Practically Motivated Matching Problems

- **Humanitarian:** “Dynamic Matching with Post-allocation Service: Application to Refugee Resettlement” (Vahideh)
- **Ride-sharing:** “Adaptive Policies and Approximation Schemes for Dynamic Matching” (Ali)
- **Education:** “Teacher Redistribution in a Public School System” (Jay)

Yale SCHOOL OF MANAGEMENT

Dynamic Matching with Post-allocation Service: Application to Refugee Resettlement

Vahideh Manshadi (Yale School of Management)

Joint work with:

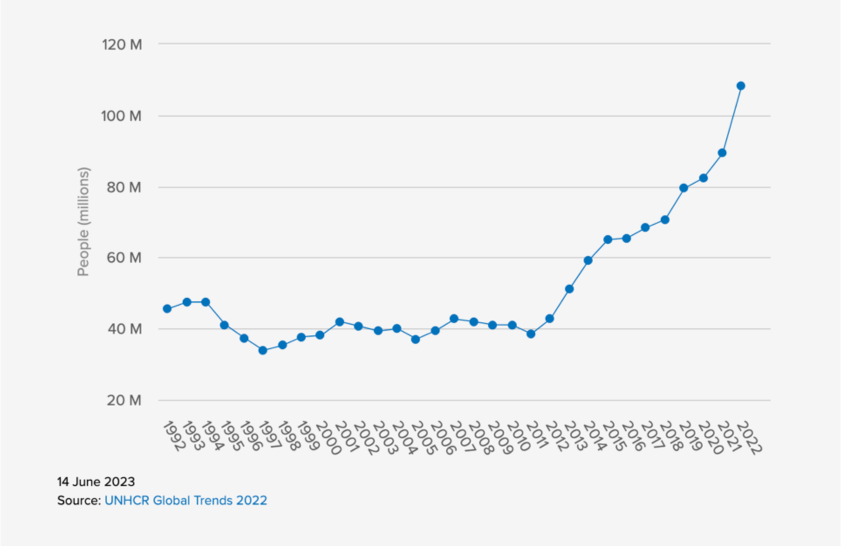
Kirk Bansak
UC Berkeley

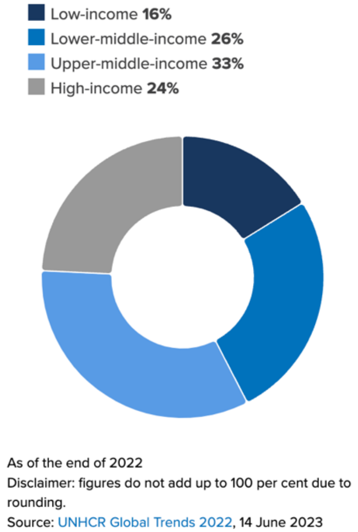
Soonbong Lee
Yale School of Management

Rad Niazadeh
UChicago Booth School of Business

Elisabeth Paulson
Harvard Business School

Refugee Resettlement: Background





- **35.3 million refugees worldwide**:
  - Syria 13.5 million
  - Ukraine 12.6 million
  - Venezuela 10.2 million
  - Afghanistan 9.7 million
  - Palestine 6.1 million (late 2023)

Refugee Resettlement: Background

Refugee resettlement: *An international effort for a durable solution*

- Relocate refugees to host countries
- Finding them a new home
- **Finding them a new job**

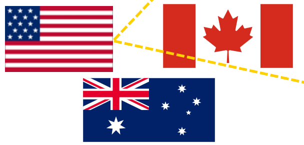
Refugee Resettlement: Background

UNHCR
The UN Refugee Agency

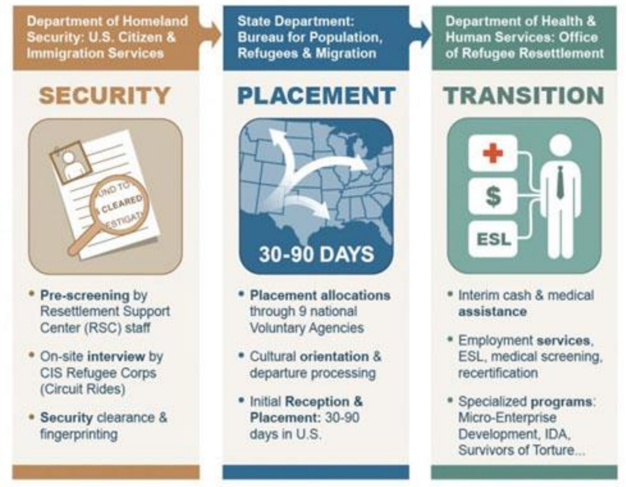
Referred to host countries

Largest hosts:

Most resettlements:



United Nations High Commissioner for Refugees



United States Refugee Admissions Program (USRAP)

Refugee resettlement: *An international effort for a durable solution*

- Relocate refugees to host countries
- Finding them a new home
- **Finding them a new job**

Refugee Resettlement in the U.S.

U.S. Citizenship and Immigration Services

+

Population, Refugees, and Migration
U.S. DEPARTMENT *of* STATE

+

U.S. Department of Health and Human Services

20k-50k resettlement cases

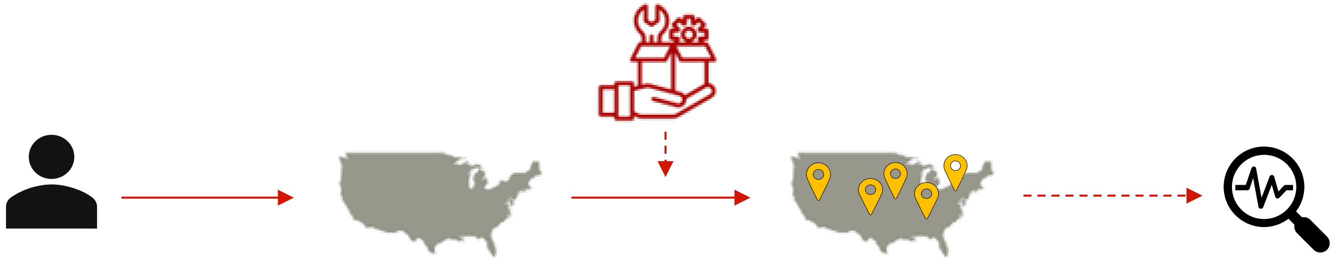
Non-profit Resettlement Partners of the U.S. Government

1. Church World Service (CWS)
2. Ethiopian Community Development Council (ECDC)
3. Episcopal Migration Ministries (EMM)
4. Hebrew Immigrant Aid Society (HIAS)
5. International Rescue Committee (IRC)
6. US Committee for Refugees and Immigrants (USCRI)
7. **Global Refuge**
   (formerly Lutheran Immigration and Refugee Services (LIRS))
8. United States Conference of Catholic Bishops (USCCB)
9. World Relief Corporation (WR)
10. Bethany Christian Services

Dynamics of Refugee Resettlement

US, Netherlands, Switzerland, Norway, Sweden, etc.

Resettlement agency



Case arrives at host country

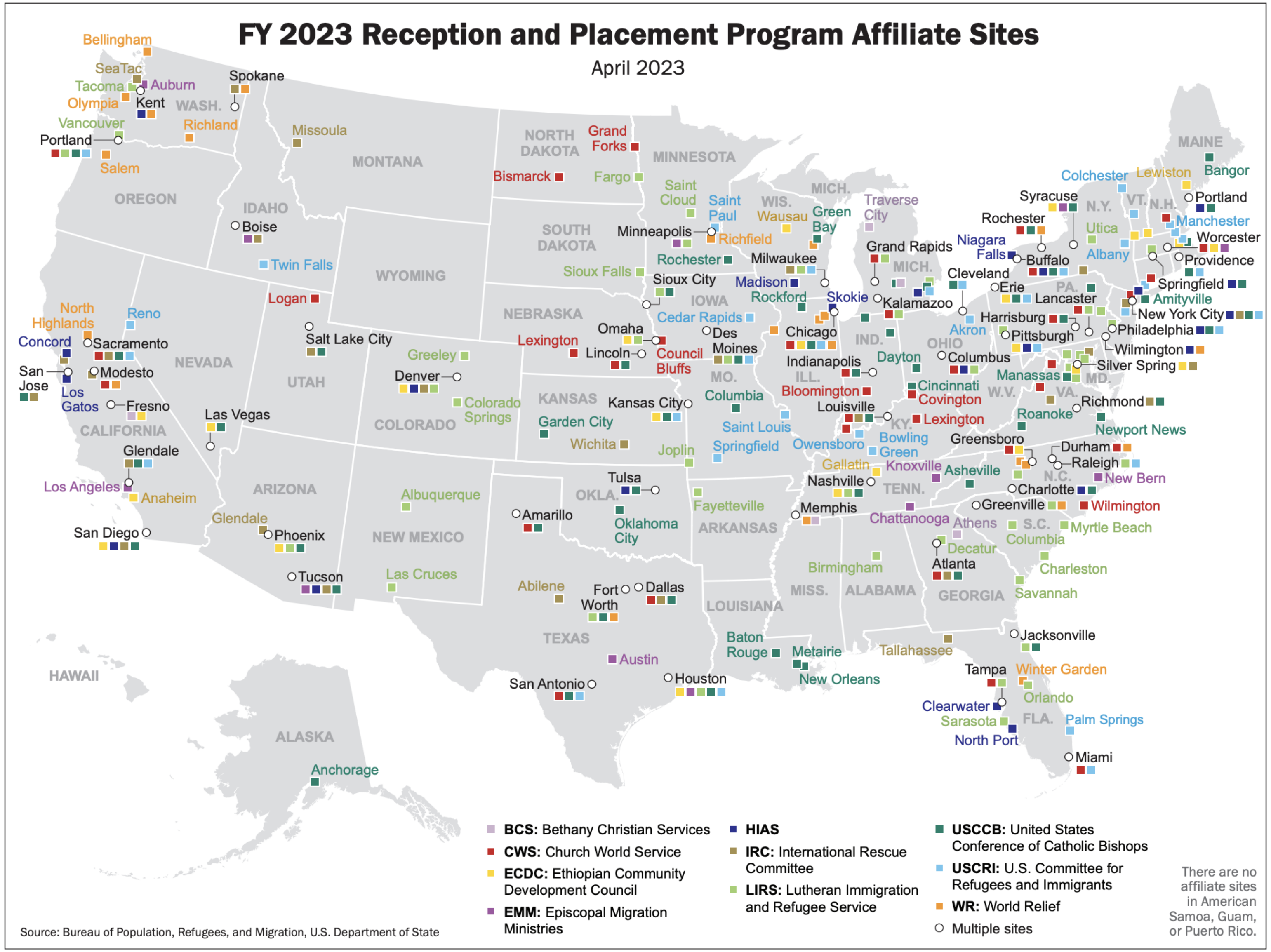
Case is assigned to a locality within host country

Outcomes observed
(e.g. employment status after X days/years)

Constraints
- Locality capacity constraints (N people/year)
- Family ties
- Medical/educational constraints

Resettlement Locations

Unclassified

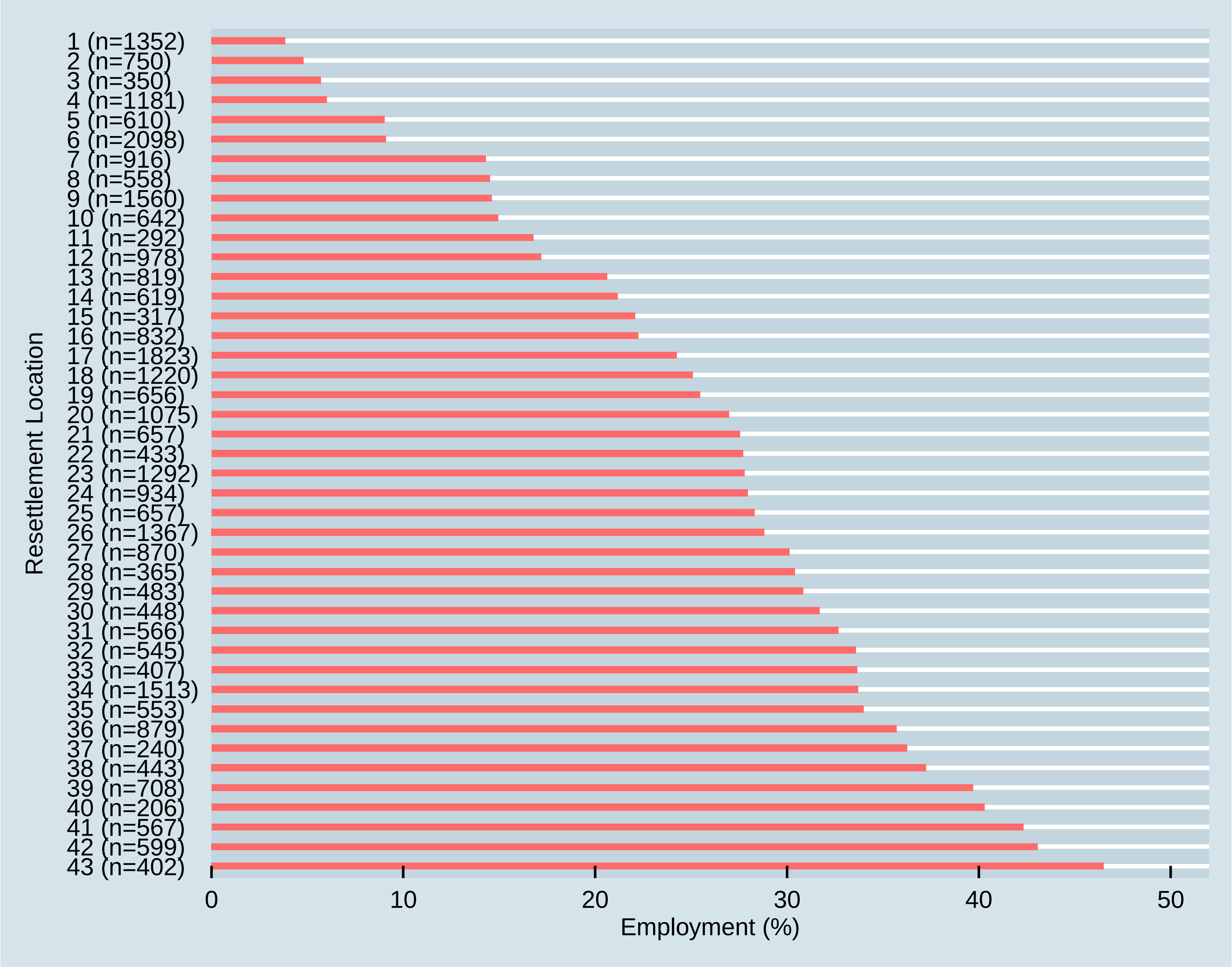


2750 4-23 STATE (INR)

