$$\begin{aligned}
&\frac{1}{\Lambda}\left(\sum_{j=1}^{J} C_j x_{0j} + \sum_{i=1}^{N}\sum_{j=1}^{J} c_{ij}x_{ij}\right) \\
=&\sum_{j=1}^{J}\frac{\lambda_j}{\Lambda}C_j - \sum_{j=1}^{J}\sum_{i=1}^{N} C_j u_i^*(0,j,1) + \sum_{j=1}^{J}\sum_{i=1}^{N} c_{ij}u_i^*(0,j,1) \\
=&\sum_{j=1}^{J}\frac{\lambda_j}{\Lambda}C_j + \sum_{j=1}^{J}\sum_{i=1}^{N}(c_{ij}-C_j)u_i^*(0,j,1),
\end{aligned}$$

which is the optimal objective value of (9). Constraint (11b) holds by construction. Next, we are left to show (x_{ij}) satisfies Constraint (11c). By Constraint (9d), we have

$$u_i^*(j,j',0) = \frac{\lambda_{j'}}{\Lambda}\left(\sum_{j''=1}^{J}\frac{\Lambda}{\Lambda+\mu_{ij}}u_i^*(j,j'',0) + \frac{\Lambda}{\Lambda+\mu_{ij}}u_i^*(0,j,1)\right),$$

which implies that

$$\sum_{j'=1}^{J} u_i^*(j,j',0) = \frac{\Lambda}{\Lambda+\mu_{ij}}\left(\sum_{j'=1}^{J} u_i^*(j,j',0) + u_i^*(0,j,1)\right),$$

and thus, by reorganizing terms, we obtain,

$$\Lambda\mu_{ij}^{-1}u_i^*(0,j,1) = \sum_{j'=1}^{J} u_i^*(j,j',0).$$

Therefore, we have

$$\sum_{j=1}^{J}\mu_{ij}^{-1}x_{ij} = \Lambda\sum_{j=1}^{J}\mu_{ij}^{-1}u_i^*(0,j,1) = \sum_{j=1}^{J}\sum_{j'=1}^{J}u_i^*(j,j',0) \le \sum_{j=1}^{J}\sum_{s,a}u_i^*(s,j,a) = B_i.$$

In conclusion, the constructed solution (x_{ij}) is feasible for the deterministic LP (11) and achieves the same objective value as the optimal objective of the Lagrangian relaxation dual LP (9), which implies that the optimal value of (11) is no larger than that of (9). Therefore, since both are lower bounds to the original Problem (4), the Lagrangian relaxation provides a tighter bound than the deterministic LP. □

A.3 Proofs and Supplementary Analysis for Section 3.3

We first establish the following lemmas, which are critical in the proof of Theorem 3.3.

Lemma A.1. *For any constant $B \ge 1$, $f(x) = \left(\sum_{n=0}^{B}\frac{x^n}{n!}\right)^{-1}\left(\frac{x^B}{B!}\right)$ is increasing in x on $[0,+\infty)$.*

Proof of Lemma A.1.

Denote $a_n(x) = \frac{x^n}{n!}$, which is a positive, monotonically increasing function for any $n \ge 0, x \ge 0$.

Therefore, the function $\frac{\sum_{n=0}^{B} a_n(x)}{a_B(x)}$ is decreasing in x. As a result

$$f(x) = \left(\frac{\sum_{n=0}^{B} a_n(x)}{a_B(x)} \right)^{-1} = \left(\sum_{n=0}^{B} \frac{x^n}{n!} \right)^{-1} \left(\frac{x^B}{B!} \right)$$

is increasing in x for $x \geq 0$. $\square$

Lemma A.2. *For any $\alpha \in (0,1]$, define $g(n,\alpha) = \frac{(\alpha n)^n}{n!} \cdot \left(\sum_{k=0}^{n} \frac{(\alpha n)^k}{k!} \right)^{-1}$. It holds that*

$$g(n,\alpha) \leq \frac{1}{\sqrt{2\pi n}} \exp\left(n\left[1-\alpha+\ln\alpha\right]\right) \cdot \frac{1}{1-\exp\left(1-\alpha+\ln\alpha\right)}, \qquad \text{if } \alpha \in (0,1),$$

$$g(n,1) = \sqrt{\frac{2}{\pi n}} + O(\frac{1}{n}), \qquad \text{if } \alpha = 1.$$

Proof of Lemma A.2.

Denoting $\lambda = \alpha n$, we have

$$g(n,\alpha) = \frac{\lambda^n}{\mathrm{e}^{\lambda} n!} \cdot \left(\sum_{k=0}^{n} \frac{\lambda^k}{\mathrm{e}^{\lambda} k!} \right)^{-1} = \frac{\mathbb{P}\left(\text{Poisson}\left(\lambda\right) = n\right)}{\mathbb{P}\left(\text{Poisson}\left(\lambda\right) \leq n\right)}.$$

For the numerator, using the Stirling approximation, we have

$$\frac{1}{\sqrt{2\pi n}} \exp\left(n\left[1-\alpha+\ln\alpha\right]\right) \mathrm{e}^{-\frac{1}{12n}} \leq \frac{\lambda^n}{\mathrm{e}^{\lambda} n!} \leq \frac{1}{\sqrt{2\pi n}} \exp\left(n\left[1-\alpha+\ln\alpha\right]\right) \mathrm{e}^{-\frac{1}{12n+1}}.$$

For the denominator, if $\alpha = 1$, then $\lambda = n$. As n approaches infinity, we have

$$\begin{aligned}
\mathbb{P}\left(\text{Poisson}(n) \leq n\right) &= \lim_{x\to\infty} \mathbb{P}(\text{Binom}(x, n/x) \leq n) \\
&= \lim_{x\to\infty} \mathbb{P}\left(\frac{\text{Binom}(x,n/x) - n}{\sqrt{n(1-n/x)}} \leq 0 \right) \\
&\geq \lim_{x\to\infty} \left(\frac{1}{2} - K \frac{\frac{n}{x}\left(1-\frac{n}{x}\right)\left(\left(1-\frac{n}{x}\right)^2 + \left(\frac{n}{x}\right)^2\right)}{\sqrt{x}\left(\frac{n}{x}\left(1-\frac{n}{x}\right)\right)^{\frac{3}{2}}} \right) \\
&\geq \frac{1}{2} - K \lim_{x\to\infty} \frac{\left(1-\frac{n}{x}\right)^2 + \left(\frac{n}{x}\right)^2}{\sqrt{n}\left(1-\frac{n}{x}\right)^{\frac{3}{2}}} = \frac{1}{2} - \frac{K}{\sqrt{n}},
\end{aligned}$$

where the first line follows from the definition of the Poisson distribution, and the fourth line follows from the Berry–Esseen theorem, where K is a constant less than 0.475. Therefore, $\frac{1}{2} - \frac{K}{\sqrt{n}} > 0$ for any $n \geq 1$.

If $\alpha \in (0,1)$, by Chernoff bound, we have

$$\mathbb{P}\left(\text{Poisson}\left(\lambda\right) \leq n\right) = 1 - \mathbb{P}\left(\text{Poisson}\left(\lambda\right) \geq n+1\right) \geq 1 - \inf_{t>0} M(t)\mathrm{e}^{-t(n+1)},$$

where $M(t) = \exp\left(\lambda\left(\mathrm{e}^t - 1\right)\right)$ is the moment-generating function of Poisson(λ). Taking $t = -\ln\alpha$, we obtain that

$$\mathbb{P}\left(\text{Poisson}\left(\lambda\right) \leq n\right) \geq 1 - \alpha \cdot \exp\left(n\left[1 - \alpha + \ln \alpha\right]\right) > 1 - \exp\left(1 - \alpha + \ln \alpha\right) > 0.$$

Finally, we obtain that if $\alpha = 1$,

$$g(n, 1) = \sqrt{\frac{2}{\pi n}} + O(\frac{1}{n}),$$

and if $\alpha \in (0, 1)$,

$$g(n, \alpha) \leq \frac{1}{\sqrt{2\pi n}} \exp\left(n\left[1 - \alpha + \ln \alpha\right]\right) \cdot \frac{1}{1 - \exp\left(1 - \alpha + \ln \alpha\right)}.$$

□

Proof of Theorem 3.3.

