Two-Stage Distributionally Robust Edge Node Placement Under Endogenous Demand Uncertainty

Jiaming Cheng, Duong Thuy Anh Nguyen, Duong Tung Nguyen

School of Electrical, Computer and Energy Engineering, Arizona State University

February 7, 2025

Outline

Overview of edge computing (EC) system

- **EC platform**: manages a set of heterogeneous edge resources (e.g., Edge nodes = ENs).
- **User**: requests are aggregated by nearby access points (e.g., Wi-Fi routers, base stations)

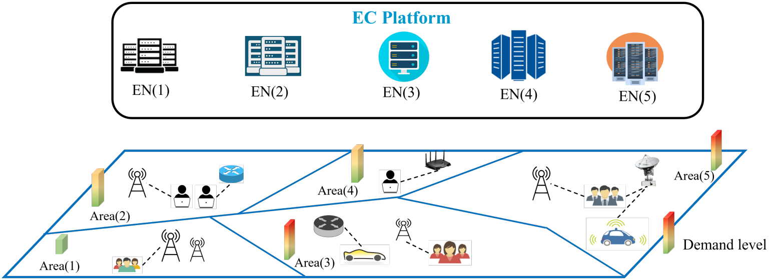

Figure: overview of edge computing (EC) system

Overview of edge computing (EC) system

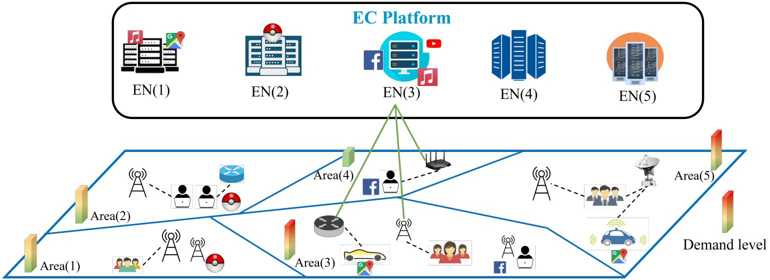

Figure: overview of edge computing (EC) system

- **Operations**: service placement, workload allocation, pricing
- **Planning**: network design, edge node placement
- **Uncertainties**: extreme weather conditions, component failures, fluctuating resource demand, user mobility, ...

Focus of today's paper

This work investigates edge node placement and resource allocation problem

The EC platform (e.g., Equinix) may decide:

- **EN placement decisions**: install ENs from a set of potential candidate locations.
- **Resource allocation decisions**: equip an appropriate amount of edge resources, given the diverse range of IoT services with varying requirements.
- **Uncertain demand**

Toy example

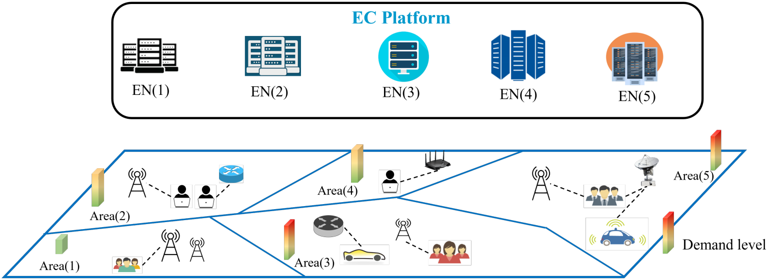

Figure: Toy example

We consider an EC system that consists of

- 5 candidate EN locations, one in each area (city metropolitan level)
- Uncertain demand: Users' demand in each area

Toy example

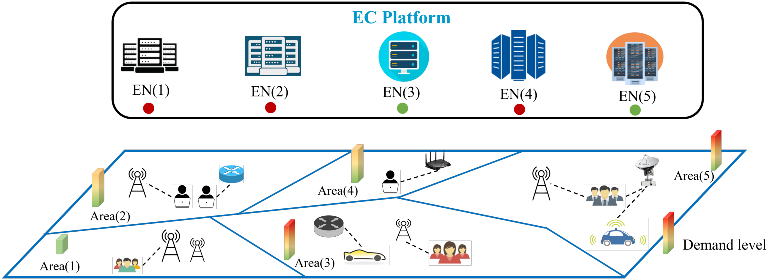

Figure: Toy example

- **EN placement**: place EN 3 and EN 5

Toy example

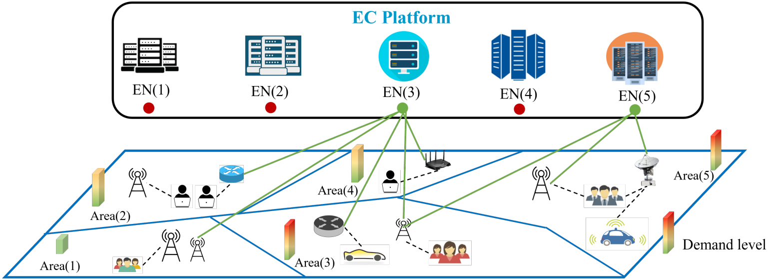

Figure: Toy example

Resource allocation decisions

- Workload from area 1, 2, 4 is allocated to EN 3; Workload from area 5 is allocated to EN 5.
- Portion of workload from area 3 is served by EN 3 and the rest of workload is served by EN 5.