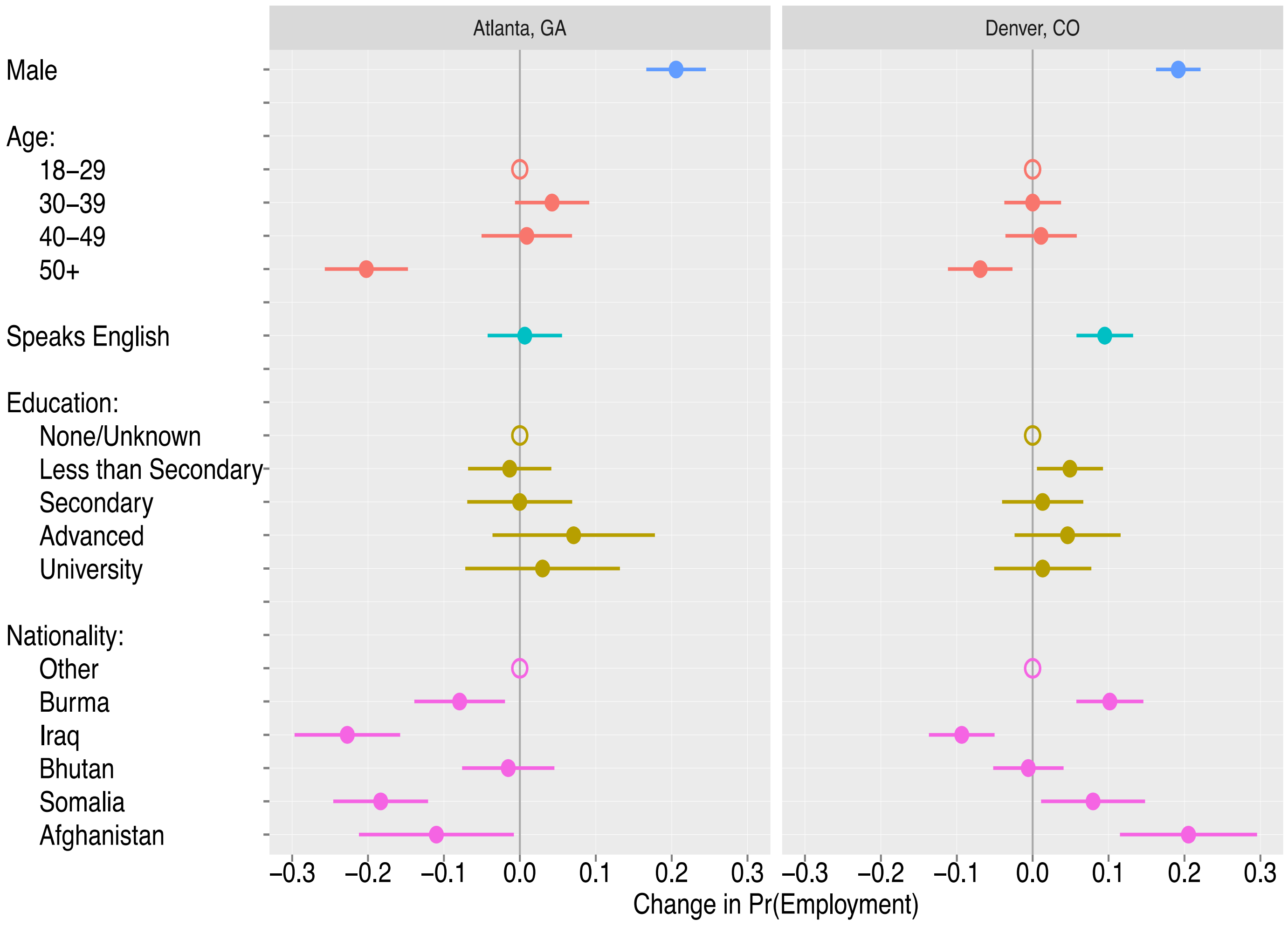
Location Matters

The initial placement of a refugee family within a host country has a **<u>significant impact on their future success</u>**

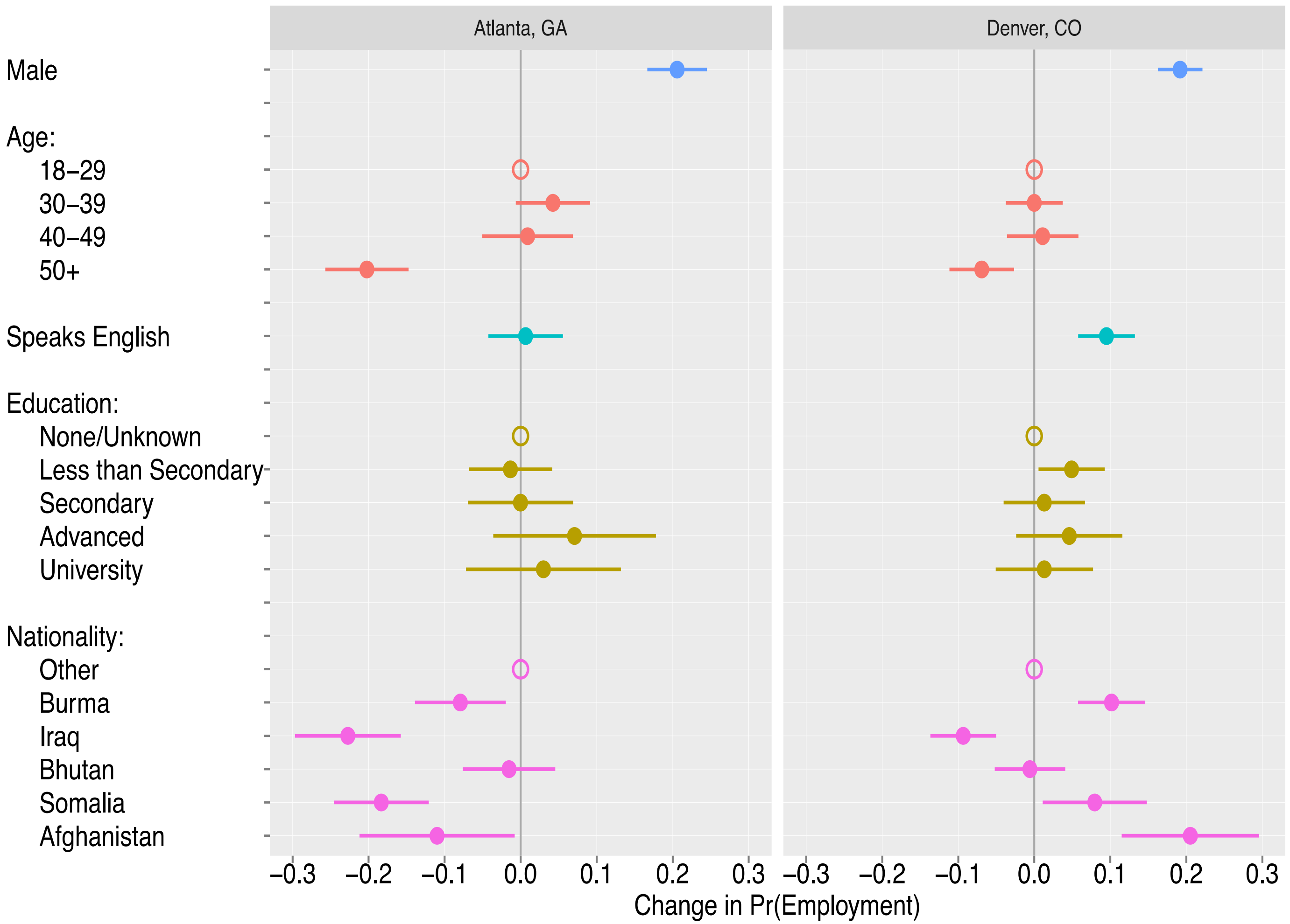
(Bansak et al. 2018)



Impact of Locations Varies Across Cases



Impact of Locations Varies Across Cases



Opportunity:

Improving outcomes through **<u>data-driven algorithmic assignment</u>**

Harnessing Big Data to Improve Refugee Resettlement

ipl immigration policy lab

Stanford | Zurich

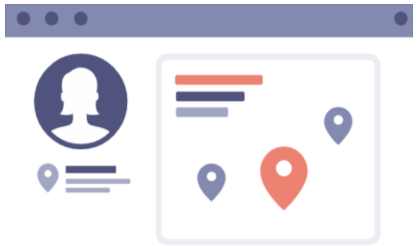
GeoMatch™

Connecting people to places

Refugee Resettlement: Value of Algorithmic Assignment

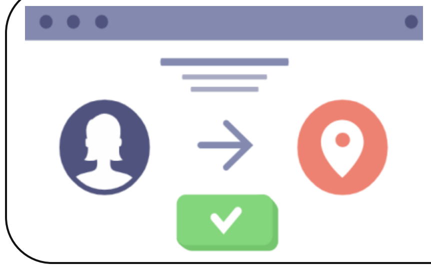
- Refugee resettlement program: relocate refugees to host country (Bansak et al. ‘18)
  - **Important decision: <u>initial geographic placement</u>** has a profound impact on economic outcome

- Opportunity: **improvement through <u>data-driven</u> <u>algorithmic</u> assignment**



Predict employment outcomes

through ML models



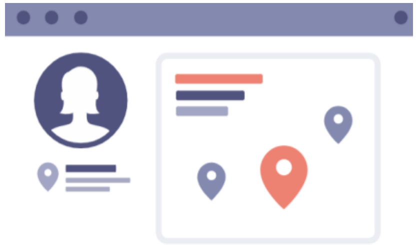
Recommend a location

through optimization algorithm

Refugee Resettlement: Value of Algorithmic Assignment

- Refugee resettlement program: relocate refugees to host country (Bansak et al. ‘18)
  - **Important decision: <u>initial geographic placement</u>** has a profound impact on economic outcome

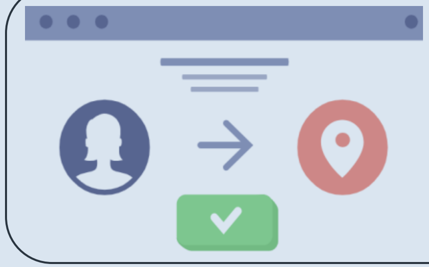
- Opportunity: **improvement through <u>data-driven</u> <u>algorithmic</u> assignment**



Predict employment outcomes

through ML models

<u>**Our paper**</u>

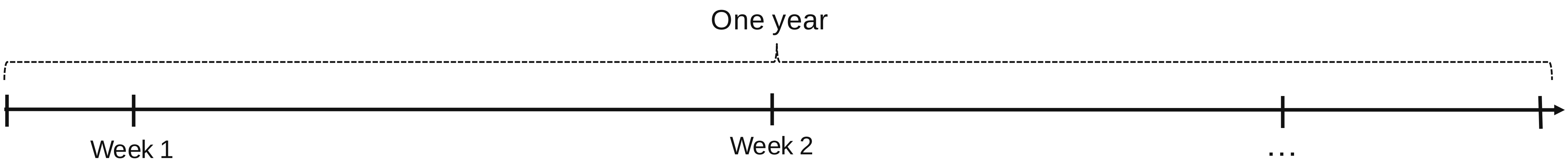


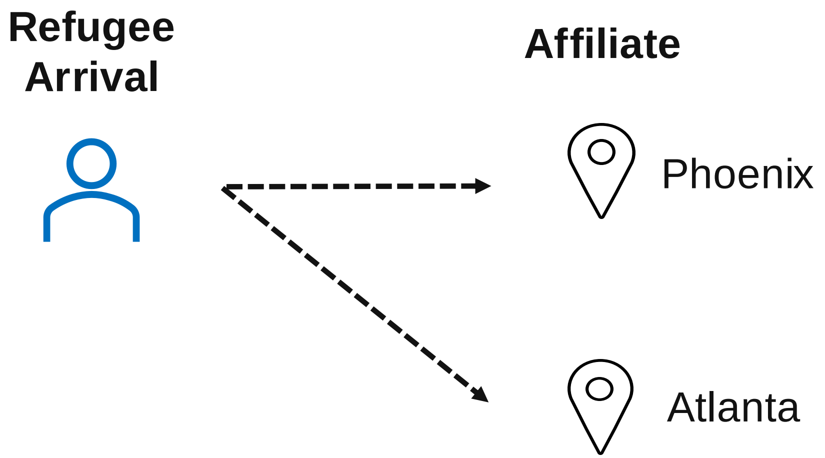
Recommend a location

through optimization algorithm

*Through **collaboration with a major U.S. resettlement agency**,*
we design placement algorithm to incorporate our partner’s <u>novel operational considerations</u>

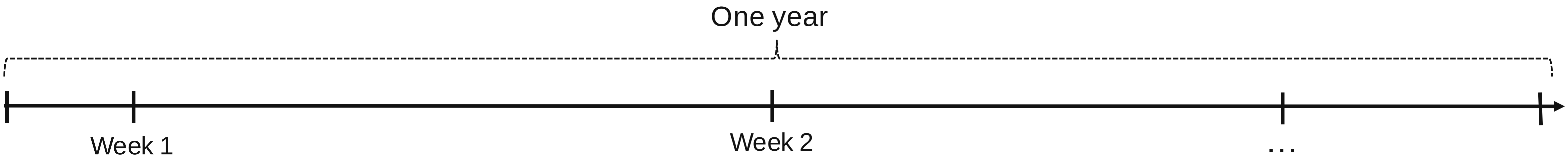
Refugee Resettlement as Dynamic Matching

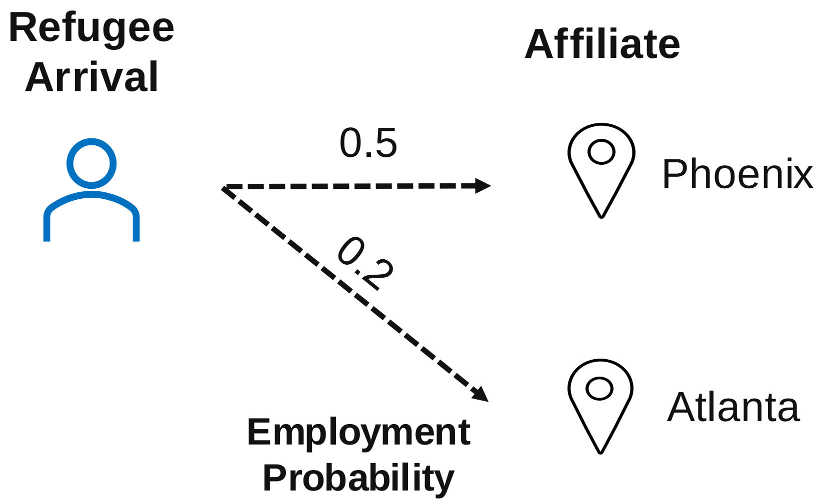




- **Dynamic matching:** refugee matched upon arrival without knowing future

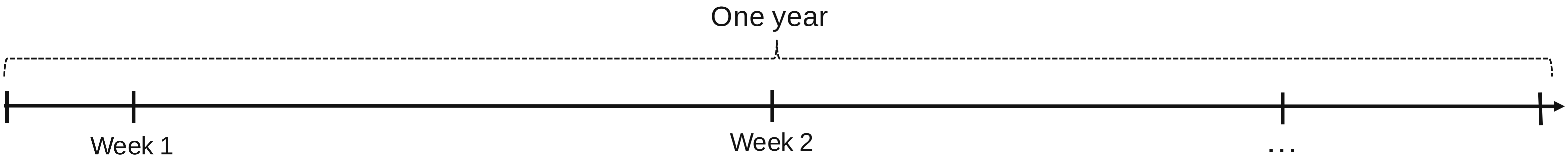
Refugee Resettlement as Dynamic Matching