(a) Denote Q_{ij} as the stationary probability that a call from sub-region j with initial station selection i but is serviced by the flexible unit; that is, Case (i) occurs. To express Q_{ij}, we denote Y_{ij} as the stationary number of calls from sub-region j serviced at station i under SAP-DLP. It then follows from the PASTA property (Wolff 1982) that for each $i = 1, \ldots, N$,

$$Q_{ij} = Q_i = \mathbb{P}\left(\sum_{j'=1}^{J} Y_{ij'} = \theta B_i\right), \forall j.$$

Then, the long-run average cost per arrival can be written as

$$\frac{1}{\Lambda^\theta} \sum_{j=1}^{J} \left[\left(x_{0j}^\theta + \sum_{i=1}^{N} x_{ij}^\theta Q_i\right) C_j + \sum_{i=1}^{N} (1 - Q_i) c_{ij} x_{ij}^\theta\right].$$

Thus, the gap between the system performance under SAP-DLP and the lower bound obtained from the deterministic LP is

$$c^\pi(\boldsymbol{x}^\theta) - \frac{1}{\Lambda^\theta}\left(\sum_{j=1}^{J} C_j x_{0j}^\theta + \sum_{i=1}^{N}\sum_{j=1}^{J} c_{ij} x_{ij}^\theta\right) \leq \frac{|\bar{C} - \underline{c}|}{\Lambda^\theta} \sum_{j=1}^{J}\sum_{i=1}^{N} x_{ij}^\theta Q_i.$$

Because (x_{ij}^θ) is non-negative, it is sufficient to bound Q_i. Following the proof of Theorem 1 in Burman et al. (1984), for station $i \in \{1, \ldots, N\}$, we consider an auxiliary system with infinite capacity, leading to the sub-region j arrives in station i in a Poisson pattern with arrival rate $\lambda_j \times \frac{x_{ij}}{\lambda_j} = x_{ij}$ and this arrival process is independent of other sub-regions. The service time in the auxiliary system follows a general distribution with mean μ_{ij}^{-1}. Let Z_{ij} be the stationary number of sub-region j calls being serviced by station i in the infinite capacity system. Then, following the standard insensitivity result for loss system (e.g., Burman et al. 1984), for $n = 0, 1, ..., \theta B_i$,

$$\mathbb{P}\left(\sum_{j=1}^{J} Y_{ij} = n\right) = \mathbb{P}\left(\sum_{j=1}^{J} Z_{ij} = n \middle| \sum_{j=1}^{J} Z_{ij} \leq \theta B_i\right) = \frac{\mathbb{P}\left(\sum_{j=1}^{J} Z_{ij} = n\right)}{\mathbb{P}\left(\sum_{j=1}^{J} Z_{ij} \leq \theta B_i\right)}.$$

In addition, each variable Z_{ij} follows an independent Poisson distribution with mean $\rho_{ij} = x_{ij}\mu_{ij}^{-1}$. Thus, we have

$$\mathbb{P}\left(\sum_{j=1}^{J} Z_{ij} = n\right) = \sum_{(z_{i1},\ldots,z_{iJ})\in\mathcal{Z}(n)} \prod_{j=1}^{J} \mathrm{e}^{-\rho_{ij}} \frac{\rho_{ij}^{z_{ij}}}{z_{ij}!} = \mathrm{e}^{-\sum_{j=1}^{J}\rho_{ij}} \frac{(\rho_{i1}+\ldots+\rho_{iJ})^n}{n!},$$

and

$$\mathbb{P}\left(\sum_{j=1}^{J} Z_{ij} \leq \theta B_i\right) = \sum_{k=0}^{\theta B_i} \mathbb{P}\left(\sum_{j=1}^{J} Z_{ij} = k\right),$$

where $\mathcal{Z}(n) = \left\{(z_{i1},\ldots,z_{iJ}) \in \mathbb{Z}_+^J : \sum_{j=1}^{J} z_{ij} = n\right\}$. This argument is also discussed in blocking-probability analyses (e.g., Eick et al. (1993), §4). Thus, $\sum_{j=1}^{J} Y_{ij}$ is a truncated Poisson random variable.

Therefore, for each $n = 0, \ldots, \theta B_i$, we obtain

$$\mathbb{P}\left(\sum_{j=1}^{J} Y_{ij} = \theta B_i\right) = \frac{\frac{(\rho_{i1}+\ldots+\rho_{iJ})^{\theta B_i}}{(\theta B_i)!}}{\sum_{n=0}^{\theta B_i} \frac{(\rho_{i1}+\ldots+\rho_{iJ})^n}{n!}} \leq \frac{\frac{(\theta B_i)^{\theta B_i}}{(\theta B_i)!}}{\sum_{n=0}^{\theta B_i} \frac{(\theta B_i)^n}{n!}} = \sqrt{\frac{2}{\pi\theta B_i}} + O\left(\frac{1}{\theta B_i}\right),$$

where the inequality follows from Lemma A.1 and Constraint (11c), and the last equality follows from Lemma A.2.

Therefore, there exists a constant K such that

$$c^{\pi}(\boldsymbol{x}^{\theta}) - \frac{1}{\Lambda^{\theta}}\left(\sum_{j=1}^{J} C_j x_{0j}^{\theta} + \sum_{i=1}^{N}\sum_{j=1}^{J} c_{ij} x_{ij}^{\theta}\right) \leq \frac{|\bar{C} - \underline{c}|}{\Lambda^{\theta}} \sum_{j=1}^{J}\sum_{i=1}^{N} x_{ij}^{\theta} Q_i \leq \frac{\left|\bar{C} - \underline{c}\right|}{\theta\Lambda} \frac{K}{\sqrt{\theta}} \sum_{j=1}^{J}\sum_{i=1}^{N} x_{ij}^{\theta} \leq \frac{\left|\bar{C} - \underline{c}\right| K}{\sqrt{\theta}},$$

where the last inequality holds due to Constraint (11b).

Next, we provide an example where the gap exactly attains the bound, showing that the bound is tight. Consider the following setting: there is only one region and one station equipped with $B = 1$ unit. The arrival rate follows the Poisson distribution with $\lambda = 1$, and the service time is exponentially distributed with a mean of $\mu^{-1} = 2$. Suppose the costs are c for the station and C for the flexible unit, with $C > c$. The deterministic linear program is as follows:

$$\underset{x_0, x_1 \geq 0}{\text{minimize}} \quad \frac{1}{\theta}(Cx_0 + cx_1) \tag{15a}$$

subject to

$$x_1 + x_0 = \theta, x_1 \leq \frac{\theta}{2} \tag{15b}$$

The solution to LP (15) is $x_0^* = x_1^* = \frac{1}{2}\theta$, and the minimal value of the objective function is $\frac{C+c}{2}$. Based on x_0^* and x_1^*, the long-run average response time per arrival under SAP-DLP π is

$$c^{\pi}(x_0^*, x_1^*) = \frac{C+c}{2} + \frac{C-c}{2}Q,$$

where Q is the stationary probability that the station has no available units when a call arrives.