Challenges

Uncertainty

Managing uncertainties is a key factor in achieving consistent performance, and superior user experience in EC.

Challenges

Uncertainty

Managing uncertainties is a key factor in achieving consistent performance, and superior user experience in EC.

Many efforts within the realm of optimization under uncertainty have been developed for EC:

- Stochastic optimization (SO): assume complete knowledge of the underlying uncertainty distribution (true distribution); require a large number of samples.

Challenges

Uncertainty

Managing uncertainties is a key factor in achieving consistent performance, and superior user experience in EC.

Many efforts within the realm of optimization under uncertainty have been developed for EC:

- Stochastic optimization (SO): assume complete knowledge of the underlying uncertainty distribution (true distribution); require a large number of samples.
- Robust optimization (RO): use a parametric set to represent uncertain parameters; robust solution can be overly conservative

Related work

Our approach

Distributionally robust optimization (DRO) optimizes decisions w.r.t worst-case distribution within a *predefined ambiguity set*.

- Moment-based ambiguity set: [Ye, 2010]
- Wasserstein-metric ambiguity set: [Kuhn, 2018]
- χ-divergence ambiguity set: [Yu, 2024]

Related work: resource management under

- **Demand uncertainty:** [Liang *et al.* 2018], [Zhang *et al.* 2020], [Chen *et al.* 2021], [Li *et al.* 2022].
- **Delay uncertainty:** [Cui *et al.* 2023]
- **Others:** renewable energy [Zhou *et al.* 2021], risk [Li, *et al.* 2023]

Motivation: endogeneity

Interdependence between decisions and uncertainty

Some random factors are substantially affected by the choice of decision, therefore referred to as **decision-dependency** or **endogeneity**. For example,

- Production decisions serve as not only an instruction to produce but also an investment to refine the information on the production cost.
- System reliability or failure rate change with respect to maintenance decision.

Motivation: endogeneity

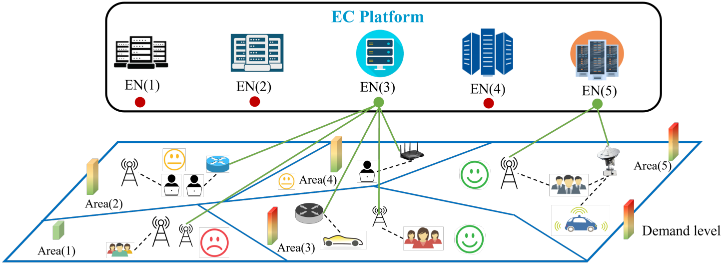

Figure: System model

- **Increased mean of demand**: *The presence of more ENs*, along with increased resource availability and reduced network delay, improving user confidence.

Motivation: endogeneity

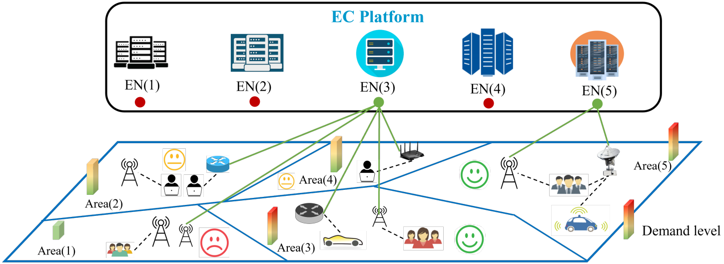

Figure: System model

- **Increased mean of demand**: *The presence of more ENs*, along with increased resource availability and reduced network delay, improving user confidence.
- **Decreased demand variance**: As more users become more confident in the reliability and availability of edge resources, their demand patterns tend to become more consistent and predictable.

Motivation

Research question:

- How to quantitatively capture this endogeneity between uncertainties and decisions?
- What are the benefits of capturing this interdependency?

Contribution

Our contributions can be summarized in three folds:

- **Modeling**: propose a novel two-stage DRO framework with a decision-dependent moment-based ambiguity set for optimal EN placement.

Contribution

Our contributions can be summarized in three folds:

- **Modeling**: propose a novel two-stage DRO framework with a decision-dependent moment-based ambiguity set for optimal EN placement.
- **Techniques**:
    - (i) develop an efficient and exact reformulation to convert the two-stage problem into a mixed integer linear programming
    - (ii) introduce an improved algorithm that generates feasibility cuts to speed up the computation.