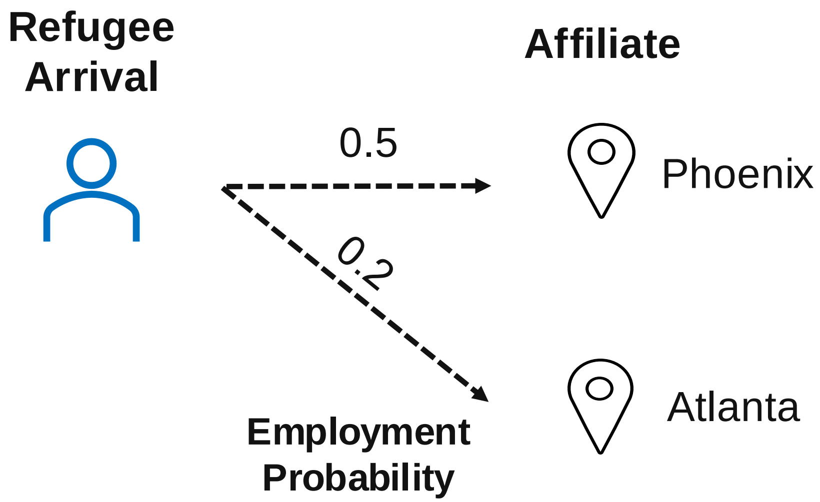




- **Dynamic matching:** refugee matched upon arrival without knowing future
  - ➢ Maximize employment outcome s.t. **resource constraint**

Refugee Resettlement as Dynamic Matching

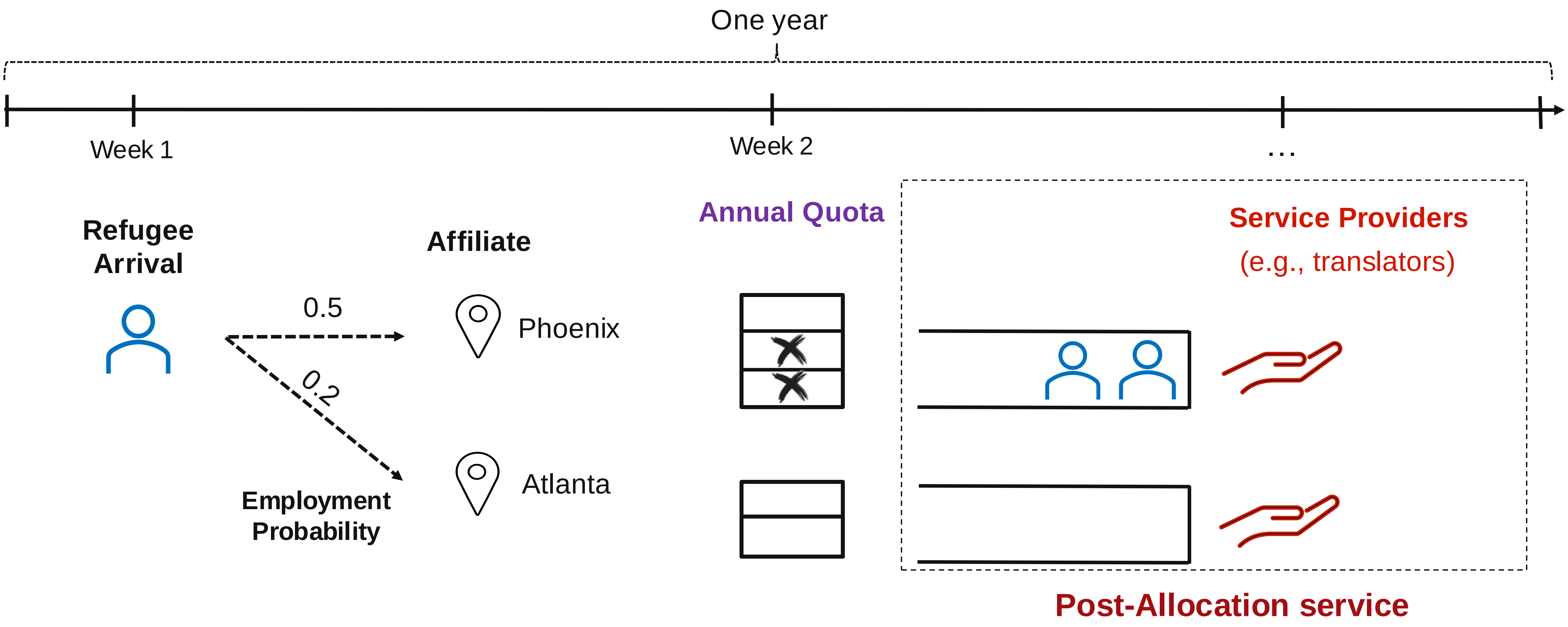




- **Dynamic matching:** refugee matched upon arrival without knowing future
  - ➢ Maximize employment outcome s.t. **resource constraint**

- **Novel aspect: post-allocation service**

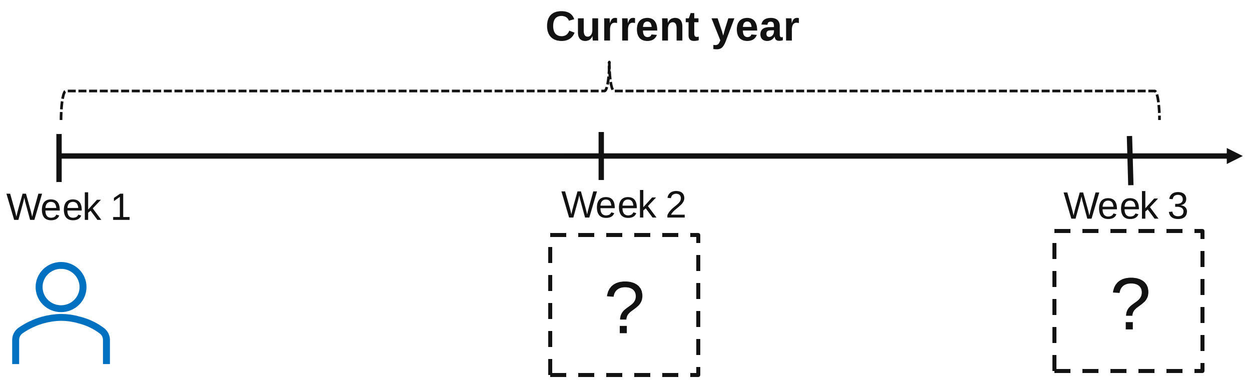
Novel Aspect: Post-Allocation Service



Key operational consideration: avoid congestion for post-allocation services

Need for “distribution-free” design

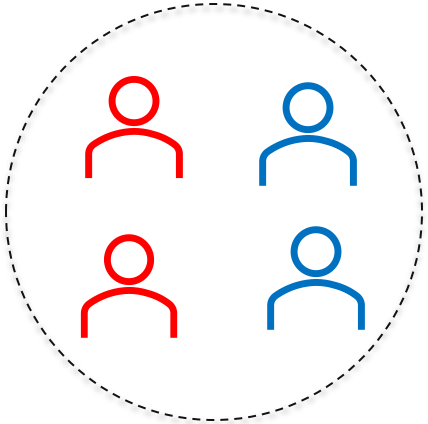
- **Dynamic matching:** refugee matched upon arrival **without knowing future**



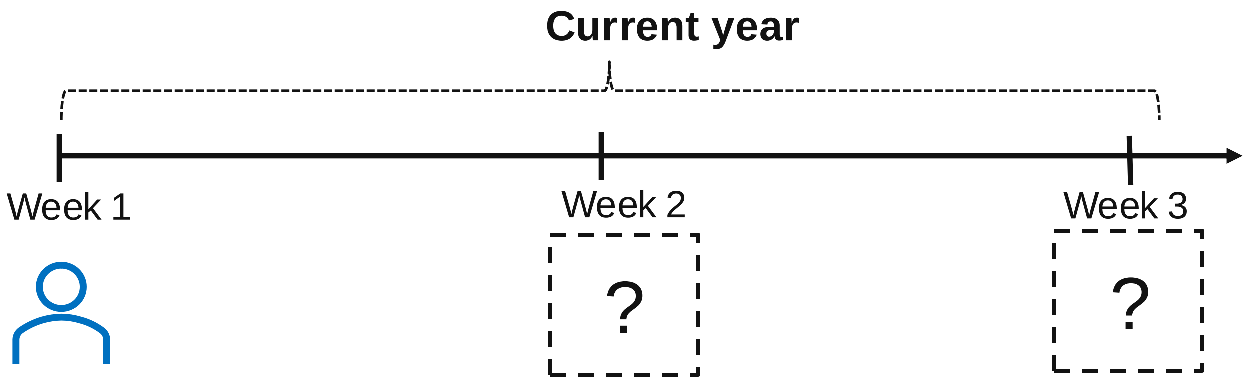
➢ Existing proposal: <u>simulate future from data of past years</u> (Bansak & Paulson ’22, Ahani et al. ’22)

Need for “distribution-free” design

- **Dynamic matching:** refugee matched upon arrival **without knowing future**



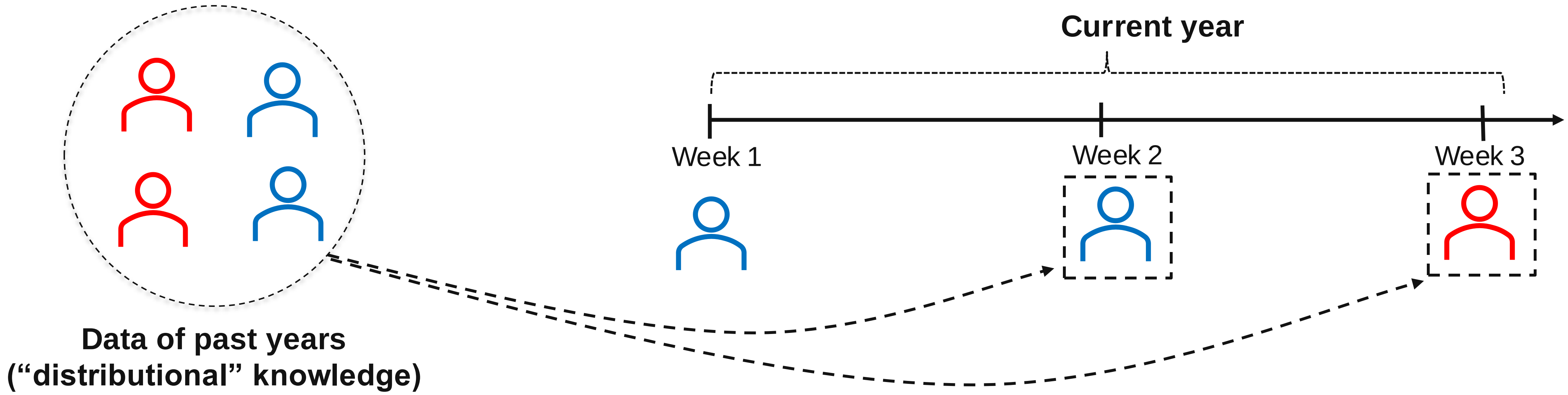
Data of past years
(“distributional” knowledge)



➢ Existing proposal: <u>simulate future from data of past years</u> (Bansak & Paulson ’22, Ahani et al. ’22)

Need for “distribution-free” design

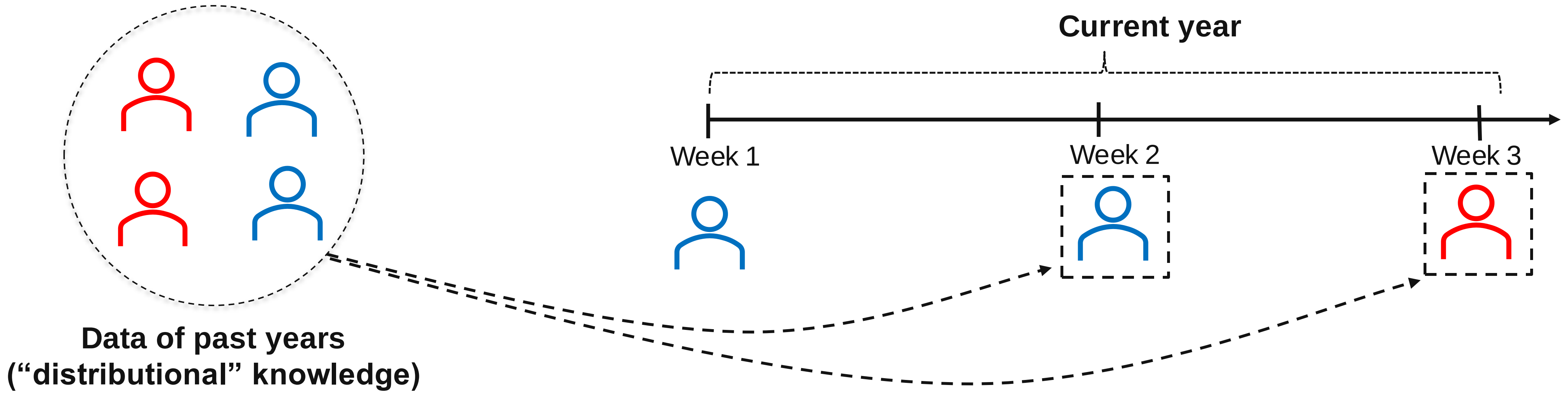
- **Dynamic matching:** refugee matched upon arrival **without knowing future**



➢ Existing proposal: <u>simulate future from data of past years</u> (Bansak & Paulson ’22, Ahani et al. ’22)

Need for "distribution-free" design

- **Dynamic matching:** refugee matched upon arrival **without knowing future**



➢ Existing proposal: simulate future from data of past years (Bansak & Paulson '22, Ahani et al. '22)

➢ **Not robust due to across year variations in the refugee pool composition!**

Evidence of across year variation

✓**Tied** cases: a pre-determined target affiliate (**family reunification policy**)