Moreover, based on the proof of Theorem 3.3, we obtain

$$Q = \frac{\frac{\theta^\theta}{\theta!}}{\sum\limits_{n=0}^{\theta} \frac{\theta^n}{n!}} = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{\theta}} + O\left(\frac{1}{\theta}\right).$$

Thus, the gap between the long-run average response time per arrival under SAP π and the optimal value of the objective function is

$$c^\pi(x_0^*, x_1^*) - \frac{1}{\theta}(Cx_0^* + cx_1^*) = \frac{C-c}{2} Q = \frac{(C-c)}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\theta}} + O\left(\frac{1}{\theta}\right).$$

(b) Similar to the proof of (a), it is sufficient to bound the blocking probability of each station i to bound the optimality gap. Note that from Constraints (11b) and (11c), we obtain for each station i

$$\sum_{j=1}^{J} \rho_{ij} = \sum_{j=1}^{J} \frac{x_{ij}}{\mu_{ij}} \le \frac{1}{\min_{ij} \mu_{ij}} \sum_{j=1}^{J} \left(\sum_{i=1}^{N} x_{ij} + x_{0j} \right) = \frac{\theta \Lambda}{\omega(\theta) \min_{ij} \mu_{ij}} = \mathcal{B}\nu^\theta.$$

Therefore, for the blocking probability of station i, it holds that

$$\mathbb{P}\left(\sum_{j=1}^{J} Y_{ij} = B_i \right) = \frac{\frac{(\rho_{i1}+\ldots+\rho_{iJ})^{B_i}}{(B_i)!}}{\sum\limits_{n=0}^{B_i} \frac{(\rho_{i1}+\ldots+\rho_{iJ})^n}{n!}} \le \frac{(\mathcal{B}\nu^\theta)^{B_i}}{(B_i)!} = \frac{1}{(B_i)!} \left(\frac{\theta}{\omega(\theta)} \right)^{B_i},$$

where the first inequality holds as Lemma A.1 and the fact that $\sum_{n=0}^{B_i} \frac{x^n}{n!} \ge 1$ for any $x \ge 0$. Therefore, it holds that

$$c^\pi(\boldsymbol{x}^\theta) - \frac{1}{\Lambda^\theta} \left(\sum_{j=1}^{J} C_j x_{0j}^\theta + \sum_{i=1}^{N} \sum_{j=1}^{J} c_{ij} x_{ij}^\theta \right) \le \frac{|\bar{C} - \underline{c}|}{\Lambda^\theta} \sum_{j=1}^{J} \sum_{i=1}^{N} x_{ij}^\theta Q_i \le \frac{|\bar{C} - \underline{c}|}{\underline{B}!} \left(\frac{\theta}{\omega(\theta)} \right)^{B}.$$

□

Notably, when $\omega(\theta)$ scales proportionally to θ, part (b) of Theorem 3.3 indicates that SAP is not asymptotically optimal in this regime. Nonetheless, a lower traffic intensity, even if strictly positive, can be expected to improve the performance of SAP, as it reduces the frequency of case (i) occurrences under the policy. To formalize and leverage this intuition, we next introduce the concept of *degree of traffic*, which quantifies system load and serves as the foundation for designing policies with provable performance guarantees.

Definition A.1. Let (x_{ij}) be an optimal solution to the deterministic fluid LP (11), the *degree of traffic* δ is defined as

$$\delta = \max_{i \in \{i:\ \delta_i > 0\}} 1 - \delta_i + \ln \delta_i, \text{ where } \delta_i = \frac{1}{B_i} \sum_{j=1}^{J} \mu_{ij}^{-1} x_{ij}.$$

The value $\delta_i \in [0,1]$ represents the average utilization rate (i.e., the fraction of busy units) at station i. For notational and computational tractability, we define $f(\delta_i) = 1 - \delta_i + \ln \delta_i$, which is non-positive and increases from $-\infty$ to 0 as δ_i increases from 0^+ to 1. Accordingly, we take $\delta = \max_i f(\delta_i)$ as the system's overall traffic intensity. We also note that in the location-independent service time setting, δ is closely linked to the classical traffic intensity $\rho = \Lambda \Big/ \sum_{i=1}^{N} B_i \mu_i$. Specifically, the stability condition $\rho < 1$ (e.g., Shortle et al. 2018) follows directly from $\delta < 0$. We next show that, although δ is defined using the optimal solution to the deterministic fluid LP, it is invariant with respect to the scaling parameter θ. Consequently, under SAP-DLP, a single δ uniformly characterizes the traffic intensity for all $\theta \geq 1$.

Lemma A.3. *For the SAP-DLP policy $\pi\left(\boldsymbol{x}^{\theta *}\right)$ constructed from the optimal solution $\boldsymbol{x}^{\theta *}$ to the deterministic fluid LP* (11)*, define for each station i,*

$$\delta_i^\theta = \frac{1}{B_i^\theta} \sum_{j=1}^{J} \mu_{ij}^{-1} x_{ij}^{\theta *}.$$

It follows that δ_i^θ is invariant with respect to the scaling parameter θ, i.e., $\delta_i^\theta = \delta_i^1$ for all $\theta \geq 1$.

Proof of Lemma A.3.
For SAP-DLP, it is sufficient to show that the optimal solution of the deterministic linear program (11) is scaled linearly with θ, i.e., $x_{ij}^{\theta *} = \theta x_{ij}^{1*}$, for any $i, j, \theta \geq 1$. We first note that when the system is scaled by a factor θ, the corresponding deterministic fluid LP is

$$\begin{aligned}
&\underset{x_{ij}}{\text{minimize}} \quad \frac{1}{\theta\Lambda}\left(\sum_{j=1}^{J} C_j x_{0j} + \sum_{i=1}^{N}\sum_{j=1}^{J} c_{ij} x_{ij}\right) \\
&\text{subject to} \\
&\sum_{i=1}^{N} x_{ij} + x_{0j} = \theta\lambda_j, \forall j = 1, \ldots, J \\
&\sum_{j=1}^{J} \mu_{ij}^{-1} x_{ij} \leq \theta B_i, \forall i = 1, \ldots, N \\
&x_{ij} \geq 0, \forall i = 0, 1, \ldots, N, j = 1, \ldots, J.
\end{aligned} \tag{16}$$

For any $\theta \geq 1$, let $z_{0j} = \frac{x_{0j}}{\theta}$ and $z_{ij} = \frac{x_{ij}}{\theta}$. Taking $\{z_{ij}\}$ into the scaled linear program (16), we obtain that

$$\begin{aligned}
&\underset{z_{ij}}{\text{minimize}} \quad \frac{1}{\Lambda}\left(\sum_{j=1}^{J} C_j z_{0j} + \sum_{i=1}^{N}\sum_{j=1}^{J} c_{ij} z_{ij}\right) \\
&\text{subject to} \\
&\sum_{i=1}^{N} z_{ij} + z_{0j} = \lambda_j, \forall j = 1, \ldots, J \\
&\sum_{j=1}^{J} \mu_{ij}^{-1} z_{ij} \leq B_i, \forall i = 1, \ldots, N \\
&z_{ij} \geq 0, \forall i = 0, 1, \ldots, N, j = 1, \ldots, J.
\end{aligned} \tag{17}$$

It indicates that the optimal objective value function of the deterministic linear program is independent of θ, and $x_{ij}^{\theta *} = \theta x_{ij}^{1*}$, for any $i, j, \theta \geq 1$. $\square$

The following corollary describes how the degree of traffic influences the optimality gap.