Contribution

Our contributions can be summarized in three folds:

- **Modeling**: propose a novel two-stage DRO framework with a decision-dependent moment-based ambiguity set for optimal EN placement.
- **Techniques**:
  - (i) develop an efficient and exact reformulation to convert the two-stage problem into a mixed integer linear programming
  - (ii) introduce an improved algorithm that generates feasibility cuts to speed up the computation.
- **Numerical results**: demonstrate the efficiency of the proposed model compared to the state of art.

System model

- λ_i: Resource demand (uncertain) in area i
- $y_j \in \{0, 1\}$: EN placement decision for an EN at location j
- $x_{i,j}$: workload generated in area i allocated to EN j
- s_i: unmet workload at each area i

Uncertainty modeling

We mainly focus on the moment-based ambiguity set, where only the mean and variance of the demand distribution are provided.

- The true distribution of demand originates from a set of possible distributions, where demand λ_i can take any value from a finite support set $\Xi = \{\xi_1, \xi_2, \ldots, \xi_N\}$ with unknown probabilities $(p_{i,1}, p_{i,2}, \ldots, p_{i,N})$

Uncertainty modeling

(1) **Exogenous stochastic demand**: when the demand is independent of the EN placement decision.

$$\mathcal{U}_1 = \Big\{ \{p\}_{i\in\mathcal{I}} : \ p_i \in \mathbb{R}_+^N, \ \sum_{n=1}^{N} p_{i,n} = 1, \ \forall i, \tag{1a}$$

$$|\sum_{n=1}^{N} p_{i,n}\xi_n - \bar{\mu}_i| \leq \Gamma_i^{\mu}, \ \forall i, \tag{1b}$$

$$(\bar{\sigma}_i^2 + \bar{\mu}_i^2)\underline{\Gamma}_i^{\sigma} \leq \sum_{n=1}^{N} p_{i,n}\xi_n^2 \leq (\bar{\sigma}_i^2 + \bar{\mu}_i^2)\bar{\Gamma}_i^{\sigma}, \ \forall i \ \Big\}. \tag{1c}$$

- (1a): probability across all areas within the support set sum up to 1
- (1b): the true mean of demand lies within an $L1$-distance Γ_i^{μ} from the empirical mean $\bar{\mu}_i$
- (1c): the actual second moment of demand must fall within the interval $[(\bar{\sigma}_i^2 + \bar{\mu}_i^2)\underline{\Gamma}_i^{\sigma}, (\bar{\sigma}_i^2 + \bar{\mu}_i^2)\bar{\Gamma}_i^{\sigma}]$.

Uncertainty modeling

(2) **Endogenous stochastic demand:** when the demand is dependent of the EN placement decision.

$$\mathcal{U}_2(y) = \Big\{ \{p_i\}_{i\in\mathcal{I}}: \ p_i \in \mathbb{R}_+^N, \ \sum_{n=1}^{N} p_{i,n} = 1, \ \forall i, \tag{2a}$$

$$|\sum_{n=1}^{N} p_{i,n}\xi_n - \mu_i(y)| \leq \Gamma_i^{\mu}, \ \forall i, \tag{2b}$$

$$\big[\sigma_i^2(y)+(\mu_i(y))^2\big]\underline{\Gamma}_i^{\sigma} \leq \sum_{n=1}^{N} p_{i,n}\xi_n^2 \leq \big[\sigma_i^2(y)+(\mu_i(y))^2\big]\bar{\Gamma}_i^{\sigma}, \forall i \Big\}. \tag{2c}$$

Remark

$\mu_i(y)$ and $\sigma_i^2(y)$ are the mean/variance of the demand, defined as a function of EN placement decisions y

Uncertainty set comparison

Endogenous stochastic demand

$$\mu_i(y) = \bar{\mu}_i \left(1 + \sum_{j \in \mathcal{J}} \Psi^{\mu}_{i,j} y_j \right), \tag{3a}$$

$$\sigma_i^2(y) = \max \left\{ \bar{\sigma}_i^2 \left(1 - \sum_{j \in \mathcal{J}} \Psi^{\sigma}_{i,j} y_j \right), (\sigma_i^{LB})^2 \right\}. \tag{3b}$$

- **Closer locations** may have **higher impacts** on demand's first and second moments, while areas farther away have less effect.
- when $\Psi^{\sigma}_{i,j} = \Psi^{\mu}_{i,j} = 0, \ \forall i, j$, the ambiguity set reduces to exogenous ambiguity set.
- In the simulation, for simplicity, we consider decreasing functions of the network delay (e.g., distance), i.e., $\exp(-\frac{d_{i,j}}{b})$.