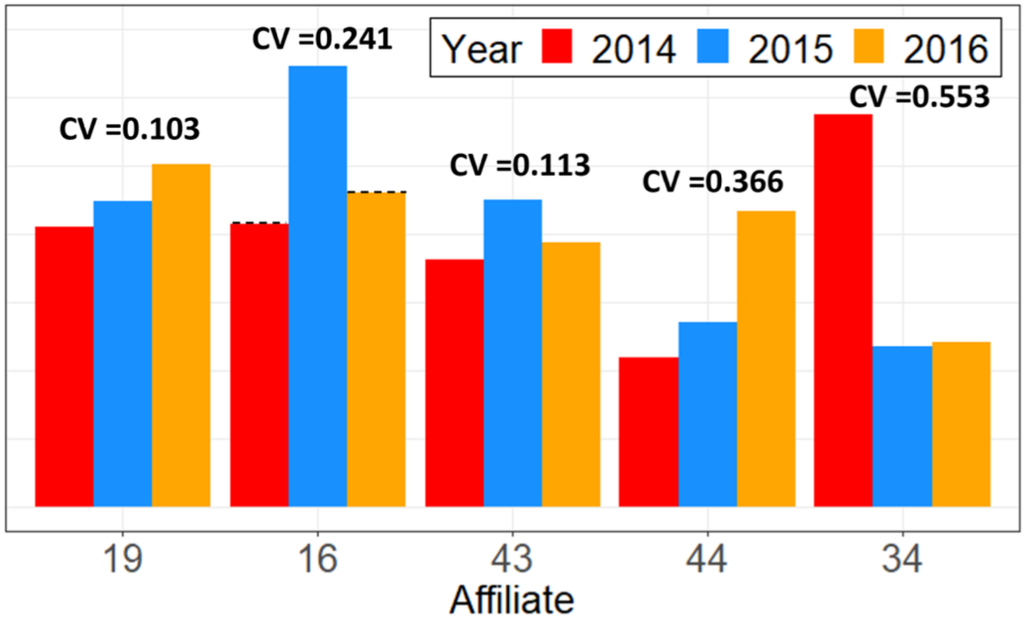
INA: ACT 411 - OFFICE OF REFUGEE RESETTLEMENT

Sec. 411. [8 U.S.C. 1521]

[...] *Refugees resettling to the United States can identify* ***friends or relatives already living in the United States*** *with whom they wish to be reunited upon arrival. Once identified, those individuals are contacted by a Resettlement Agency to confirm if they would like to have* ***the refugees resettled nearby****. If they agree, these individuals are considered* ***U.S. ties.*** *The Resettlement Agency will not share refugees' personal details, such as medical information, with their U.S. tie. However, they will coordinate with the U.S. tie to prepare for the refugees' arrival in the U.S [...]*

Evidence of across year variation

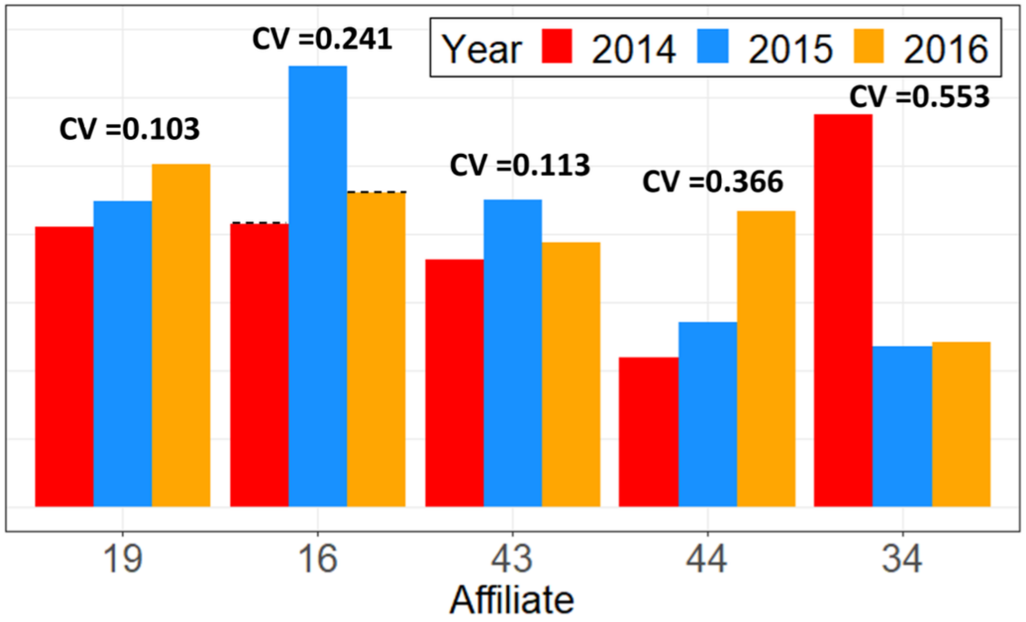
Normalized Number of the Tied Cases across Years



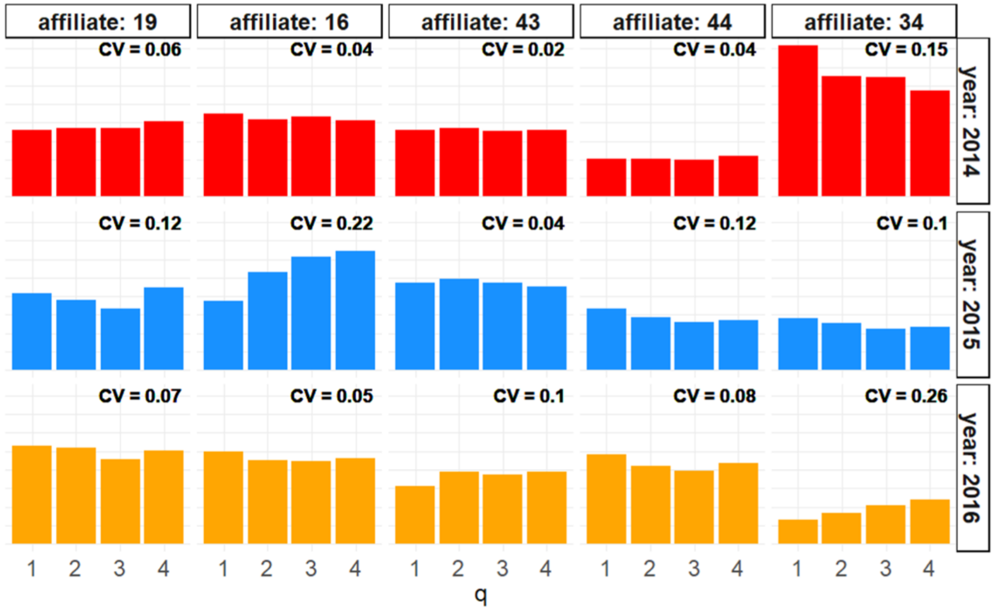
Normalized # of tied cases varies significantly across the years…

Evidence of across year variation

Normalized Number of the Tied Cases across Years



Normalized Number of the Tied cases Within Each Year



Normalized # of tied cases varies significantly across the years… but less so within a year

Research Question & Contribution

Research Question

How to design a dynamic matching algorithm that optimizes for employment outcome, given specific "operational considerations"?

(1) Respects annual quota & avoid congestion for post-allocation service

(2) Does not require distributional knowledge (e.g., past years' data)

Contribution

(1) Develop a model of dynamic matching with post-allocation service

(2) Design learning-based algorithms

- ✓ Distribution-free (no reliance on past years' data) & easy-to-implement

(3) Case study on refugee resettlement data

- ✓ Improving performance over existing proposals

Research Question & Contribution

Research Question

How to design a dynamic matching algorithm that optimizes for employment outcome, given specific "operational considerations"?

(1) Respects annual quota & avoid congestion for post-allocation service

(2) Does not require distributional knowledge (e.g., past years' data)

Contribution

(1) Develop a model of dynamic matching with post-allocation service

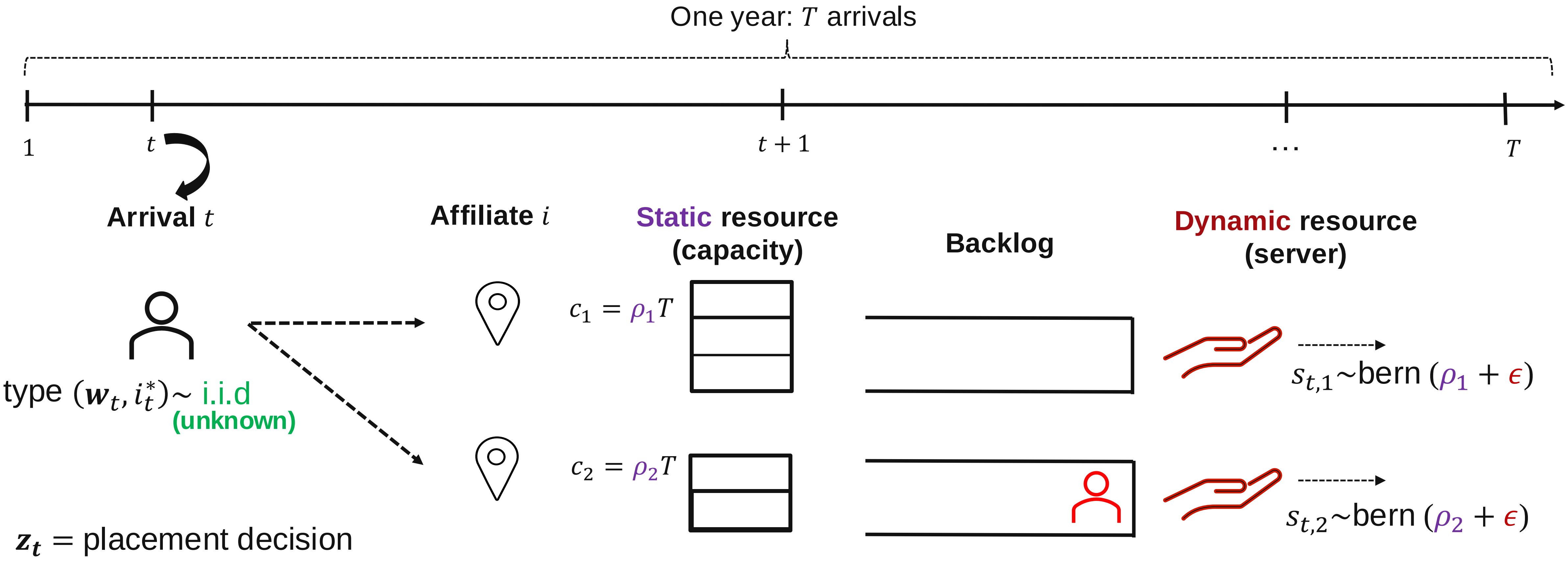
(2) Design new learning-based algorithms

- ✓ Distribution-free, near-optimal performance guarantee, & computationally fast

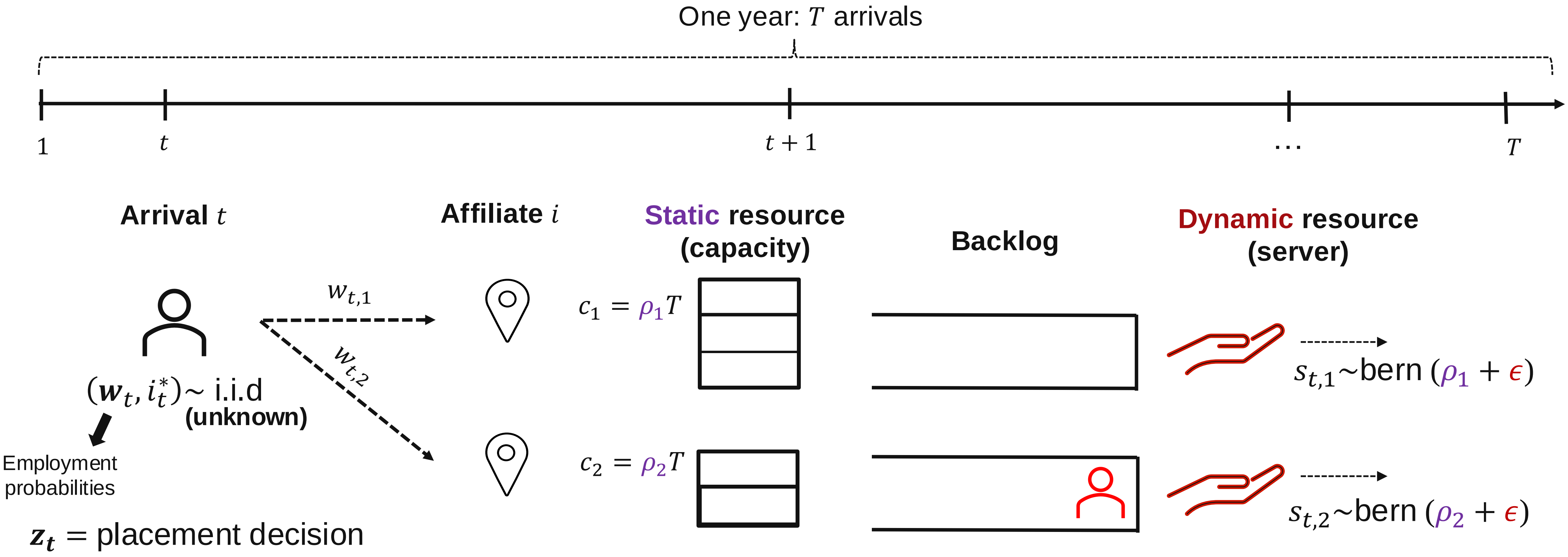
(3) Case study on refugee resettlement data

- ✓ Improving performance over existing proposals

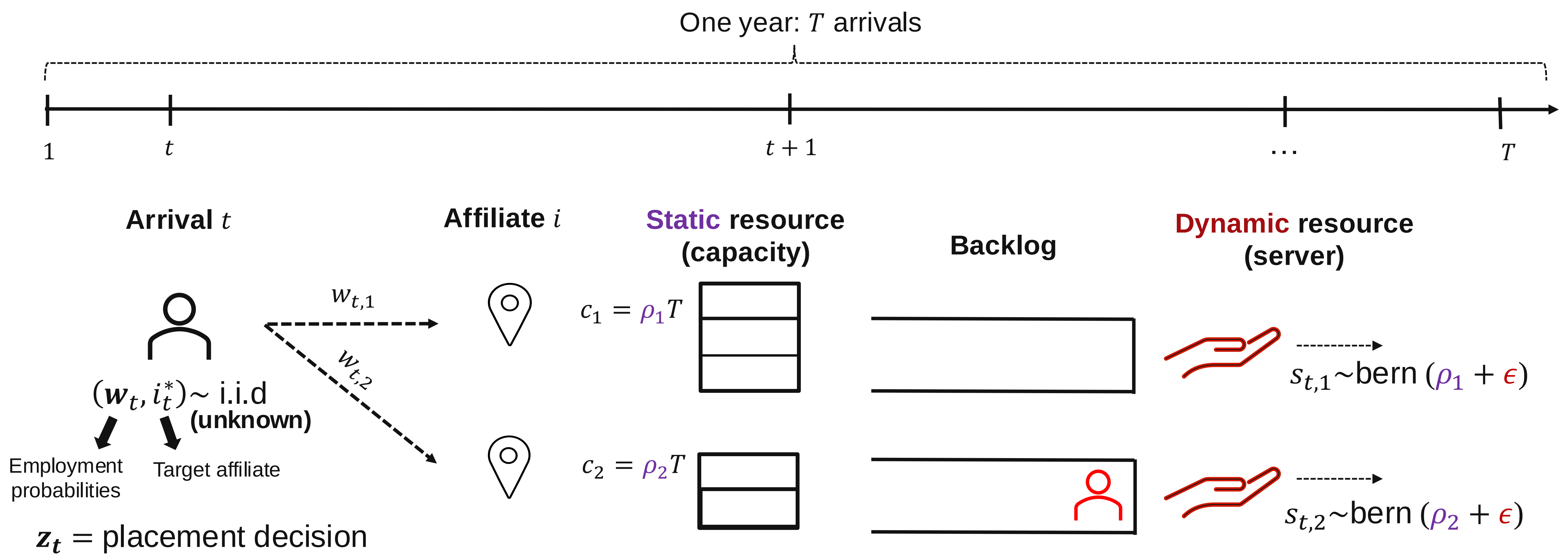
Model: Dynamic Matching with Post-Allocation Service