Corollary A.4. *For SAP-DLP $\pi\left(\boldsymbol{x}^{\theta}\right)$ generated from an optimal solution $\boldsymbol{x}^{\theta}$ from the deterministic linear program, when both λ_j^θ and B_i^θ scale linearly with θ, the degree of traffic δ remains a constant,* *and there exists a constant K independent of θ such that*

$$C^{\theta}(\pi^*) - \rho^{\theta}(x^{\theta}) \le C^{\theta}(\pi(x^{\theta})) - \rho^{\theta}(x^{\theta}) \le K \exp(\theta\delta)\theta^{-1/2}.$$

Corollary A.4 explicitly describes how the traffic load of the system improves the performance of SAP-DLP. It maintains an $O(\sqrt{1/\theta})$ optimality gap when the system is in heavy traffic (i.e., $\delta = 0$), improving smoothly to $O(\exp(-\theta))$ as the degree of traffic intensity δ decreases.

Proof of Corollary A.4.
When $0 < \delta_i < 1$, we have

$$\mathbb{P}\left(\sum_{j=1}^{J} Y_{ij} = \theta B_i\right) = \frac{\frac{(\rho_{i1}+\ldots+\rho_{iJ})^{\theta B_i}}{(\theta B_i)!}}{\sum_{n=0}^{\theta B_i} \frac{(\rho_{i1}+\ldots+\rho_{iJ})^n}{n!}} \le \frac{\frac{(\delta_i \theta B_i)^{\theta B_i}}{(\theta B_i)!}}{\sum_{n=0}^{\theta B_i} \frac{(\delta_i \theta B_i)^n}{n!}} \le K_i \exp\left(\theta\left[1 - \delta_i + \ln\delta_i\right]\right)\theta^{-1/2},$$

where the first inequality follows from Lemma A.1 and Lemma A.3, and the last inequality follows from Lemma A.2, and K_i is a constant independent of θ. Note that $1 - \delta_i + \ln\delta_i < 0$ for any $\delta_i \in (0,1)$ and increasing in $\delta_i \in (0,1)$. There exists a constant K independent of θ such that

$$\begin{aligned} & c^{\pi}(\boldsymbol{x}^{\theta}) - \frac{1}{\Lambda^{\theta}}\left(\sum_{j=1}^{J} C_j x_{0j}^{\theta} + \sum_{i=1}^{N}\sum_{j=1}^{J} c_{ij} x_{ij}^{\theta}\right) \\ \le & \frac{|\bar{C} - \underline{c}|}{\Lambda^{\theta}} \sum_{j=1}^{J}\sum_{i=1}^{N} x_{ij}^{\theta} Q_i \le \frac{|\bar{C} - \underline{c}|}{\theta\Lambda} K \exp(\theta\delta)\,\theta^{-1/2} \sum_{j=1}^{J}\sum_{i=1}^{N} x_{ij}^{\theta} \le \left|\bar{C} - \underline{c}\right| K \exp(\theta\delta)\,\theta^{-1/2}, \end{aligned}$$

where the last inequality follows from Constraint (11b). □

Unlike the deterministic fluid relaxation (11), where θ is only a common scaling factor and disappears after normalization, the Lagrangian fluid LP (9) preserves the expected state transition of each ambulance between consecutive requests. As a result, θ affects the transition through the relative time scale between arrivals and completion, so changing θ alters ambulance availability and the corresponding optimal stationary occupancy measure. Consequently, the scaling property in Lemma A.3 does not extend to the SAP-LA policy in general, and whether the SAP-LA admits a refined optimality gap similar to Corollary A.4 remains open.

A.4 Proofs in Section 3.4

Proof of Theorem 3.4.

For a system with a general service time distribution, we consider an auxiliary system where the service time follows an exponential distribution with the same mean rate, keeping the rest of the system identical. We denote the optimal long-run average cost in the auxiliary system scaled by θ as $\widehat{\mathcal{C}}^{\theta}(\pi^*)$. The deterministic LPs for both systems are identical, and we denote the optimal value by $\rho^{\theta}(x^{\theta *})$ in the scaled system. Next, for the Lagrangian relaxation dual LP (9) of the scaled auxiliary system, the optimal value, $\rho^{\theta}(u_i^{\theta *}(s,j,a))$, satisfies,

$$\rho^{\theta}(u_i^{\theta *}(s,j,a)) - \rho^{\theta}(x^{\theta *}) \le \widehat{\mathcal{C}}^{\theta}(\pi^*) - \rho^{\theta}(x^{\theta *}) \le \widehat{\mathcal{C}}^{\theta}(\pi(\boldsymbol{x}^{\theta *})) - \rho^{\theta}(x^{\theta *}) \le \frac{K_1}{\sqrt{\theta}}, \tag{18}$$

where K_1 is a constant independent of θ. The last inequality follows from the part concerning SAP-DLP in Theorem 3.3. This implies that the optimal value of Lagrangian relaxation dual LP (9) converges to the optimal value of the LP (11) with a rate of $O(1/\sqrt{\theta})$ for any service time distribution. Therefore, under a system with a general service time distribution, we formulate the Lagrangian relaxation dual LP (9) of its auxiliary system. Using its optimal solution $u_i^{\theta *}(s,j,a)$, we construct a feasible solution $\boldsymbol{x}^\theta = (x_{ij}^\theta)$ to the LP (11) following Theorem 3.2, and the corresponding SAP-LA $\pi(u_i^{\theta *}(s,j,a))$ satisfy

$$\begin{aligned} &\mathcal{C}^\theta(\pi(u_i^{\theta *}(s,j,a))) - \mathcal{C}^\theta(\pi^*) \leq \mathcal{C}^\theta(\pi(u_i^{\theta *}(s,j,a))) - \rho^\theta(\boldsymbol{x}^{\theta *}) \\ =&\mathcal{C}^\theta(\pi(u_i^{\theta *}(s,j,a))) - \rho^\theta(u_i^{\theta *}(s,j,a)) + \rho^\theta(u_i^{\theta *}(s,j,a)) - \rho^\theta(\boldsymbol{x}^{\theta *}) \leq \frac{K_2}{\sqrt{\theta}}. \end{aligned}$$

where K_2 is a constant independent of θ. The last inequality follows from the part concerning SAP-LA in Theorem 3.3 and equality (18). $\square$

A.5 Analysis of Example 4.1

For policy π_S, let Q_S denote the long-run probability of arrivals served by the flexible unit and let c_S denote the long-run average cost per arrival. Then, the optimization problem is

$$\min_{S \subseteq [J]} c_S = \min_{S \subseteq [J]} C \cdot Q_S + \sum_{j \in S} c_j \cdot \frac{\lambda_j}{\lambda_S}(1 - Q_S),$$

where $\lambda_S = \sum\limits_{j \in S} \lambda_j$, and when $S = \emptyset$, the long-run average cost per arrival is C. To analyze the long-run behavior under π_S, we view the system at the epochs when the regular unit completes a service and becomes idle. These epochs form regeneration points, because once the regular unit returns to the idle state, the future evolution of the arrival and dispatching process is independent of the past and identically distributed. Hence, the process is regenerative, and we can apply renewal-cycle arguments.