Problem formulation

The proposed two-stage decision-dependent DRO problem of the EC platform for EN placement and resource allocation is:

$$(\mathcal{P}_1) \quad \min_{y\in\{0,1\}^J} \underbrace{\sum_j f_j y_j}_{(i)} + \max_{p\in\mathcal{U}(y)} \min_{\mathbf{x},\mathbf{u}} \mathbb{E}_p \underbrace{\left[\rho \sum_{i,j} d_{i,j} x_{i,j} + \sum_i s_i u_i\right]}_{(ii)} \tag{4}$$

- (i): total EN placement cost (planning cost).
- (ii): network delay penalty & unmet demand.
- This problem is a tri-level optimization problem (“**min-max-min**”).

Problem formulation

$$\text{s.t.} \quad \Omega_1(y) = \left\{ \sum_{j \in \mathcal{J}} f_j y_j \leq B; \quad \sum_{j \in \mathcal{J}} y_j \geq K^{\text{min}} \right\} \tag{4a}$$

$$\Omega_2(y, \lambda) = \Bigg\{ \; 0 \leq x_{i,j} \leq C_{i,j} y_j, \;\forall i, j \tag{4b}$$

$$u_i + \sum_j x_{i,j} = \lambda_i(y), \;\forall i \tag{4c}$$

$$\sum_j d_{i,j} x_{i,j} \leq \Delta_i \lambda_i(y), \;\forall i \Bigg\}. \tag{4d}$$

- $\Omega_1(y)$: includes the budget and reliability constraints
- $\Omega_2(\lambda, y)$: includes the capacity; supply-demand; delay constraints
- The demand (i.e., $\lambda(y)$) is a function of the first-stage decision in the planning stage.

Exact monolithic reformulation: *Exact OPT-Placement*

Core ideas of the algorithm:

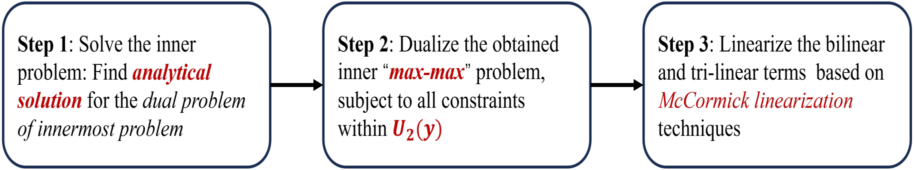

Figure: Flow chart of **Exact OPT-Placement**

Exact monolithic reformulation: *Exact OPT-Placement*

Step 1: Decompose each inner problem based on each area i:
$g(y,\lambda) = \sum_{i \in \mathcal{I}} g_i(y,\lambda)$,i.e.,

$$g_i(y,\lambda) = \min_{\mathbf{x},\mathbf{u}} \rho \sum_j d_{i,j} x_{i,j} + s_i u_i \tag{5a}$$

$$\text{s.t.} \quad x_{i,j} \le C_{i,j} y_j, \ \forall j \qquad (v_{i,j}) \tag{5b}$$

$$u_i + \sum_j x_{i,j} = \lambda_i(y), \qquad (\alpha_i) \tag{5c}$$

$$\sum_j d_{i,j} x_{i,j} \le \Delta_i \lambda_i(y), \qquad (\beta_i) \tag{5d}$$

Exact monolithic reformulation: *Exact OPT-Placement*

The dual problem of $g_i(y, \lambda)$, for area $i \in \mathcal{I}$ is:

$$\max_{v_{i,j},\alpha_i,\beta_i} \quad \sum_j C_{i,j} y_j v_{i,j} + \left[\alpha_i + \beta_i \Delta_i\right] \lambda_i(y) \tag{5a}$$

$$\text{s.t.} \quad v_{i,j} + \alpha_i + \beta_i d_{i,j} \leq \rho d_{i,j}, \ \forall j \tag{5b}$$

$$\alpha_i \leq s_i, \ \beta_i \leq 0; \ \ v_{i,j} \leq 0, \ \forall j. \tag{5c}$$

Goal

- Identify the **closed-form expression** for the optimal objective value of the dual problem in each area, considering the extreme points and rays of the feasible region.
- The second-stage problem converts from (“**max-min**”) to (“**max-max**”).