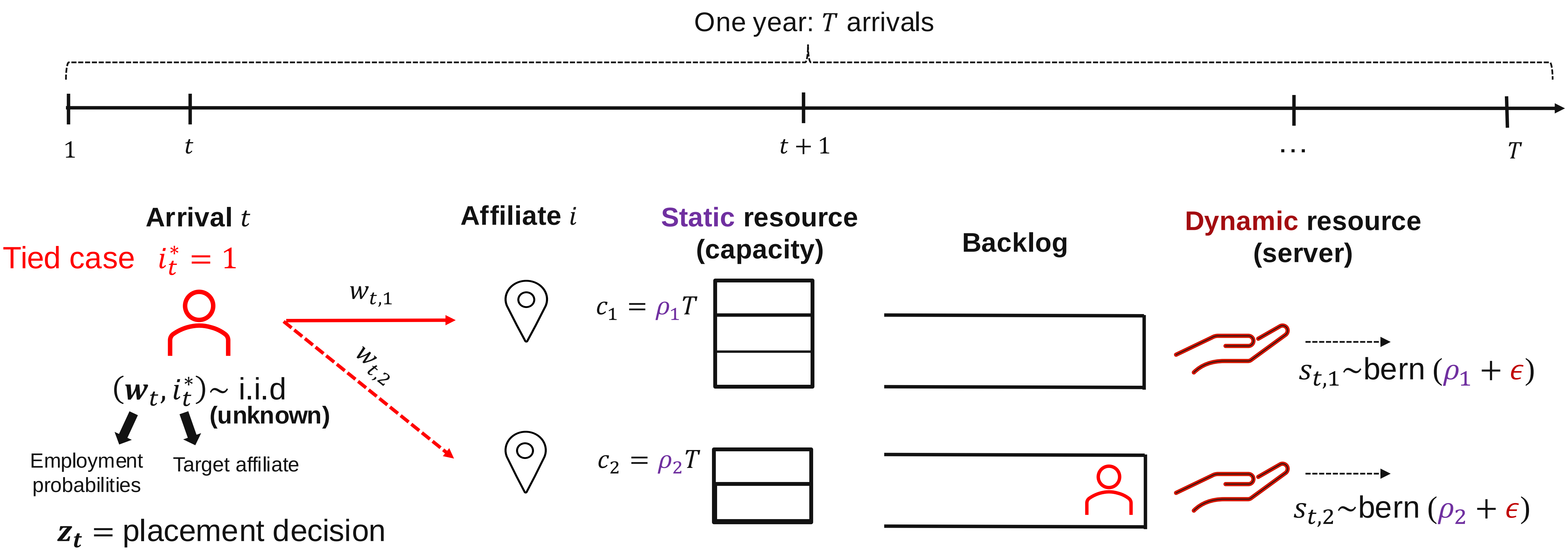
Model: Dynamic Matching with Post-Allocation Service



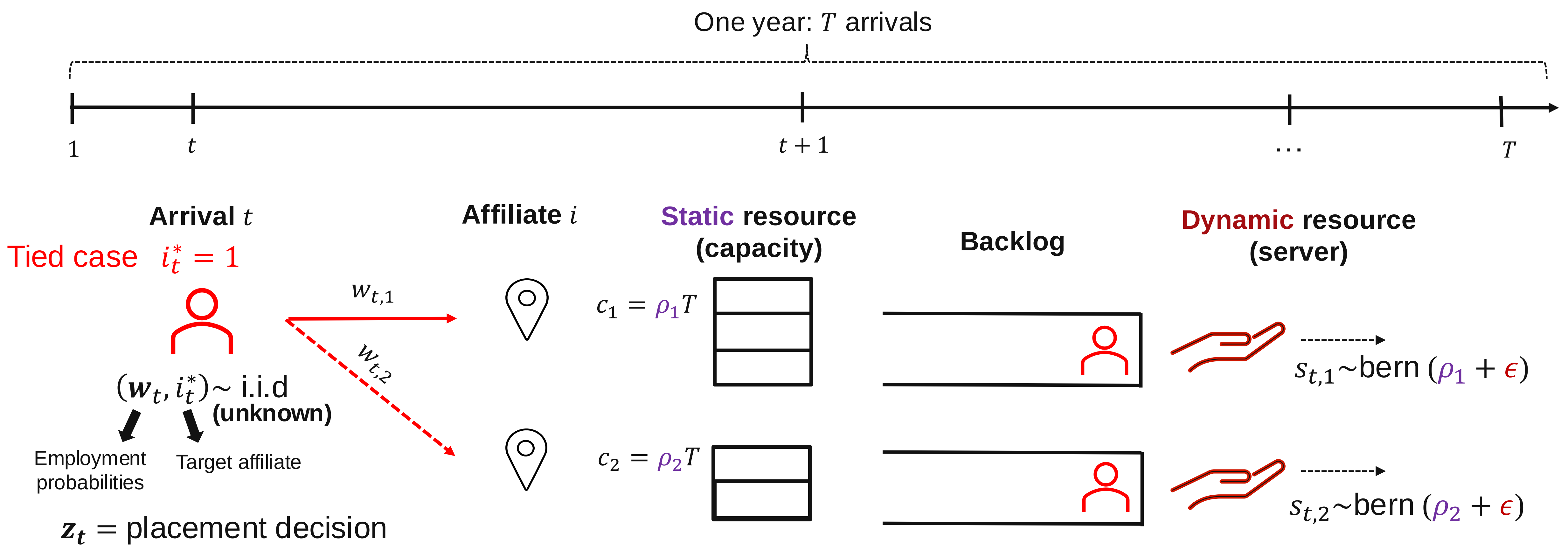
Model: Dynamic Matching with Post-Allocation Service



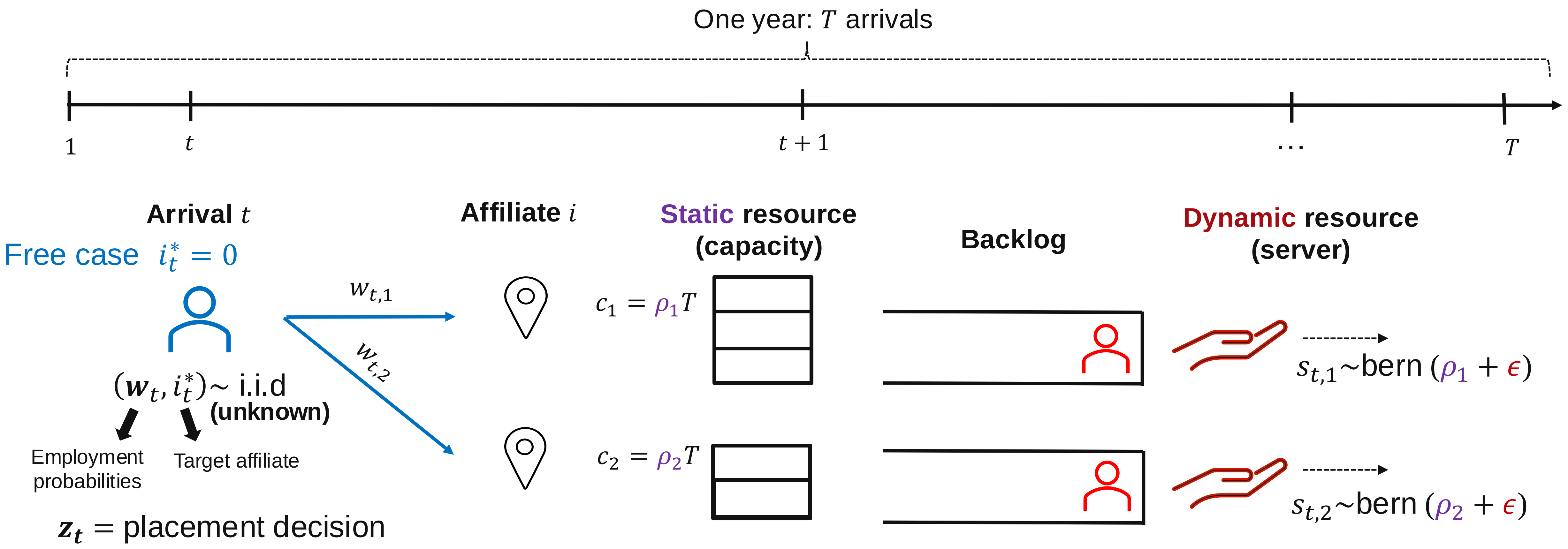
Model: Dynamic Matching with Post-Allocation Service



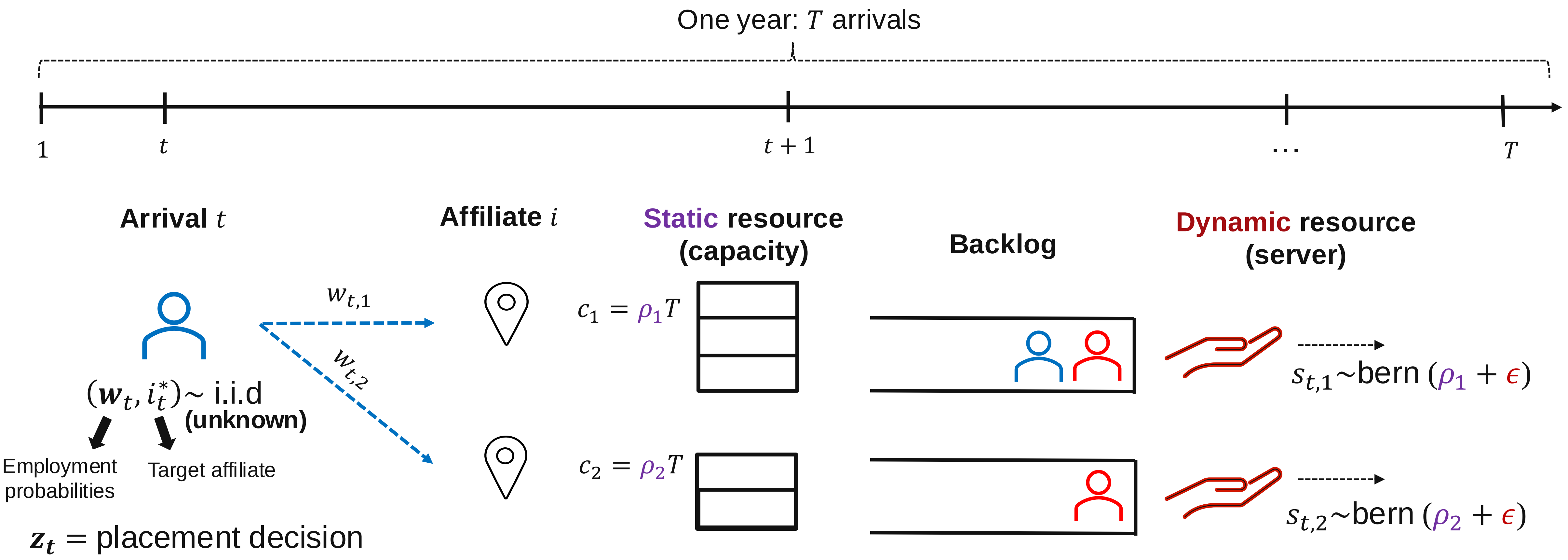
Model: Dynamic Matching with Post-Allocation Service



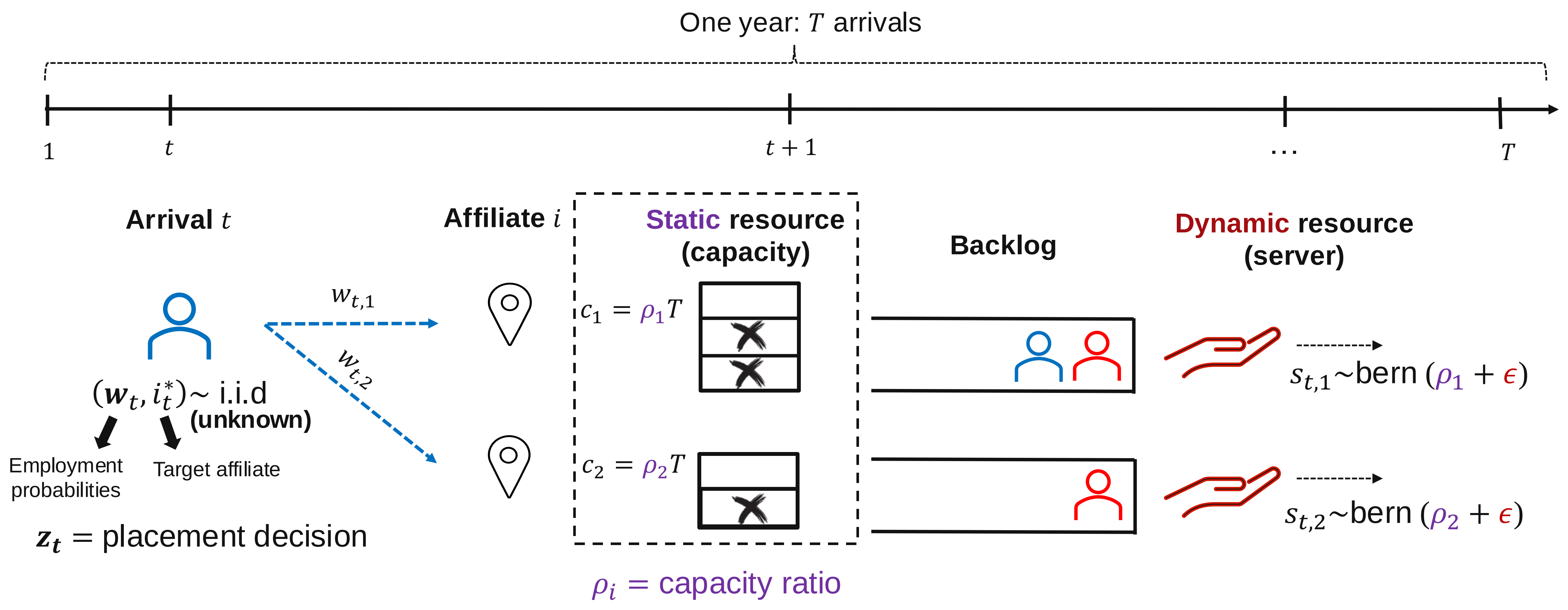
Model: Dynamic Matching with Post-Allocation Service