In each renewal cycle, the regular unit remains idle until the first arrival from S occurs. The time to the first such arrival is exponentially distributed with rate $\lambda_S := \sum_{j \in S} \lambda_j$, and therefore the expected number of arrivals during this idle period is Λ/λ_S. The arrival that ends the idle period is necessarily served by the regular unit. Conditional on belonging to region $j \in S$, its probability is λ_j/λ_S. If such a call comes from region j, its regular-unit service time has mean $1/\mu_{1j}$; therefore, the expected number of arrivals during the subsequent service time equals

$$\sum_{j \in S} \frac{\lambda_j}{\lambda_S} \cdot \Lambda \cdot \frac{1}{\mu_{1j}} = \frac{\Lambda}{\lambda_S} \sum_{j \in S} \frac{\lambda_j}{\mu_{1j}}.$$

Letting $\rho_S = \sum\limits_{j \in S} \frac{\lambda_j}{\mu_{1j}}$, the expected number of arrivals in one renewal cycle is

$$\mathbb{E}[\text{arrivals per cycle}] = \frac{\Lambda}{\lambda_S}(1 + \rho_S).$$

Since each cycle contains exactly one arrival that is admitted to the regular unit (the first arrival from S), the long-run probability that a random arrival is both from region $j \in S$ and served by the regular unit is given by the renewal-reward identity:

$$\Pr(\text{arrival from } j \text{ served by regular}) = \frac{\mathbb{E}[\text{such arrivals per cycle}]}{\mathbb{E}[\text{arrivals per cycle}]} = \frac{(\lambda_j/\lambda_S)}{(\Lambda/\lambda_S)(1 + \rho_S)} = \frac{\lambda_j/\Lambda}{1 + \rho_S}.$$

Consequently, the long-run fraction of arrivals served by the regular unit is

$$1 - Q_S = \sum_{j \in S} \frac{\lambda_j}{\Lambda} \cdot \frac{1}{1+\rho_S} = \frac{\lambda_S/\Lambda}{1+\rho_S},$$

and hence the long-run blocking probability under policy π_S is

$$Q_S = 1 - \frac{\lambda_S/\Lambda}{1+\rho_S} = \frac{\lambda_{\bar{S}}/\Lambda + \rho_S}{1+\rho_S} = \frac{\lambda_{\bar{S}}}{\Lambda} + \frac{\lambda_S}{\Lambda} \cdot \frac{\rho_S}{1+\rho_S}.$$

In particular, for policy $\pi_{[J]}$ it holds that $Q_{[J]} = \rho_{[J]}/(1+\rho_{[J]})$.

The formulation above has a direct probabilistic interpretation. The first term, $\frac{\lambda_{\bar{S}}}{\Lambda}$, is the probability that an arriving call belongs to $\bar{S}$ and is therefore assigned to the flexible unit by design, and the second term, $(\lambda_S/\Lambda) \cdot \rho_S/(1+\rho_S)$, is the probability that an arrival comes from S and finds the regular unit busy, in which case it is also served by the flexible unit. Hence, removing a sub-region j from S has two countervailing effects: it increases the direct-dispatch probability but decreases the busy probability of the regular unit. Comparing these marginal changes yields the stated characterization of when excluding $\bar{S}$ decreases Q_S.

A.6 Analysis of Example 4.2

(1). For the deterministic LP (11), when the service time is location independent, the constraint (11c) can be simplified as $\sum_{j=1}^{J} x_{1j} \leq B\mu$. To show $x^*_{0j} = 0$, it is equivalent to show that we can find a feasible solution that contains $x_{0j} = 0$, for all j. It holds that

$$x_{0j} = 0 \Leftrightarrow \sum_{j=1}^{J} x_{1j} = \Lambda \Leftrightarrow \Lambda \leq B\mu.$$

(2). For the Lagrangian LP (9), when the service time is location independent, the Lagrangian LP (9) can be simplified as

$$\underset{u(s,j,a) \geq 0}{\text{minimize}} \quad \sum_{j=1}^{J} \frac{\lambda_j}{\Lambda} C_j + \sum_{j=1}^{J} (c_{1j} - C_j) u(0,j,1) \tag{19a}$$

subject to

$$\sum_{j=1}^{J} [u(1,j,0) + u(0,j,1) + u(0,j,0)] = B, \tag{19b}$$

$$u(0,j,1) \leq \frac{\lambda_j}{\Lambda}, \ \forall j, \tag{19c}$$

$$u(0,j,1) + u(0,j,0) = \frac{\lambda_j}{\Lambda} \sum_{j'=1}^{J} \left[\frac{\mu}{\Lambda+\mu} \big(u(1,j',0) + u(0,j',1) \big) + u(0,j',0) \right], \ \forall j, \tag{19d}$$

$$u(1,j,0) = \frac{\lambda_j}{\Lambda} \sum_{j'=1}^{J} \left[\frac{\Lambda}{\Lambda+\mu} \big(u(0,j',1) + u(1,j',0) \big) \right], \ \forall j, \tag{19e}$$

To show $u^*(0,j,1) = \frac{\lambda_j}{\Lambda}$ for all j, it is equivalent to show that we can find a feasible solution that contains $u(0,j,1)=\frac{\lambda_j}{\Lambda}$ for all j. From constraint (19d), we can obtain that

$$\begin{aligned}u(0,j,1)+u(0,j,0)=&\frac{\lambda_j}{\Lambda}\cdot\frac{\mu}{\Lambda+\mu}\sum_{j'=1}^{J}\left[u(1,j',0)+u(0,j',1)+u(0,j',0)\right]+\frac{\lambda_j}{\Lambda+\mu}\sum_{j'=1}^{J}u(0,j',0)\\=&\frac{\lambda_j}{\Lambda+\mu}\cdot\frac{B\mu}{\Lambda}+\frac{\lambda_j}{\Lambda+\mu}\sum_{j'=1}^{J}u(0,j',0).\end{aligned}$$

Taking summation over j, it holds that $\sum_{j=1}^{J}u(0,j,1)=\frac{\mu}{\Lambda+\mu}\left[B-\sum_{j=1}^{J}u(0,j,0)\right]$. From constraint (19e), we can obtain that $u(1,j,0)=\frac{\lambda_j}{\Lambda+\mu}\left[B-\sum_{j'=1}^{J}u(0,j',0)\right]$. Taking summation over j, it holds that $\sum_{j=1}^{J}u(1,j,0)=\frac{\Lambda}{\Lambda+\mu}\left[B-\sum_{j=1}^{J}u(0,j,0)\right]$.

Then, if there exists a feasible solution to the Lagrangian LP (9) that contains $u(0,j,1)=\frac{\lambda_j}{\Lambda}$ for all j, we can obtain that $\sum\limits_{j=1}^{J}u(0,j,1)=1$, and $\sum\limits_{j=1}^{J}u(0,j,0)=\frac{1}{\mu}\left[(B-1)\mu-\Lambda\right]$. It implies that for any j,

$$\begin{aligned}u(0,j,0)&=\frac{\lambda_j}{\Lambda+\mu}\cdot\left[\frac{B\mu}{\Lambda}+\frac{(B-1)\mu-\Lambda}{\mu}\right]-\frac{\lambda_j}{\Lambda},\\u(1,j,0)&=\frac{\lambda_j}{\Lambda+\mu}\cdot\left[B-\frac{(B-1)\mu-\Lambda}{\mu}\right].\end{aligned}$$

Therefore, for any $j=1,\dots,J$,

$$\begin{aligned}u(0,j,1)&=\frac{\lambda_j}{\Lambda},\\u(0,j,0)&=\frac{\lambda_j}{\Lambda+\mu}\cdot\left[\frac{B\mu}{\Lambda}+\frac{(B-1)\mu-\Lambda}{\mu}\right]-\frac{\lambda_j}{\Lambda},\\u(1,j,0)&=\frac{\lambda_j}{\Lambda+\mu}\cdot\left[B-\frac{(B-1)\mu-\Lambda}{\mu}\right],\end{aligned}$$

is a feasible solution to the Lagrangian LP (9) if and only if $u(s,j,a)\geq 0\Leftrightarrow(B-1)\mu\geq\Lambda$.
(3). The loss process of the regular station is a delayed regenerative process. Formally, define the regeneration epochs

$$\begin{aligned}&T_0=0;\\&T_k=\inf\left\{t>T_{k-1}:X(t^-)=\theta B-1,\text{ and a request arrives at }t\text{ such that }X(t)=\theta B\right\},\ k\geq 1,\end{aligned}$$

where $X(t)$ is the number of busy regular units at time t. Then, for $k\geq 2$, the intervals $\mathcal{I}_k=[T_{k-1},T_k)$, form i.i.d. cycles. Each cycle splits into two independent phases:

- *Blocking period* $\mathcal{L}_k$: the time spent in state θB until the first service completion;
- *Return period* $\mathcal{R}_k$: the time from that first drop to $\theta B-1$ until the next entry into state θB.