Exact monolithic reformulation: *Exact OPT-Placement*

Step 2: Dualize the obtained inner "min-max" problem subject to all constraints within $\mathcal{U}_2(y)$. ($\theta_{i,n}(y)$ is obtained from Step 1)

$$\max_{p_{i,n}} \quad \sum_{i\in\mathcal{I}}\sum_{n=1}^{N} p_{i,n}\theta_{i,n}(y) \tag{6a}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} p_{i,n}\xi_n = 1, \ \forall i \quad (\omega_i) \tag{6b}$$

$$\Gamma_i^{\mu} - \mu_i(y) \le \sum_{n=1}^{N} p_{i,n}\xi_n \le \Gamma_i^{\mu} + \mu_i(y), \ \forall i \ (\delta_i^2, \delta_i^1) \tag{6c}$$

$$\left(\sigma_i^2 + (\mu_i(y))^2\right)\underline{\Gamma}_i^{\sigma} \le \sum_{n=1}^{N} p_{i,n}\xi_n^2 \le \left(\sigma_i^2 + (\mu_i(y))^2\right)\bar{\Gamma}_i^{\sigma}, \ \forall i \ (\gamma_i^2, \gamma_i^1) \tag{6d}$$

Exact monolithic reformulation: *Exact OPT-Placement*

Step 2: Dualize the obtained inner "min-max" problem subject to all constraints within $\mathcal{U}_2(y)$. ($\theta_{i,n}(y)$ is obtained from Step 1)

$$\min_{\omega,\delta,\gamma} \quad \sum_i \omega_i + \delta_i^1(\mu_i(y) + \Gamma_i^\mu) - \delta_i^2(\mu_i(y) - \Gamma_i^\mu)$$

$$+ \left(\sigma_i^2(y) + ((\mu_i(y))^2\right)\bar{\Gamma}_i^\sigma \gamma_i^1 - \left(\sigma_i^2(y) + (\mu_i(y))^2\right)\underline{\Gamma}_i^\sigma \gamma_i^2 \tag{6a}$$

$$\text{s.t.} \quad \omega_i + (\delta_i^1 - \delta_i^2)\xi_n + (\gamma_i^1 - \gamma_i^2)\xi_n^2 \geq \theta_{i,n}(y), \ \forall i, n \tag{6b}$$

$$\delta_i^1, \delta_i^2, \gamma_i^1, \gamma_i^2 \geq 0, \ \forall i. \tag{6c}$$

Exact monolithic reformulation: *Exact OPT-Placement*

Step 2: Dualize the obtained inner "min-max" problem subject to all constraints within $\mathcal{U}_2(y)$. ($\theta_{i,n}(y)$ is obtained from Step 1) After step 2,

- $\mathcal{P}_1$ ("*min-max-min*") was reduced to a single-stage minimization problem.
- However, it is still a single-stage mixed integer non-linear programming problem (MINLP).

Exact monolithic reformulation: *Exact OPT-Placement*

Step 3: Linearize the bilinear and trilinear terms by McCormick linearization.

- $\mu_i(y)$ and $\sigma_i^2(y)$ are affine function of the placement decision y.
- Bilinear terms involve the product of binary variables and a non-negative continuous variable $(\kappa^r = \gamma^r y)$.
- To linearized the bilinear term, $\mathcal{M}_{\kappa,y,\gamma}$ denotes the set involving the McCormick inequalities for linearizing any bilinear term, where $y \in \{0,1\}$, and γ^r is non-negative.

$$\mathcal{M}_{\kappa,y,\gamma} = \Big\{(\kappa,\gamma,y):\ \underline{\gamma}^r y \le \kappa^r \le \bar{\gamma}^r y, \underline{\gamma}^r \le \gamma^r \le \bar{\gamma}^r \\ \gamma^r - (1-y)\bar{\gamma}^r \le \kappa^r \le \gamma^r - (1-y)\underline{\gamma}^r\Big\}, \qquad (6)$$

Improved variants

According to **Algorithm 1**, this computation time of solving this large-scale MILP can be sensitive to the network size.

Improved variants

Core idea:

- The problem after **Step 2** is feasible within a region satisfying associated inequalities:

$$\omega_i + \underbrace{(\delta_i^1 - \delta_i^2)}_{\delta_i} \xi_n + \underbrace{(\gamma_i^1 - \gamma_i^2)}_{\gamma_i} \xi_n^2 \geq \theta_{i,n}(y), \ \forall i, n \tag{7a}$$

$$\delta_i^1, \delta_i^2, \gamma_i^1, \gamma_i^2 \geq 0, \ \forall i. \tag{7b}$$

- δ_i and γ_i are unbounded: identifying the extreme points to achieve the optimal objective might be time-consuming.
- The goal is to find a set of extreme rays (ω_i, δ_i^1, δ_i^2, γ_i^1, γ_i^2) that can represent the feasible region defined by (7)

Improved variants

- To identify the extreme rays, we will solve the following inequality systems for $k, l \in \{1, 2, \ldots N\}$.