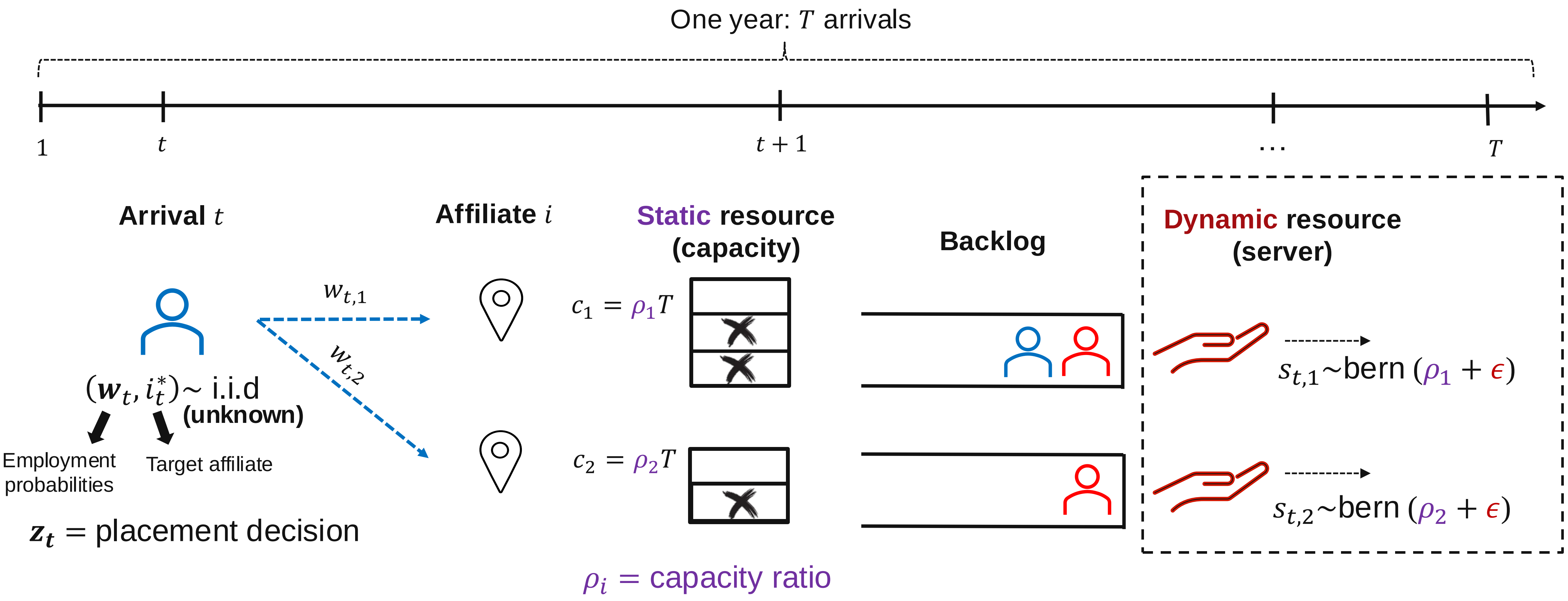
Model: Dynamic Matching with Post-Allocation Service



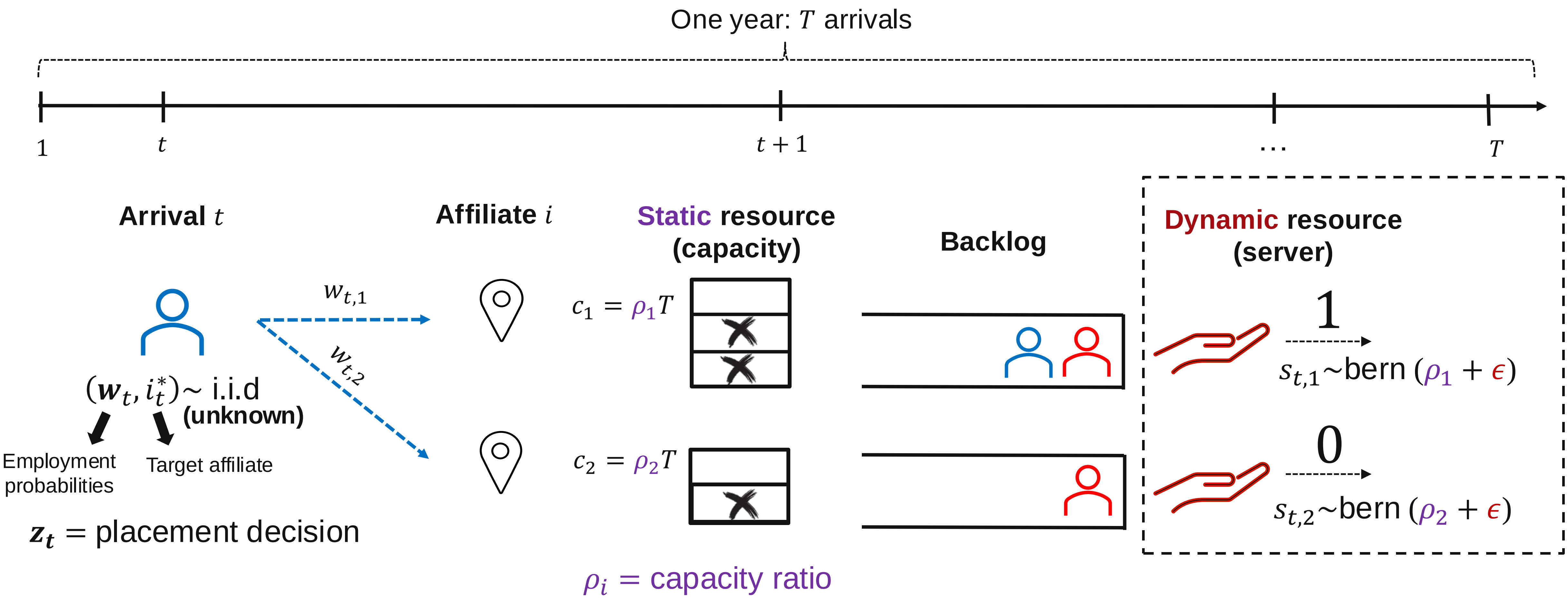
Model: Dynamic Matching with Post-Allocation Service



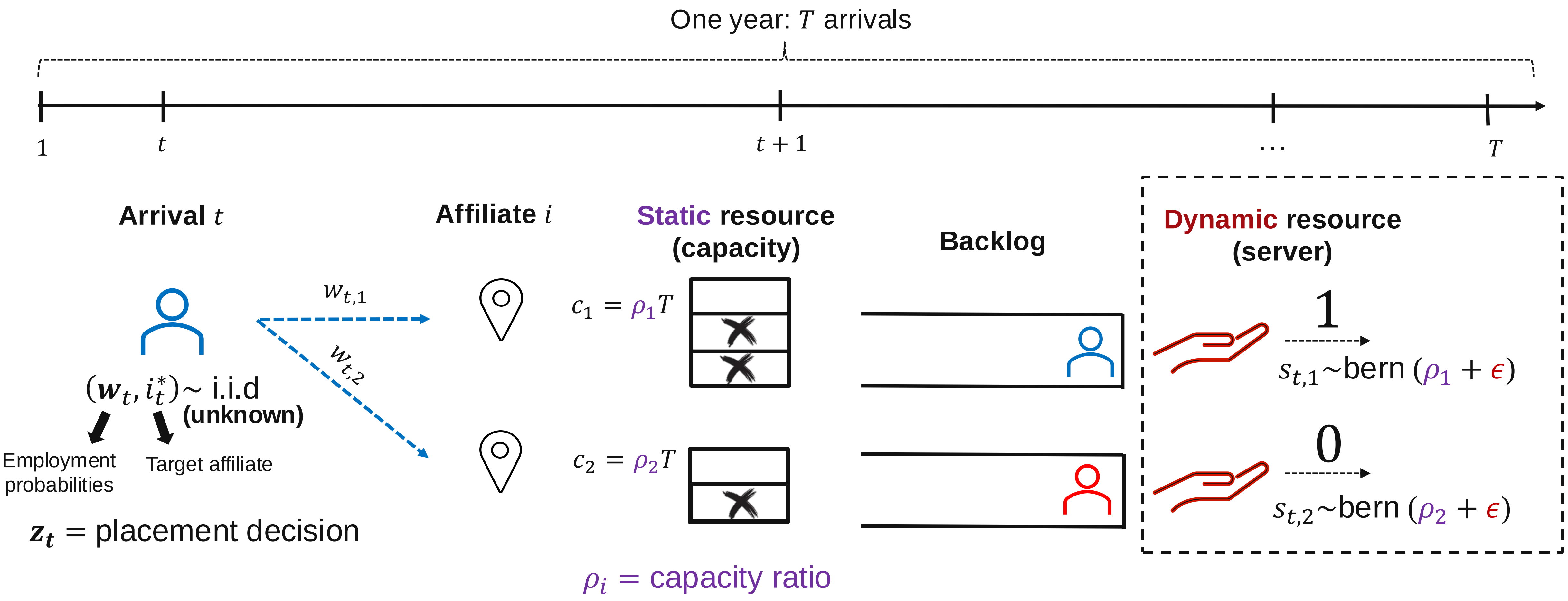
Model: Dynamic Matching with Post-Allocation Service



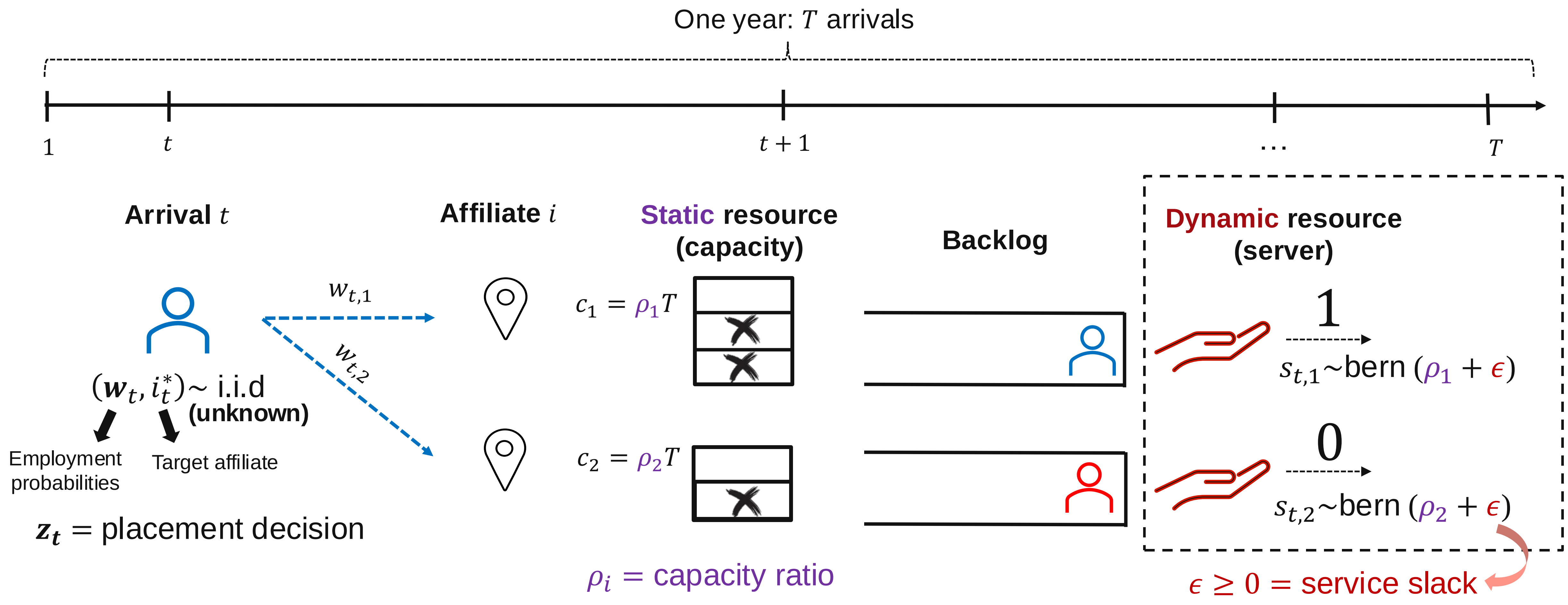
Model: Dynamic Matching with Post-Allocation Service



Model: Dynamic Matching with Post-Allocation Service

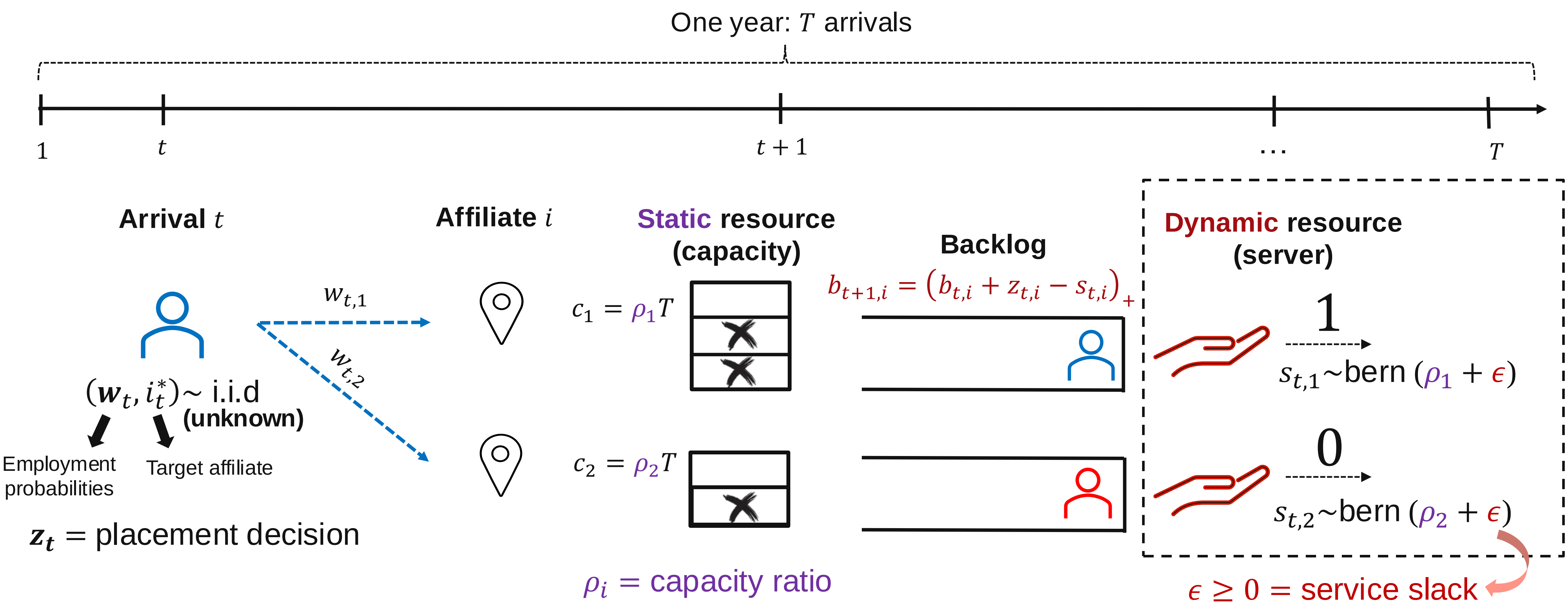


Model: Dynamic Matching with Post-Allocation Service



Note: (1) endogenous arrival rate to queues!
(2) ρ_i = baseline arrival rates

Model: Dynamic Matching with Post-Allocation Service



Model: Dynamic Matching with Post-Allocation Service

Objective = $$Max_{z_t} \underbrace{\sum_{t=1}^{T}\sum_{i=1}^{m} w_{ti} z_{ti}}_{\textit{Total employment}} - \alpha \times \underbrace{\sum_{i=1}^{m}\left(\sum_{t=1}^{T} z_{ti} - c_i\right)_+}_{\substack{\textit{Over-allocation} \\ \textit{(due to tied cases)}}} - \gamma \times \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m} b_{t,i}\right)}_{\substack{\textit{Average backlog} \\ \textit{(due to “bursty” matching)}}}$$

s.t $$\sum_{\tau=1}^{t} z_{\tau i}\mathbb{I}[i_\tau^* = 0] \le \left(c_i - \sum_{\tau=1}^{t}\mathbb{I}[i_\tau^* = i]\right)_+ ,\forall t \in [T], \forall i \in [m]$$