Within each cycle, only those arrivals occurring during the blocking period $\mathcal{L}_k$ are lost and instead routed into the flexible unit. Consequently, the loss process during the blocking period follows a Poisson process with rate $\theta\Lambda$. Denoting L_k as the length of the blocking period, and ℓ_k as the number of lost arrivals in cycle k. L_k is the minimum of θB independent $\text{Exp}(\mu)$ service-completion times, and hence $\mathbb{E}\left[L_k\right]=\frac{1}{\theta B\mu}$, which indicates that $\mathbb{E}\left[\ell_k\right]=\frac{\Lambda}{B\mu}$.

Besides the number of lost arrivals in each cycle, the loss rate also depends on the average length of the cycle. Denote R_k as the length of the return period $\mathcal{R}_k$ in cycle k, then R_k is the time from the first drop below θB until the next hitting of θB. By the renewal-reward theory (e.g., Ross 1996), the steady-state loss rate, denoted by $\bar{\ell}$, can be expressed as

$$\bar{\ell} = \frac{\mathbb{E}[\ell_k]}{\mathbb{E}[T_k - T_{k-1}]} = \frac{\frac{\theta\Lambda}{\theta B\mu}}{\mathbb{E}[L_k] + \mathbb{E}[R_k]} = \theta\Lambda \cdot \left[\frac{\mathbb{E}[L_k]}{\mathbb{E}[L_k] + \mathbb{E}[R_k]}\right] = \theta\Lambda Q(\theta\Lambda, \theta B).$$

Moreover, following the proof of Theorem 3.3, it holds that $\bar{\ell} = \theta\Lambda Q(\theta\Lambda, \theta B) \leq \Lambda\sqrt{\frac{2\theta}{\pi B}} + O(1)$.

B Approximate Dynamic Programming with Value Function Approximation (VFA)

For each state $(\boldsymbol{s}, j)$, we approximate the optimal relative value function $V^*(\boldsymbol{s}, j)$ by a linear combination of the basis functions, i.e., $V^*(\boldsymbol{s}, j) \approx \alpha_0 + \sum_{k=1}^{K} \alpha_k \phi_k(\boldsymbol{s}, j)$. We next show how we construct six basis functions following Nasrollahzadeh et al. (2018).

B.1 Basis Functions

We first introduce some useful notations for brevity. We say a sub-region j is covered by a station i, if the response time is less than or equal to a threshold Δ, i.e., $c_{ij} \leq \Delta$. Let $M_j(\boldsymbol{s})$ denote the set of available ambulances that cover sub-region j, i.e., $M_j(\boldsymbol{s}) = \{(i,k) | s_{ik} = 0, c_{ij} \leq \Delta\}$, and $N_j(\boldsymbol{s})$ denote the number of available ambulances that cover sub-region j when the state of the system is $\boldsymbol{s}$, i.e., $N_j(\boldsymbol{s}) = |M_j(\boldsymbol{s})|$.

Response Time. This basis function estimates the expected response time of a call from sub-region j, when the state of the system is $(\boldsymbol{s}, j)$. As the DM is required to take immediate action by dispatching an ambulance to carry out the rescue operation, if there are no available ambulances covering the sub-region, it will incur a C_j cost for the flexible unit. We then define the basis function in regard to response time as $\phi_1(\boldsymbol{s}, j) = \mathbb{1}_{\{M_j(\boldsymbol{s}) = \emptyset\}} C_j + \mathbb{1}_{\{M_j(\boldsymbol{s}) \neq \emptyset\}} \left[\min_{(i,k) \in M_j(\boldsymbol{s})} c_{ij}\right]$.

Future Response Time This basis function estimates the future response time of the next arriving call. Specifically, we define $\phi_2(\boldsymbol{s}, j) = \mathbb{E}[\phi_1(\boldsymbol{s}', j') | (\boldsymbol{s}, j, \tilde{\boldsymbol{a}})]$, where $\tilde{\boldsymbol{a}}$ satisfies if $M_j(\boldsymbol{s}) \neq \emptyset$, then dispatch the unit in $M_j(\boldsymbol{s})$ with the minimum cost, else, dispatch the flexible unit.

Uncovered Calls The third basis function describes whether the arriving call can be covered currently given the state $\boldsymbol{s}$. We define the basis function regarding the uncovered calls as $\phi_3(\boldsymbol{s}, j) = \mathbb{1}_{\{M_j(\boldsymbol{s}) = \emptyset\}}$.

Future Uncovered Calls The fourth basis function describes the conditional probability of whether the next arriving call cannot be covered. We have $\phi_4(\boldsymbol{s}, j) = \mathbb{P}\left(M_{j'}(\boldsymbol{s'}) = \emptyset \middle| \boldsymbol{s}, j, \tilde{\boldsymbol{a}}\right)$.

Unreachable Calls The fifth basis function quantifies the number of calls for which an ambulance is assigned but fails to reach the scene within the time threshold Δ, i.e., $\phi_5(\boldsymbol{s}, j) = \sum_{i=1}^{N} \sum_{k=1}^{B_i} \mathbb{1}_{\{s_{ik} \neq 0,\ c_{i,s_{ik}} > \Delta\}}$, where $s_{ik} \neq 0$ indicates that ambulance k from station i is currently servicing a call from sub-region s_{ik}, and $c_{i,s_{ik}}$ means the response time from station i to sub-region s_{ik} assuming $s_{ik} \neq 0$.

Aggregated Delay Time The final basis function calculates the aggregated delay time for calls assigned to an ambulance, where the response time exceeds the time threshold Δ, i.e., $\phi_6(\boldsymbol{s}, j) = \sum_{i=1}^{N} \sum_{k=1}^{B_i} \mathbb{1}_{\{s_{ik} \neq 0\}} \mathbb{1}_{\{c_{i,s_{ik}} > \Delta\}} \left(c_{i,s_{ik}} - \Delta\right)$.

B.2 Approximate Policy Iteration Algorithm

We next show how to use an approximate policy iteration algorithm (see Algorithm 1) proposed in Nasrollahzadeh et al. (2018) to obtain the optimal weight $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_6)$. Note that each VFA policy corresponds to a set of $\boldsymbol{\alpha}$. Therefore, to update $\boldsymbol{\alpha}$, Nasrollahzadeh et al. (2018) proposed a linear model, i.e., to estimate α, solving the following optimization problem

$$\min_{\boldsymbol{\alpha}} \sum_{(s,j) \in \hat{S}} \sum_{r=1}^{R_s} \left(C^r(\boldsymbol{s}, j) - \alpha_0 - \sum_{k=1}^{K} \alpha_k \phi_k(\boldsymbol{s}, j) \right)^2, \tag{20}$$

where $C^r(\boldsymbol{s}, j)$ denotes the relative value function of the realization of the system for initial state $(\boldsymbol{s}, j)$ in replication r, i.e., the simulated relative value function for state $(\boldsymbol{s}, j)$. To simulate $C^r(\boldsymbol{s}, j)$, we use the finite horizon cost to approximate the relative value function (details are shown in Appendix C.3). Let R_s denote the total number of replications of the Monte Carlo simulation for state $(\boldsymbol{s}, j)$, and $\hat{S}$ denote a pre-determined set of states under which the policy will be evaluated in each iteration.