$$\omega_i + \delta_i \xi_k + \gamma_j \xi_k^2 = 0, \ \forall i, k \tag{8a}$$

$$\omega_i + \delta_i \xi_l + \gamma_j \xi_l^2 = 0, \ \forall i, l \tag{8b}$$

$$\omega_i + \delta_i \xi_n + \gamma_j \xi_n^2 \geq 0, \ \forall n \in \{1, 2, \ldots, N\} \setminus \{l, k\}. \tag{8c}$$

Improved variants

- To identify the extreme rays, we will solve the following inequality systems for $k, l \in \{1, 2, \ldots N\}$.

$$\omega_i + \delta_i \xi_k + \gamma_j \xi_k^2 = 0, \ \forall i, k \tag{8a}$$

$$\omega_i + \delta_i \xi_l + \gamma_j \xi_l^2 = 0, \ \forall i, l \tag{8b}$$

$$\omega_i + \delta_i \xi_n + \gamma_j \xi_n^2 \geq 0, \ \forall n \in \{1, 2, \ldots, N\} \setminus \{l, k\}. \tag{8c}$$

- W.l.o.g, we assume that $\xi_k < \xi_l$. Define $\{\xi_{(1)}, \xi_{(2)}, \ldots, \xi_{(N)}\}$ as a ordered support for the random demand.
- **Goal:** determine the relationship between ξ_k, ξ_l, and the other instances $\xi_n, n \in \{1, 2, \ldots, N\} \setminus \{k, l\}$.

Performance comparison

In this section, we compare the performance of the proposed **DRO-DDU** with the following benchmarks:

- **HEU**: Choose a subset of ENs according to demand, giving priority to areas with higher demand until the available budget is fully utilized.
- **BSPA**: Deploy as many ENs as possible within the budget.
- **DET**: Deterministic EN placement problem.
- **SO**: Two-stage SO with uniform in-sample distribution.
- **DRO-DIU**: $\Psi^{\mu}_{i,j} = \Psi^{\sigma}_{i,j} = 0$. The original problem reduces to a two-stage DRO with exogenous stochastic demand.

Performance analysis

Impacts of the EN placement cost

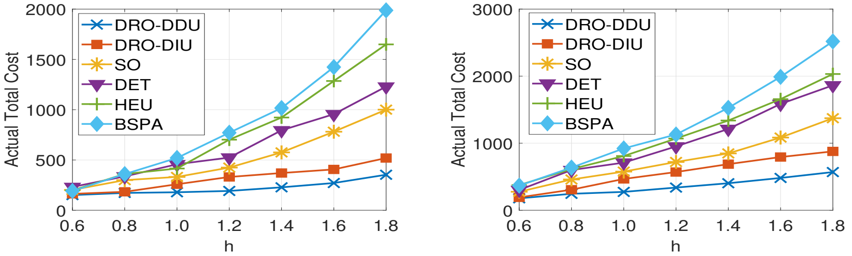

Figure: h_i: Scaling factor for EN placement cost

- **Stability:** DRO-based models show increased stability compared to other schemes, especially with higher h_i.

Performance analysis

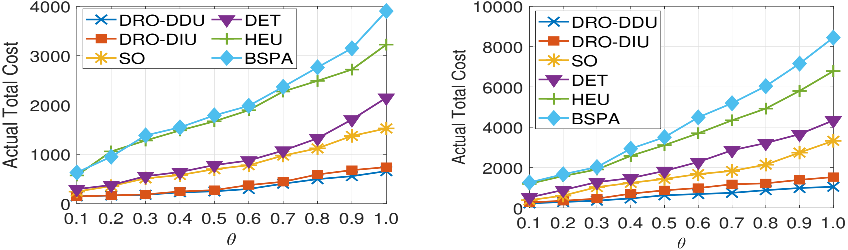

Figure: Ratio of variation at each area: $\theta_i = \frac{\bar{\sigma}_i}{\bar{\mu}_i}, \forall i$

- **Robustness:** as θ increases, the gap between these schemes widens due to the significant deviation of actual demand from its mean.

Sensitivity analysis

Choice of decision-dependency:

- **Uni**: *uniform impact* overall areas ($\Psi_{i,j} = \frac{1}{J}$);
- **No**: *No impact:* reduce the problem to the traditional DRO problem with a decision-independent ambiguity set;
- **Max**: Maximum impact on the closest area only ($\min_i d_{i,j}$).
- **Decrease**:decreasing function of the network delay $\exp(-\frac{d_{i,j}}{b})$.