Hard constraint: no over-allocation can occur from free cases

Regret = $\sup_{arrival\ dist} E[\text{Optimal "Offline"}] - E[ALG]$

Goal: online algorithm with $o(T)$ regret

Algorithm Design

total employment

$\boldsymbol{\theta^*}$**: over allocation**

$$Max_{\boldsymbol{z}_t} \quad \sum_{t=1}^{T}\sum_{i=1}^{m} w_{ti}z_{ti} - \alpha \times \sum_{i=1}^{m}\left(\sum_{t=1}^{T} z_{ti} - c_i\right)_+ - \gamma \times \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m} b_{t,i}\right)$$

s.t

$$\sum_{t=1}^{T} z_{ti}\mathbb{I}[i_t^* = 0] \leq \left(c_i - \sum_{t=1}^{T}\mathbb{I}[i_t^* = i]\right)_+$$

$\boldsymbol{\lambda^*}$**: (relaxed) capacity constraint**

$$b_{t+1,i} \geq b_{t,i} + z_{ti} - s_{t,i}, \quad b_{t+1,i} \geq 0$$

$\{\boldsymbol{\beta_t^*}\}_{t=1}^{T}$**: backlog dynamics**

- **High-level Idea:** *learn (update)* the dual variables & design a score-based matching rule

Algorithm Design

total employment

$\boldsymbol{\theta}^*$**: over allocation**

$$Max_{\boldsymbol{z}_t} \quad \sum_{t=1}^{T}\sum_{i=1}^{m} w_{ti} z_{ti} - \alpha \times \sum_{i=1}^{m}\left(\sum_{t=1}^{T} z_{ti} - c_i\right)_+ - \gamma \times \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m} b_{t,i}\right)$$

$$\text{s.t} \quad \sum_{t=1}^{T} z_{ti}\mathbb{I}[i_t^* = 0] \leq \left(c_i - \sum_{t=1}^{T}\mathbb{I}[i_t^* = i]\right)_+$$

$\boldsymbol{\lambda}^*$**: (relaxed) capacity constraint**

$$b_{t+1,i} \geq b_{t,i} + z_{ti} - s_{t,i}, \quad b_{t+1,i} \geq 0$$

$\{\boldsymbol{\beta}_t^*\}_{t=1}^{T}$**: backlog dynamics**

- **High-level Idea:** *learn (update)* the dual variables & design a score-based matching rule

(Agrawal & Devanur ’14, Balseiro et al. ’21)

➢ **Time-invariant** $(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*)$ → direct learning via adversarial online learning (& stationary arrivals)

Algorithm Design

total employment

θ^*: over allocation

$$Max_{z_t} \quad \sum_{t=1}^{T}\sum_{i=1}^{m} w_{ti} z_{ti} - \alpha \times \sum_{i=1}^{m}\left(\sum_{t=1}^{T} z_{ti} - c_i\right)_+ - \gamma \times \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m} b_{t,i}\right)$$

s.t

$$\sum_{t=1}^{T} z_{ti}\mathbb{I}[i_t^* = 0] \le \left(c_i - \sum_{t=1}^{T}\mathbb{I}[i_t^* = i]\right)_+$$

λ^*: (relaxed) capacity constraint

$$b_{t+1,i} \ge \ b_{t,i} + z_{ti} - s_{t,i}, \quad b_{t+1,i} \ge 0$$

$\{\beta_t^*\}_{t=1}^T$: backlog dynamics

- **High-level Idea:** *learn (update)* the dual variables & design a score-based matching rule

(Agrawal & Devanur ’14, Balseiro et al. ’21)

- **Time-invariant** (θ^*, λ^*) → direct learning via adversarial online learning (& stationary arrivals)
- **Time-varying** $(\beta_t^*)_{t=1}^T$: too many duals to learn!

Algorithm Design

total employment

θ^*: over allocation

$$Max_{z_t} \quad \sum_{t=1}^{T}\sum_{i=1}^{m} w_{ti} z_{ti} - \alpha \times \sum_{i=1}^{m}\left(\sum_{t=1}^{T} z_{ti} - c_i\right)_+ - \gamma \times \left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m} b_{t,i}\right)$$

s.t

$$\sum_{t=1}^{T} z_{ti}\mathbb{I}[i_t^* = 0] \le \left(c_i - \sum_{t=1}^{T}\mathbb{I}[i_t^* = i]\right)_+$$

λ^*: (relaxed) capacity constraint

$$b_{t+1,i} \ge b_{t,i} + z_{ti} - s_{t,i}, \quad b_{t+1,i} \ge 0$$

$\{\beta_t^*\}_{t=1}^T$: backlog dynamics

- **High-level Idea:** *learn (update)* the dual variables & design a score-based matching rule

(Agrawal & Devanur ’14, Balseiro et al. ’21)

- **Time-invariant** (θ^*, λ^*) → direct learning via adversarial online learning (& stationary arrivals)
- **Time-varying** $(\beta_t^*)_{t=1}^T$: too many duals to learn!
  - **Congestion-aware (CA) algorithm**: requires real-time backlog information
  - **Congestion-oblivious (CO) algorithm**: does not require any backlog information

Algorithm Design

- **High-level Idea:** *learn (update)* the dual variables & design a score-based matching rule

(Agrawal & Devanur '14, Balseiro et al. '21)

- **Time-invariant** $(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*)$ $\rightarrow$ direct learning via adversarial online learning (& stationary arrivals)
- **Time-varying** $(\boldsymbol{\beta}_t^*)_{t=1}^T$: too many duals to learn!

Congestion-aware (CA) algorithm

High-level: *use backlog information to penalize bursty matching while learning & optimizing*

- Directly learning time-invariant dual variables $(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*)$ via online learning
- Indirectly learning time-variant dual variables
  - Connections between backlog dynamics & sub-gradient descent in a surrogate dual problem → *(scaled) current backlog* = dual estimate!

Algorithm Design

- **High-level Idea:** *learn (update)* the dual variables & design a score-based matching rule

(Agrawal & Devanur '14, Balseiro et al. '21)

- **Time-invariant** $(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*)$ → direct learning via adversarial online learning (& stationary arrivals)
- **Time-varying** $(\boldsymbol{\beta}_t^*)_{t=1}^T$: too many duals to learn!

Congestion-aware (CA) algorithm

High-level: *use backlog information to penalize bursty matching while learning & optimizing*

- Directly learning time-invariant dual variables $(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*)$ via online learning
- Indirectly learning time-variant dual variables
  - Connections between backlog dynamics & sub-gradient descent in a surrogate dual problem → *(scaled) current backlog* = dual estimate!

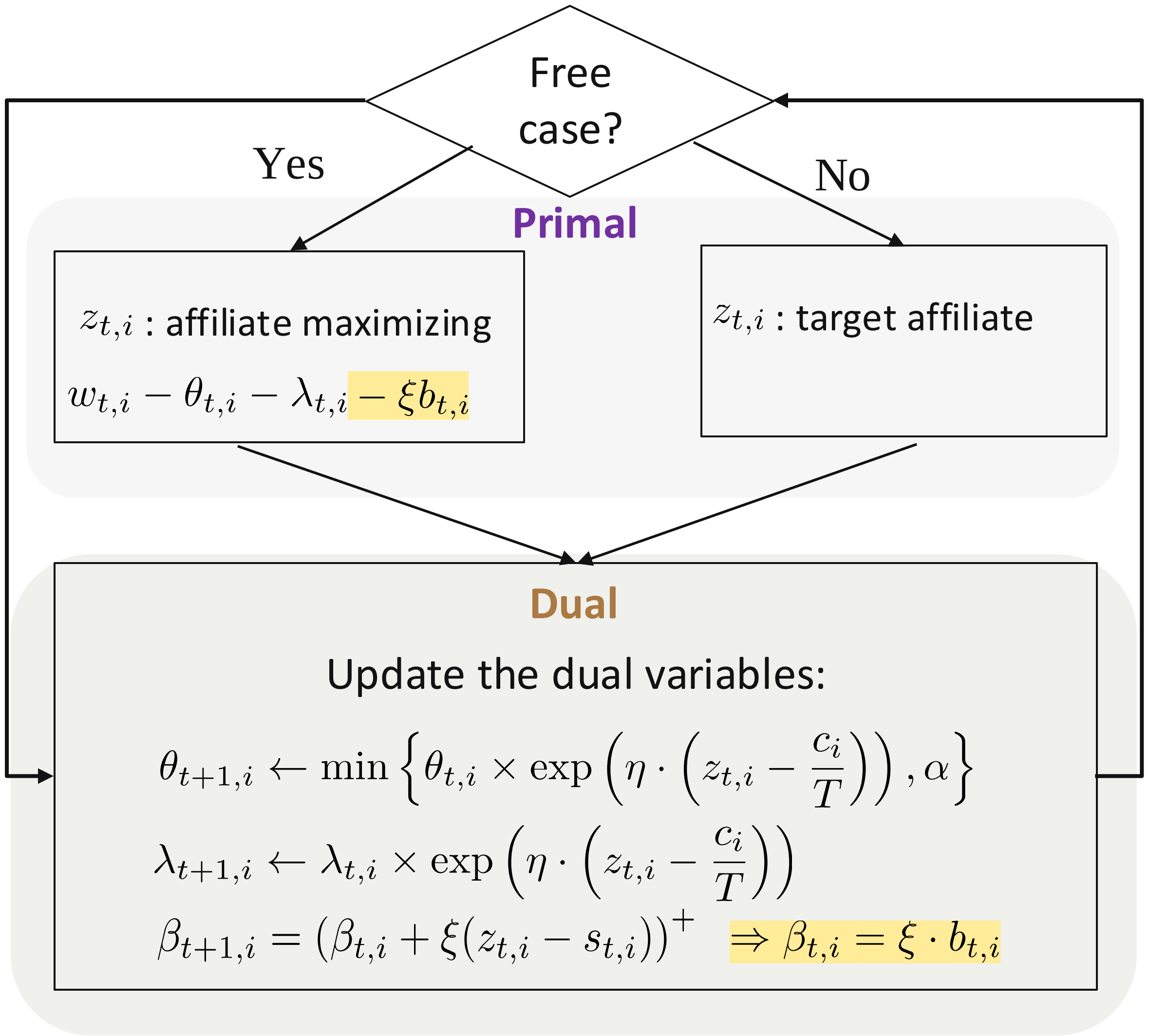
Congestion-oblivious (CO) algorithm

High-level: *control backlog by ensuring fast and high prob. convergence of endogenous arrival rates while learning & optimizing*

- Surrogate-primal program ignore backlog
- Directly learning time-invariant dual variables $(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*)$ via online learning …
  - but this time, with *time-varying learning rates* which we prove results in high-probability last-iterate convergence of both duals and endogenous arrival rates!

Congestion-aware (CA) Algorithm

- Affiliate i **chosen** $\Rightarrow$ dual variables $\uparrow$
- Affiliate i not chosen $\Rightarrow$ dual variables $\downarrow$



High-level: *use backlog information to penalize bursty matching while learning & optimizing*

- Directly learning time-invariant dual variables $(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*)$ via multiplicative update rules
- Indirectly learning time-variant dual variables $(\boldsymbol{\beta}_t^*)_{t=1}^T$: projected gradient descent = scaled current backlog

Theoretical Results

Congestion-aware (CA) algorithm

Theorem [Main Result I]

For $\forall\ \epsilon \geq 0$, **CA algorithm** obtains a regret

$$\min\left\{\mathcal{O}\left(\sqrt{T}+\frac{\gamma}{\epsilon}\right), \mathcal{O}\left(\sqrt{\gamma T}\right)\right\}$$

Proposition [Lower-bound I] If $\gamma = \Omega(T)$, no online algorithm can achieve $o(T)$ regret

Theoretical Results

Congestion-aware (CA) algorithm

Theorem [Main Result I]

For $\forall\, \epsilon \geq 0$, **CA algorithm** obtains a regret

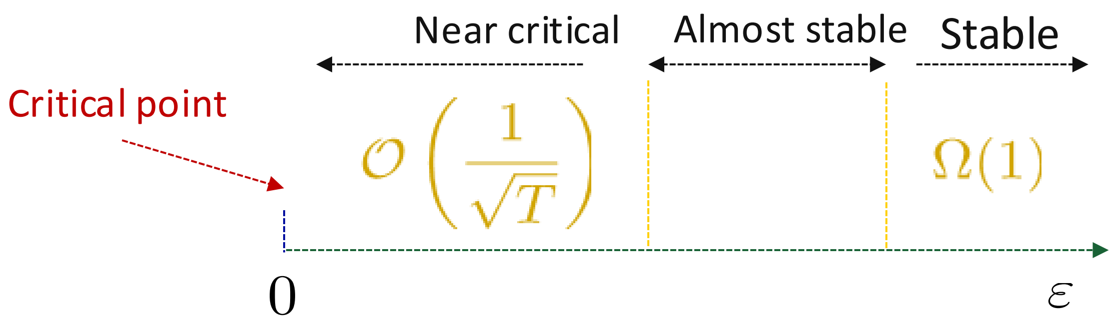
$$\min\left\{\mathcal{O}\left(\sqrt{T}+\frac{\gamma}{\epsilon}\right), \mathcal{O}(\sqrt{\gamma T})\right\}$$

Congestion-oblivious (CO) algorithm

Theorem [Main Result II]

If $\epsilon = \Omega(1)$, **CO algorithm** obtains a regret of:

$$\mathcal{O}\left(\sqrt{T}+\frac{\gamma}{\epsilon}\right)$$



Proposition [Lower-bound I] If $\gamma = \Omega(T)$, no online algorithm can achieve $o(T)$ regret