Algorithm 1: Approximate policy iteration

Set $n = 1$, $n_{\max} > 0$, $\epsilon > 0$ and $\alpha = \alpha^0$.
while $\|\hat{V}^n(\boldsymbol{s}, j) - \hat{V}^{n-1}(\boldsymbol{s}, j)\|_2 > \epsilon$ *and* $n < n_{\max}$ **do**
  Policy improvement: Find a myopic policy induced by $\hat{V}^n(\boldsymbol{s}, j)$ by solving formulation (14).
  Policy evaluation: Use Monte Carlo to simulate the system; find actions for each state visited by the simulation via solving formulation (14); calculate $C^r(\boldsymbol{s}, j)$, the relative value function for the initial state $(\boldsymbol{s}, j)$ and replication r.
  Projection: Use $C^r(\boldsymbol{s}, j)$ from the Monte Carlo simulation and solve formulation (20) to estimate α^{n+1} for the next iteration.
  Set $n \leftarrow n + 1$
end

B.3 Training Details

We report the training details of the VFA policy. We first generate a fixed evaluation set by uniformly sampling 50 states from the state space and use this set, denoted by $\hat{S}$, throughout training to evaluate policy updates. At iteration k, we update the coefficient of basis functions from $\boldsymbol{\alpha}^{k-1}$ to $\boldsymbol{\alpha}^k$ through the following three steps: (i) *Policy improvement*: We compute a myopic policy based on the expected value-to-go function, $\mathbb{E}\left[\hat{V}^k(\boldsymbol{s}, j)\right]$. The expectation is approximated via Monte Carlo simulation with 50 samples. Specifically, starting from state $\boldsymbol{s}$ and taking action $\boldsymbol{a}$, we simulate one step to obtain the next state $\boldsymbol{s}'$ and evaluate the cost-to-go using $\hat{V}^k(\boldsymbol{s}', j) = \alpha_0 + \sum_{\ell=1}^{6} \alpha_\ell \phi_\ell(\boldsymbol{s}', j)$, where $\phi_\ell(\cdot, \cdot)$ denote the predefined basis functions. (ii) *Policy evaluation*: For each state in $\hat{S}$, we evaluate the resulting myopic policy and approximate the relative value $C(\boldsymbol{s}, j)$ by the average cost over a 50-period rollout. (iii) *Projection*: we update the coefficient vector $\boldsymbol{\alpha}^k$ by solving the linear regression problem

$$\min_{\boldsymbol{\alpha}} \sum_{(s,j)\in\hat{S}} \sum_{r=1}^{R_s} \left(C^r(\boldsymbol{s}, j) - \alpha_0 - \sum_{\ell=1}^{6} \alpha_\ell \phi_\ell(\boldsymbol{s}, j) \right)^2,$$

that projects the simulated value estimates onto the basis-function space.

C Supplementary Materials of Sections 4 and 5

C.1 Numerical Settings and Sensitivity Analysis

We summarize the parameter settings used in Sections 4 and 5. Specifically, the capacity at each station (B_i) is fixed at 3 when the scaling parameter equals 1, and the service rates (μ_{ij}) are specified in Table 5.

μ_{ij}	$i = 1$	$i = 2$
$j = 1$	26	34
$j = 2$	39	30
$j = 3$	48	27

Table 5: Service Rate μ_{ij}

We consider two sets of arrival rates, $(\lambda_1^{\mathrm{L}}, \lambda_2^{\mathrm{L}}, \lambda_3^{\mathrm{L}}) = (27, 27, 22)$ and $\lambda_j^{\mathrm{H}} = 3\lambda_j^{\mathrm{L}}$, representing light- and heavy-traffic regimes, respectively. We also consider two levels of flexible-unit costs, $C_j^{\mathrm{L}} = 41$ and $C_j^{\mathrm{H}} = 3C_j^{\mathrm{L}}$, to assess the impact of cost differences between regular and flexible units. For Figure 4, Table 1, and Table 2, we use heavy traffic λ_j^{H} with low flexible-unit cost C_j^{L}. For Figure 3, we use light traffic λ_j^{L} with low flexible-unit cost C_j^{L}. Table 6 reports performance across all parameter settings.

We identify scenarios in which the proactive deployment of flexible units delivers performance gains as follows. Under heavy traffic, particularly in larger systems, proactive use of flexibility yields gains across all policies, and these gains are robust to the level of flexible-unit cost. Under light traffic or in smaller-scale systems, additional gains remain attainable but typically require more targeted, state-dependent deployment of flexible units, as illustrated by resolving-SAP and VFA. When traffic is light and flexible units are inexpensive, the myopic policy is already near-optimal, and proactive deployment provides limited incremental value.

We also observe that when the cost of flexible unit is high, the optimality gap relative to the DLP bound becomes particularly large, especially in small-scale systems. The main reason is that the deterministic fluid relaxation is looser in this regime. In the fluid system, regular units are sufficient to serve all the requests in expectation, so the optimal solution often avoids using the flexible unit, i.e., $x_{0j} = 0$ for all j. In the actual system, however, random fluctuations in arrivals and unit availability make the use of the flexible unit unavoidable, even when its average utilization is relatively low. As a result, the DLP bound significantly underestimates the true operating cost in the low-traffic regime.

C.2 Reformulation of the Original Problem and its Optimal Policy

To facilitate computing the optimal policy in a small-scale problem, we start by reformulating the problem as follows. Let $\boldsymbol{m}_i = (m_{i1}, \ldots, m_{iJ})$ represent the state of station i, where m_{ij} denotes the number of units in station i that are serving calls from sub-region j. Note that if $\sum_{j=1}^{J} m_{ij} = B_i$, it indicates that there are no available units in station i. We denote the state space of all stations

Total Number of Regular Units			12 units ($\theta = 2$)			48 units ($\theta = 8$)			192 units ($\theta = 32$)		
Penalty Cost	Arrival Rate	Policy	GAP	#Reg	#Flex	GAP	#Reg	#Flex	GAP	#Reg	#Flex
C_j^{L}	λ_j^{L}	SAP-DLP	1.93	5.04	0.43	0.09	18.3	0.04	0.03	68.31	0
		SAP-LA	1.93	5.04	0.43	0.09	18.30	0.04	0.03	68.31	0
		RS-DLP	0.67	5.37	0.005	0.05	18.08	0	0.03	68.31	0
		VFA	2.42	5.20	0.10	2.02	16.88	0.29	1.89	68.40	1.13
		Myopic	0.67	5.36	0.007	0.04	17.46	0	0.03	66.57	0
	λ_j^{H}	SAP-DLP	12.99	9.54	3.87	6.50	41.48	7.71	2.89	175.24	12.90
		SAP-LA	11.45	8.88	4.66	6.30	41.03	8.32	3.46	176.21	12.90
		RS-DLP	11.17	9.89	3.57	6.16	42.53	7.32	2.22	176.30	12.23
		VFA	13.81	10.57	3.48	9.03	44.54	9.79	5.52	184.83	20.96
		Myopic	14.52	10.72	3.76	8.75	44.67	8.70	5.28	184.38	19.61
C_j^{H}	λ_j^{L}	SAP-DLP	18.20	5.04	0.43	0.62	18.3	0.04	0.03	68.31	0
		SAP-LA	18.20	5.04	0.43	0.62	18.3	0.04	0.03	68.31	0
		RS-DLP	1.04	5.37	0.005	0.05	18.08	0	0.03	68.31	0
		VFA	2.43	5.18	0.004	1.72	17.74	0	1.41	69.61	0
		Myopic	1.12	5.36	0.007	0.05	17.46	0	0.03	66.57	0
	λ_j^{H}	SAP-DLP	91.40	9.54	3.86	46.89	41.48	7.71	20.59	175.24	12.90
		SAP-LA	102.02	8.88	4.66	49.23	41.03	8.32	23.88	176.21	14.53
		RS-DLP	90.54	9.75	3.90	44.76	41.70	7.33	21.07	176.50	13.27
		VFA	86.39	10.65	3.29	56.42	44.75	8.22	32.28	185.75	19.97
		Myopic	90.02	10.72	3.76	53.45	44.67	8.71	31.46	184.38	19.61