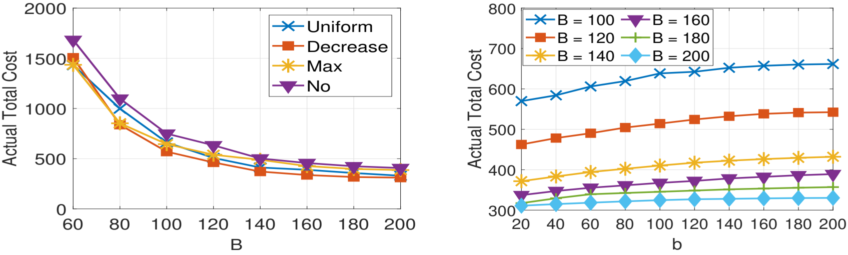

Figure: Choice of decision-dependency: b: decaying factor

Network size	Standard	Improved
$I = 10; J = 10$	31.31s	21.68s
$I = 20; J = 10$	66.88s	62.95s
$I = 20; J = 20$	404.11s	331.79s
$I = 30; J = 20$	1314.8s	901.8s
$I = 40; J = 20$	3357.2s	2178.28s

Table: Runtime comparison

Conclusions

- Who will benefit from this framework?
  - Edge infrastructure provider (e.g., Equinix, AT&T): long-term planning, data center capacity expansion
  - They can proactively control uncertainties and obtain a more accurate representation of uncertainty through the lens of *Endogeneity*.
- The ambiguity set in DRO framework can be based on different metrics. This endogeneity can be also considered in those metrics.

Exact monolithic reformulation: *Exact OPT-Placement*

Case 1: $\alpha_\mathbf{i} = \mathbf{s_i}$

- $v_{i,j} \leq \rho d_{i,j}\alpha_i + \beta_i d_{i,j}$. As $v_{i,j} \leq 0$, the extreme point of $v_{i,j}$ can occur at
  - (i) $v_{i,j} = 0$
  - (ii) $v_{i,j} = \rho d_{i,j}\alpha_i + \beta_i d_{i,j}$ if $\rho d_{i,j} - s_i - \beta_i d_{i,j} < 0$

Exact monolithic reformulation: *Exact OPT-Placement*

Case 1: $\boldsymbol{\alpha_i = s_i}$

- $v_{i,j} \leq \rho d_{i,j}\alpha_i + \beta_i d_{i,j}$. As $v_{i,j} \leq 0$, the extreme point of $v_{i,j}$ can occur at
  (i) $v_{i,j} = 0$
  - $\rho - \frac{s_i}{d_{i,j}} < \beta_i \leq 0$, then $v_{i,j} \leq \rho d_{i,j} - s_i < 0$ due to assumption $s_i > \rho d_{i,j}$. $v_{i,j} \leq 0$ becomes redundant and $v_{i,j} = \rho d_{i,j-s_i}$ is the extreme point.
  - The optimal value of the objective function is

$$s_i\lambda_i(y) + \sum_j C_{i,j}y_j(\rho d_{i,j} - s_i), \ \forall i. \tag{9}$$

Exact monolithic reformulation: *Exact OPT-Placement*

Case 1: $\alpha_\mathbf{i} = \mathbf{s_i}$

- $v_{i,j} \le \rho d_{i,j}\alpha_i + \beta_i d_{i,j}$. As $v_{i,j} \le 0$, the extreme point of $v_{i,j}$ can occur at
  (ii) $v_{i,j} = \rho d_{i,j}\alpha_i + \beta_i d_{i,j}$ if $\rho d_{i,j} - s_i - \beta_i d_{i,j} < 0$
  - $\beta_i \le \rho - \frac{s_i}{d_{i,j}} < 0$, the inequality $\rho d_{i,j} - s_i - \beta_i d_{i,j} > 0$ holds true. Thus, we have the $v_{i,j} \le \rho d_{i,j} - s_i - \beta_i d_{i,j}$ becomes redundant and $v_{i,j}$ represent the extreme point.
  - The optimal value of the objective function is

$$\left[s_i + \left(\rho - \frac{s_i}{d_i^{\text{min}}}\right)\Delta_i\right]\lambda_i(y),\ \forall i. \tag{9}$$

where $d^{\text{min}} = \min_{j' \in \mathcal{J}} d_{i,j'}\,\forall i$, we have $\beta_i = \rho - \frac{s_i}{d_i^{\text{min}}}$

Exact monolithic reformulation: *Exact OPT-Placement*

Case 2: $\alpha_{\mathbf{i}} < \mathbf{s_i}$

- Similarly, $v_{i,j}$ reaches its extreme point at either
    - (i) $v_{i,j} = 0$
    - (ii) $v_{i,j} = \rho d_{i,j} - \alpha_i - \beta_i d_{i,j}$