Proposition [Lower-bound II] If $\epsilon = \mathcal{O}(1/\sqrt{T})$ & $\gamma = \Omega(\sqrt{T})$, **CO algorithm** cannot achieve $o(T)$ regret

Theoretical Results

Congestion-aware (CA) algorithm

Theorem [Main Result I]

For $\forall\, \epsilon \geq 0$, **CA algorithm** obtains a regret

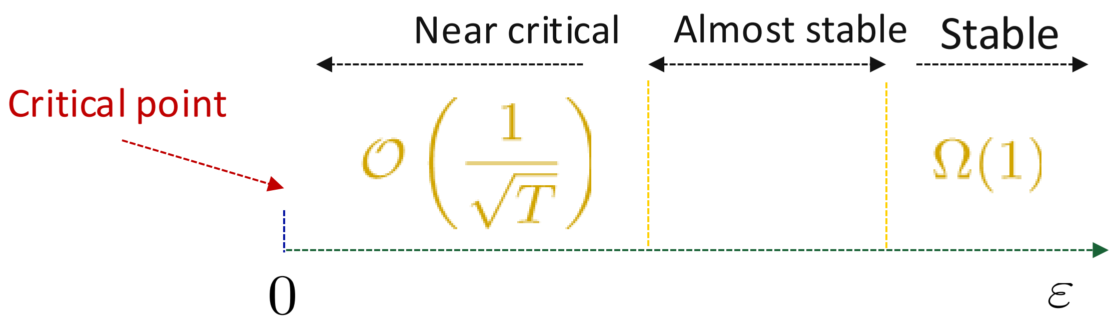
$$\min\left\{\mathcal{O}\left(\sqrt{T}+\frac{\gamma}{\epsilon}\right), \mathcal{O}(\sqrt{\gamma T})\right\}$$

Congestion-oblivious (CO) algorithm

Theorem [Main Result II]

If $\epsilon = \Omega(1)$, **CO algorithm** obtains a regret of:

$$\mathcal{O}\left(\sqrt{T}+\frac{\gamma}{\epsilon}\right)$$



Proposition [Lower-bound I] If $\gamma = \Omega(T)$, no online algorithm can achieve $o(T)$ regret

Proposition [Lower-bound II] If $\epsilon = \mathcal{O}(1/\sqrt{T})$ & $\gamma = \Omega(\sqrt{T})$, **CO algorithm** cannot achieve $o(T)$ regret

Takeaways:

- CA achieves sublinear regret whenever possible
- CO cannot achieve sublinear regret in near critical regime & γ "sufficiently" large

Theoretical Results

Congestion-aware (CA) algorithm

Theorem [Main Result I]

For $\forall\ \epsilon \geq 0$, **CA algorithm** obtains a regret

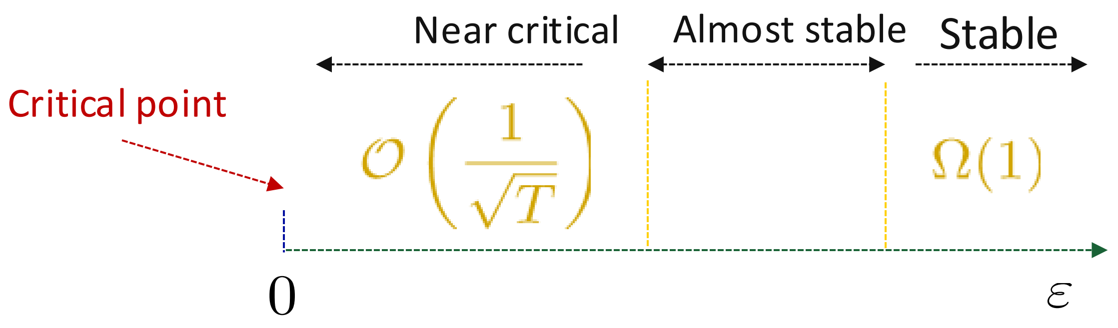
$$\min\left\{\mathcal{O}\left(\sqrt{T}+\frac{\gamma}{\epsilon}\right), \mathcal{O}\left(\sqrt{\gamma T}\right)\right\}$$

Congestion-oblivious (CO) algorithm

Theorem [Main Result II]

If $\epsilon = \Omega(1)$, **CO algorithm** obtains a regret of:

$$\mathcal{O}\left(\sqrt{T}+\frac{\gamma}{\epsilon}\right)$$



Proposition [Lower-bound I] If $\gamma = \Omega(T)$, no online algorithm can achieve $o(T)$ regret

Proposition [Lower-bound II] If $\epsilon = \mathcal{O}(1/\sqrt{T})$ & $\gamma = \Omega(\sqrt{T})$, **CO algorithm** cannot achieve $o(T)$ regret

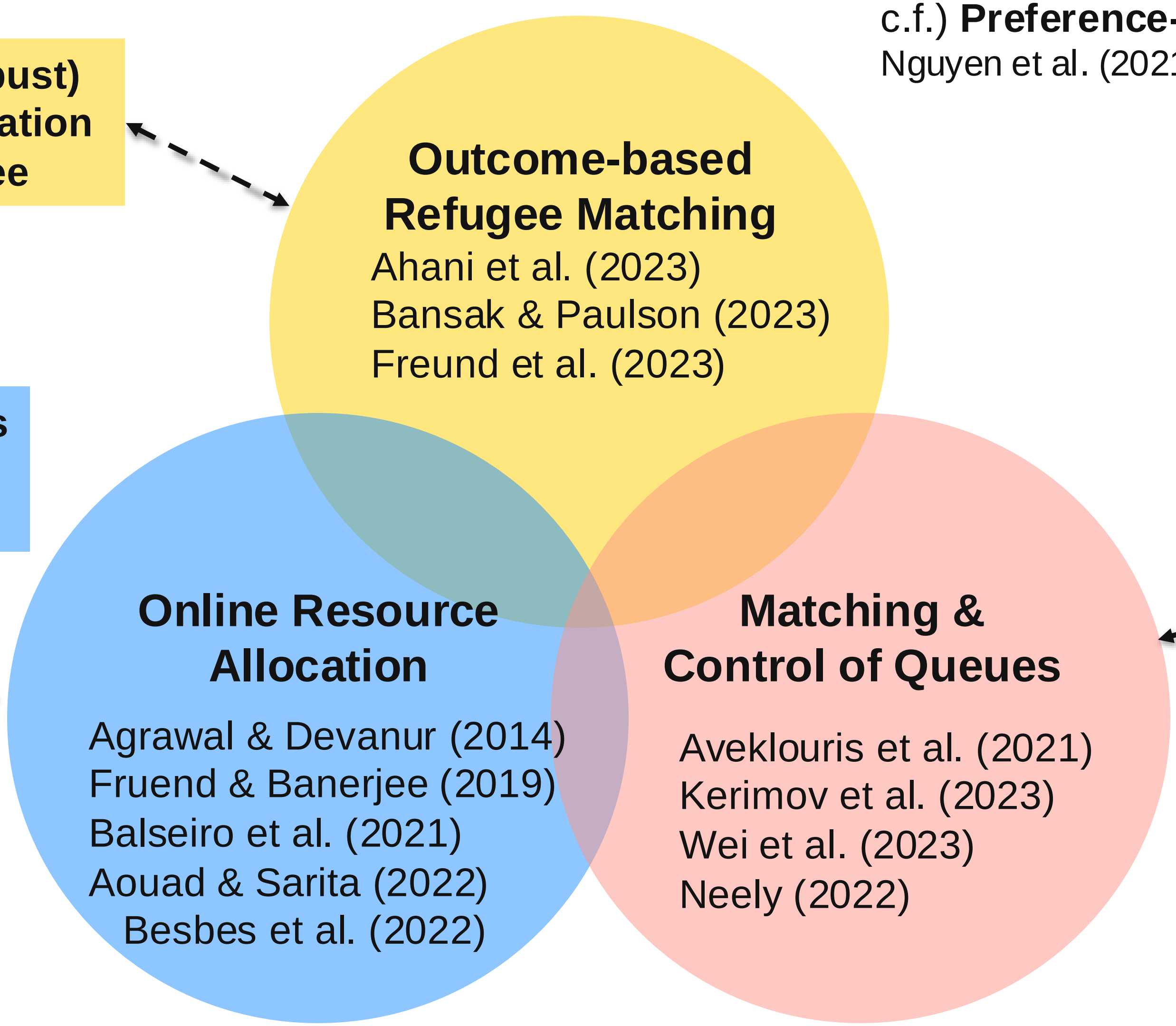
High-level Proof Ideas:

- CA: combines adversarial online learning & drift-analysis
- CO: establishes negative drift for backlog w.h.p. by proving high-probability last-iterate convergence of dual variables (Harvey et al. '19)

Related Literature

c.f.) **Preference-based Refugee Matching**:
Nguyen et al. (2021), Delacretaz et al. (2023)

✓ **Distribution-free (⇒ robust)**
✓ **Tied cases & over-allocation**
✓ **Theoretical guarantee**



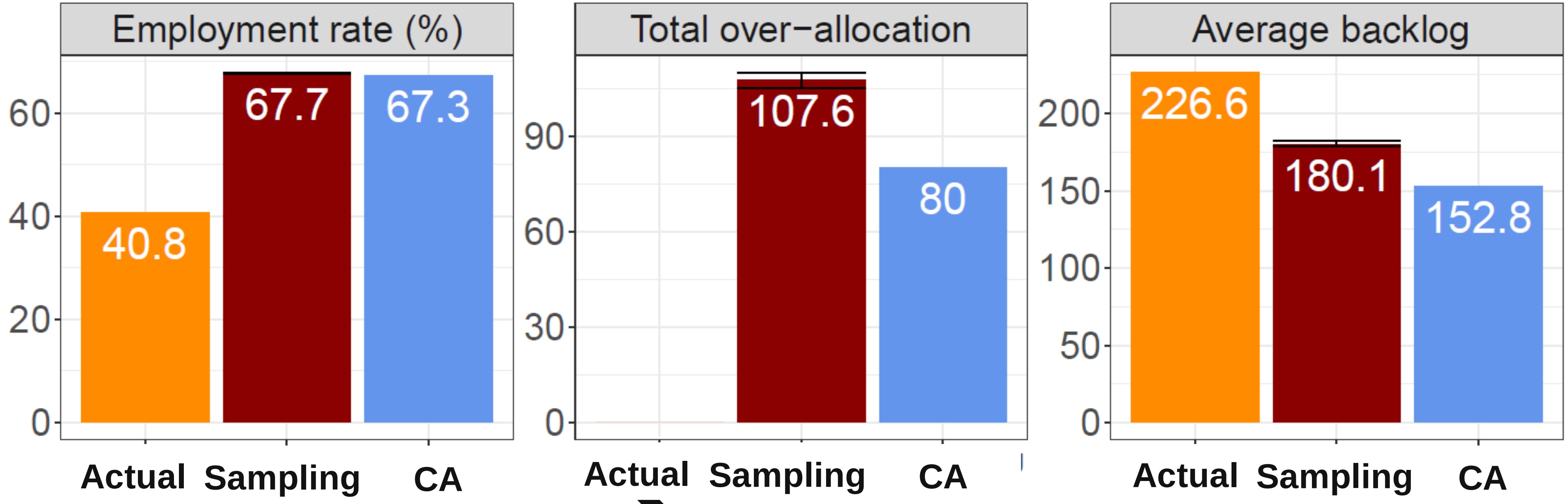
✓ **Post-allocation services (→ time-varying nature of dual problem)**

✓ **Reverse order of matching & waiting**

Case study

- Data: actual arrival sequence to resettlement agency in Year 2015
- Benchmarks: (1) **Actual** (actual historical placement) (2) **Sampling** (Bansak & Paulson '22) (simulate future arrival patterns from previous year's data) (3) **CA** (our algorithm)

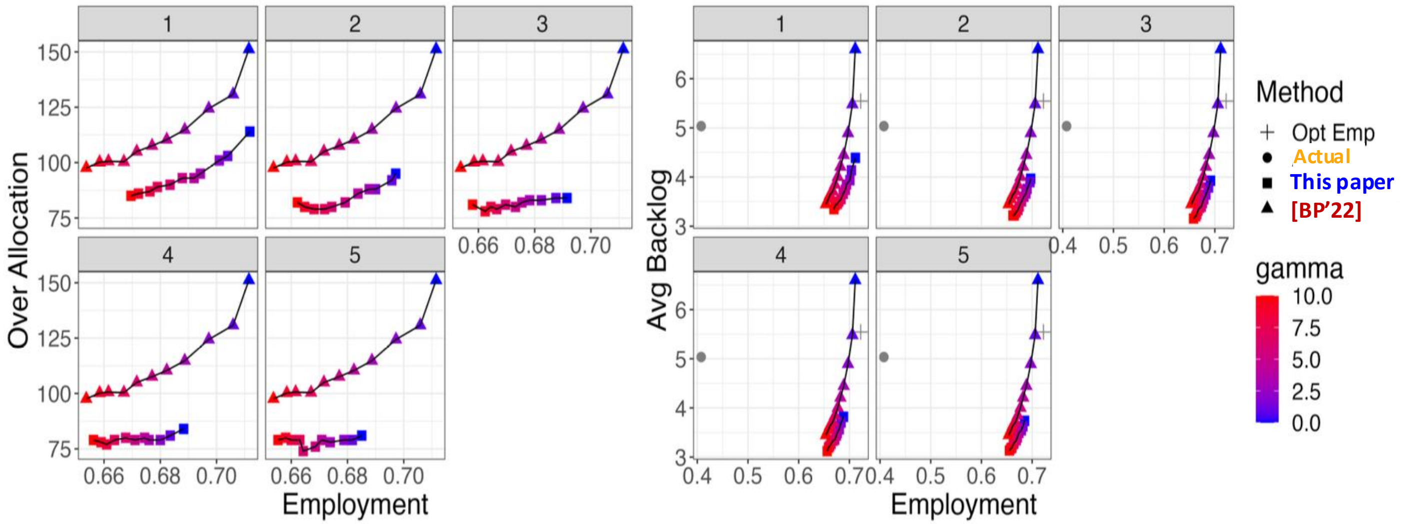
✓ **Significant improvement upon current practice & existing proposal**



(capacity = actual # of refugees resettled)

LIRS is becoming Global Refuge

Case study



(a) Year 2015

We can improve one outcome without hurting the other two!

- **Other practical benefits:**

	[this paper]	Existing Proposals
Robust (Free of history-based projection)?	Yes	No
Computationally fast?	Yes	No

Conclusion & future directions

Summary

- **Dynamic matching with post-allocation service**
  - Refugee resettlement: helping refugees & avoiding overburdening the service providers

- **Developing learning-based algorithms**
  - Distribution-free & near-optimal performance guarantee
  - Performance improvement over existing proposals + other practical benefits (check our paper!)

Future directions

- Beyond refugee matching: other applications of managing post-allocation service & congestion in healthcare (e.g. Shi et al. (2016); post-(bed) allocation service) & humanitarian services

Thank you!

Check out the paper for more details!

QR code for SSRN link