Table 6: Sensitivity Analysis. GAP: relative gap (%) between the policy performance and the lower bound derived from the deterministic fluid LP; #Reg (#Flex): average number of busy regular (flexible) units.

as vector $\boldsymbol{m} = (\boldsymbol{m}_1, \dots, \boldsymbol{m}_N)$. The system state is defined by the states of all stations $\boldsymbol{m}$, and the arriving call j. Given a system state, the DM must decide which available unit to dispatch to serve the arriving call. Formally, when a call from sub-region j arrives, let the actions selected for all stations be $\boldsymbol{a} = (a_1, \dots, a_N)$, where $a_i = 1$ if the DM decides to dispatch the unit from this station and $a_i = 0$, otherwise. Denote the action space in state $\boldsymbol{m}$ by $\widehat{\boldsymbol{A}}(\boldsymbol{m}) = \otimes_{i=1}^N \hat{A}_i(\boldsymbol{m}_i)$, where $\hat{A}_i(\boldsymbol{m}_i) = \left\{ a \in \{0,1\} : \sum_{j=1}^J m_{ij} + a \le B_i \right\}$. Note that for each arrival, the call can be assigned to only one available unit (including flexible units). We denote the set of feasible actions at the decision epoch when an arrival occurs as

$$\widehat{\mathcal{A}}(\boldsymbol{m}) = \left\{ \boldsymbol{a} \in \widehat{\boldsymbol{A}}(\boldsymbol{m}) : \sum_{i=1}^N a_i \le 1 \right\}.$$

As service times for each unit follow independent exponential distributions, the component of $\boldsymbol{m}_i$ evolves randomly according to a binomial distribution. Formally, denote $\boldsymbol{m}'_i$ as the state of station i at the beginning of the next period. Then, for an arrival call from sub-region j, m'_{ij} follows the Binomial distribution with parameter $n = m_{ij} + a_i, p = \frac{\Lambda}{\mu_{ij}+\Lambda}$, and for $j' \ne j$, $m'_{ij'}$ follows the Binomial distribution with parameter $n = m_{ij}, p = \frac{\Lambda}{\mu_{ij}+\Lambda}$.

Note that for all initial states and all policies, the system state $(\boldsymbol{0}, j)$ for any j is always visited with positive probability in any decision epoch, which implies that every policy is unichain. Therefore, by standard DP arguments (e.g., Bertsekas 2017; Puterman 1994), the following holds: The optimal average cost c^* is the same for all initial states, and with optimal relative value functions $V^*(\boldsymbol{m}, j)$, it satisfies Bellman's equations for any state $(\boldsymbol{m}, j)$,

$$V^*(\boldsymbol{m}, j) + c^* = \min_{\boldsymbol{a} \in \widehat{\mathcal{A}}(\boldsymbol{m})} \left\{ \sum_{i=1}^N c_{ij} a_i + C_j \left(1 - \sum_{i=1}^N a_i \right) + \mathbb{E}\left[V^*(\boldsymbol{m}', j') \mid (\boldsymbol{m}, j, \boldsymbol{a}) \right] \right\} \tag{21}$$

Consider that the size of the state space of the reformulated problem in (21) is given by $J \times \prod_{i=1}^N \sum_{b=0}^{B_i} \binom{b+J-1}{J-1}$, whereas the size of the state space of the original problem (4) is $J \times (J+1)^{\mathcal{B}}$, where $\mathcal{B} = \sum_{i=1}^{N} B_i$. It is noteworthy that when B_i is relatively small, the former state space is significantly smaller than the latter. For example, in the case of our numerical setting $N = 2, J = 3$ and $B_1 = B_2 = 3$, $\prod_{i=1}^{N} \sum_{b=0}^{B_i} \binom{b+J-1}{J-1} = 400$, whereas $(J+1)^{\mathcal{B}} = 4096$. Consequently, the reformulated problem in (21) facilitates a more efficient way for determining the optimal policy.

It is worth noting that the reformulated problem remains a weakly coupled dynamic program framework, where all stations are coupled through the linking constraint $\sum_{i=1}^{N} a_i \leq 1$. However, the Lagrangian relaxation method previously proposed is no longer applicable to the reformulated problem, as the Lagrangian dual problem remains intractable for large B_i, due to the combinatorial growth in the size of the state space of each subproblem as B_i increases.

C.3 Approximate One-step Improvement Policy

In Section 5, we introduce the one-step improvement policy based on the SAP-LA (OSI-LA), which improves upon SAP-LA by first estimating the relative value function $V^{\pi}(\boldsymbol{s}, j)$ across all states and then, for each state $(\boldsymbol{s}, j)$, selecting the feasible action that minimizes

$$\min_{\boldsymbol{a} \in \mathcal{A}(\boldsymbol{s})} \left\{ \sum_{i=1}^{N} \sum_{k=1}^{B_i} c_{ij} a_{ik} + C_j \left(1 - \sum_{i=1}^{N} \sum_{k=1}^{B_i} a_{ik} \right) + \mathbb{E} \left[V^{\pi}(\tilde{\boldsymbol{s}}, \tilde{j}) \middle| (\boldsymbol{s}, j, \boldsymbol{a}) \right] \right\}.$$

The main implementation challenges for large-scale systems arise from: (i) evaluating the relative value function $V(\boldsymbol{s}, j)$, and (ii) computing the expectation over next-period states. To address this challenge, we (i) replace the full expectation with a Monte Carlo simulation of one-step-ahead states, and (ii) estimate the relative value function using the finite horizon cost under SAP-LA. Specifically, by standard DP arguments (e.g., Section 8.2 Puterman 1994), for any state $(\boldsymbol{s}, j)$

$$V^{\pi}(\boldsymbol{s}, j) = \lim_{T \to \infty} \mathbb{E} \left(\sum_{t=1}^{T} [c_t(\boldsymbol{s}_t^{\pi}, \boldsymbol{a}_t^{\pi}, j_t) - c^{\pi}] \middle| (\boldsymbol{s}_0, j_0) = (\boldsymbol{s}, j) \right),$$

where $c_t(\boldsymbol{s}_t, \boldsymbol{a}_t, j_t)$ denotes the per-period cost given state and action $(\boldsymbol{s}_t, \boldsymbol{a}_t, j_t)$. We select a sufficiently large finite horizon T_0 such that, for any state $(\boldsymbol{s}, j)$,

$$V^{\pi}(\boldsymbol{s}, j) + T_0 c^{\pi} \approx \mathbb{E} \left(\sum_{t=1}^{T_0} c_t(\boldsymbol{s}_t^{\pi}, \boldsymbol{a}_t^{\pi}, j_t) \middle| (\boldsymbol{s}_0, j_0) = (\boldsymbol{s}, j) \right).$$

This transforms the original optimization problem to

$$\min_{\boldsymbol{a} \in \mathcal{A}(\boldsymbol{s})} \left\{ \sum_{i=1}^{N} \sum_{k=1}^{B_i} c_{ij} a_{ik} + C_j \left(1 - \sum_{i=1}^{N} \sum_{k=1}^{B_i} a_{ik} \right) + \mathbb{E} \left[V^{\pi}(\tilde{\boldsymbol{s}}, \tilde{j}) + T_0 c^{\pi} \middle| (\boldsymbol{s}, j, \boldsymbol{a}) \right] \right\}.$$

Thus, to approximately evaluate the relative value function of the sampled states, we simulate SAP-LA over T_0 periods starting from these states. The finite-horizon cost approximation is a standard technique for estimating the cost-to-go function in infinite-horizon settings (Section 6 of Bertsekas and Tsitsiklis 1996).