Exact monolithic reformulation: *Exact OPT-Placement*

Case 2: $\alpha_\mathbf{i} < \mathbf{s_i}$

- Similarly, $v_{i,j}$ reaches its extreme point at either
  (i) If $v_{i,j} = 0$ for some j, it must hold that $\rho d_{i,j} - \alpha_i - \beta_i d_{i,j} \geq 0$, i.e., $\alpha_i \leq d_{i,j}(\rho - \beta_i)$
  - we aim to find extreme points for β_i such that

$$s_i > \left\{ \max_{\beta_i} (\rho - \beta_i) d_{i,j}, \forall j, \qquad \text{s.t} \qquad \beta_i \leq 0 \right\}. \tag{10}$$

  - Notably, $(\rho - \beta_i) d_{i,j} > s_i$ when $\beta_i \to -\infty, \forall i$. Thus, $\alpha_i = \rho d_{i,j}$ and $\beta_i = 0$ represent the extreme points.
  - The optimal value of the objective function is

$$s_i \lambda_i(y) + \sum_j C_{i,j} y_j (\rho d_{i,j} - s_i), \ \forall i. \tag{11}$$

Exact monolithic reformulation: *Exact OPT-Placement*

Case 2: $\alpha_\mathbf{i} < \mathbf{s_i}$

- Similarly, $v_{i,j}$ reaches its extreme point at either (ii) $v_{i,j} = \rho d_{i,j} - \alpha_i - \beta_i d_{i,j}$, it implies that the constraint $v_{i,j} \leq \rho d_{i,j} - \alpha_i - \beta_i d_{i,j}$ is binding, i.e.,

$$\rho d_{i,j} - \alpha_i - \beta_i d_{i,j} \leq 0. \quad (10)$$

  - Since $\beta_i \leq 0$, $\beta_i = 0$ represents the extreme point that ensures above constraints holds. Thus α_i must satisfy $\rho d_{i,j} \leq \alpha_i < s_i$ for all j
  - The optimal value of the objective function is

$$\rho d_{i,j^*} \lambda_i(y) + \sum_{j:d_{i,j} < d_{i,j^*}} C_{i,j} \rho (d_{i,j} - d_{i,j^*}) y_j, \ \forall i. \quad (11)$$

Exact monolithic reformulation: *Exact OPT-Placement*

Case 2: $\alpha_\mathbf{i} < \mathbf{s_i}$

- since $s_i > \rho d_{i,j^*}$ and $\rho - \frac{s_i}{d_i^{\text{min}}} < \rho - \frac{\rho d_{i,j^*}}{d_i^{\text{min}}}$
- For a given j^*, we have a closed form expression:

$$\begin{cases} \rho d_{i,j^*}\lambda_i(y) + \sum_{j:d_{i,j}<d_{i,j^*}} C_{i,j} y_j \rho (d_{i,j} - d_{i,j^*}) \\ \rho d_{i,j^*}\lambda_i(y) + \left[\left(\rho - \frac{\rho d_{i,j^*}}{d_i^{\text{min}}}\right)\Delta_i\right]\lambda_i(y). \end{cases}$$

Exact monolithic reformulation: *Exact OPT-Placement*

Case 2: $\alpha_\mathbf{i} < \mathbf{s_i}$

- since $s_i > \rho d_{i,j^*}$ and $\rho - \frac{s_i}{d_i^{\text{min}}} < \rho - \frac{\rho d_{i,j^*}}{d_i^{\text{min}}}$
- For a given j^*, we have a closed form expression:

$$\begin{cases} \rho d_{i,j^*}\lambda_i(y) + \sum_{j:d_{i,j}<d_{i,j^*}} C_{i,j}y_j\rho(d_{i,j} - d_{i,j^*}) \\ \rho d_{i,j^*}\lambda_i(y) + \left[\left(\rho - \frac{\rho d_{i,j^*}}{d_i^{\text{min}}}\right)\Delta_i\right]\lambda_i(y). \end{cases}$$

- For each area i, the optimal inner problem $g_i(y,\lambda)$ corresponding to the actual realization ξ_n with probability $p_{i,n}$ can be written as

$$\theta_{i,n}(y) = \max_{j^*\in\mathcal{J}} \quad \rho d_{i,j^*}\xi_n + \max\left\{[\rho - \frac{\rho d_{i,j^*}}{d_i^{\text{min}}})\Delta_i]\xi_n, \sum_{j:d_{i,j}<d_{i,j^*}} C_{i,j}\rho(d_{i,j} - d_{i,j^*})y_j\right\}, \ \forall i,n. \tag{12}$$

Exact monolithic reformulation: *Exact OPT-Placement*

Intuitive ideas of step 1

The inner obj determines which one of these negative terms imposes a more stringent requirement, either in terms of the capacity constraint or the delay constraint.

The whole problem now becomes the "*min-max*" problem.