UNIVERSITY OF TWENTE.

DINALOG
Dutch Institute
for Advanced Logistics

ANTICIPATORY SCHEDULING OF FREIGHT IN A SYNCHROMODAL TRANSPORT NETWORK

Arturo E. Pérez Rivera & Martijn R.K. Mes

Department of Industrial Engineering and Business Information Systems
University of Twente, The Netherlands

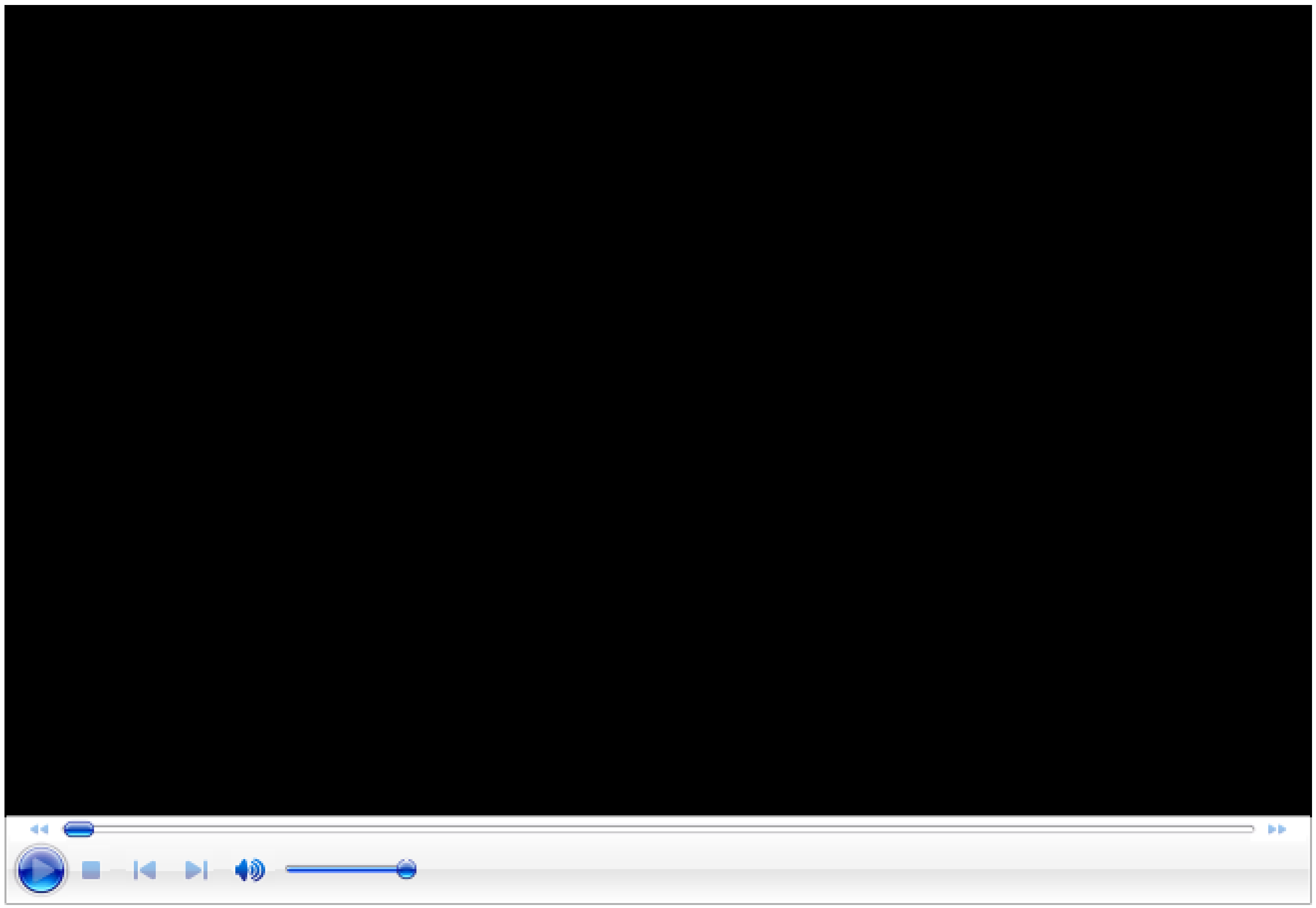
Beta Symposium 2017 - Tuesday, September 26th
Eindhoven, The Netherlands

CONTENTS

- Synchromodal transport
- Anticipatory scheduling problem:
  - *Markov Decision Process model*
- Heuristic policy:
  - *Approximate Dynamic Programming algorithm*
- Numerical experiments:
  - *Tuning and benchmark experiments*
- What to remember

SYNCHROMODAL TRANSPORT

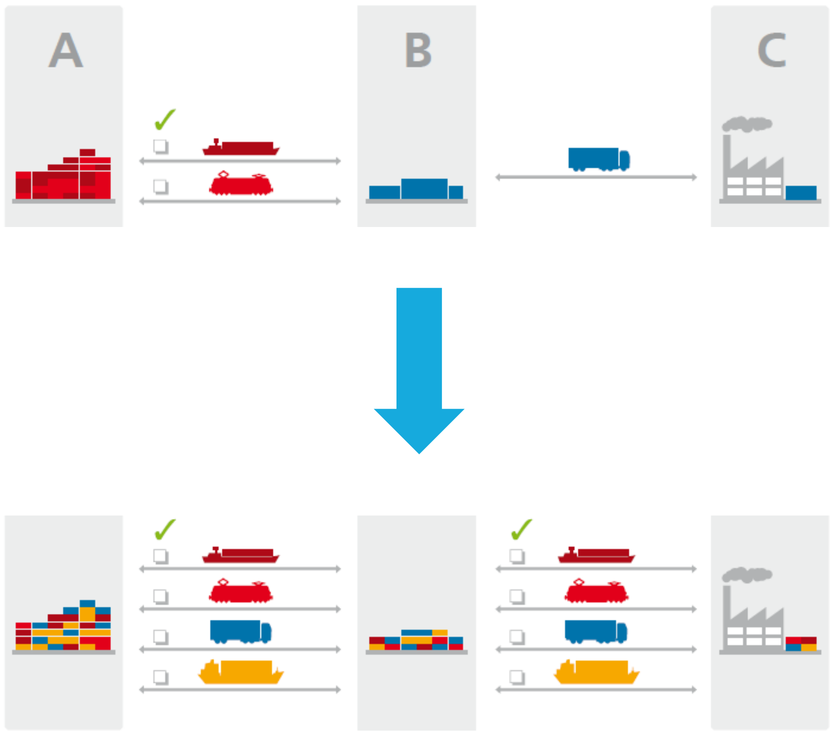
WHAT IS SYNCHROMODALITY?



**Source of video: Dutch Institute for Advanced Logistics (DINALOG) www.dinalog.nl*

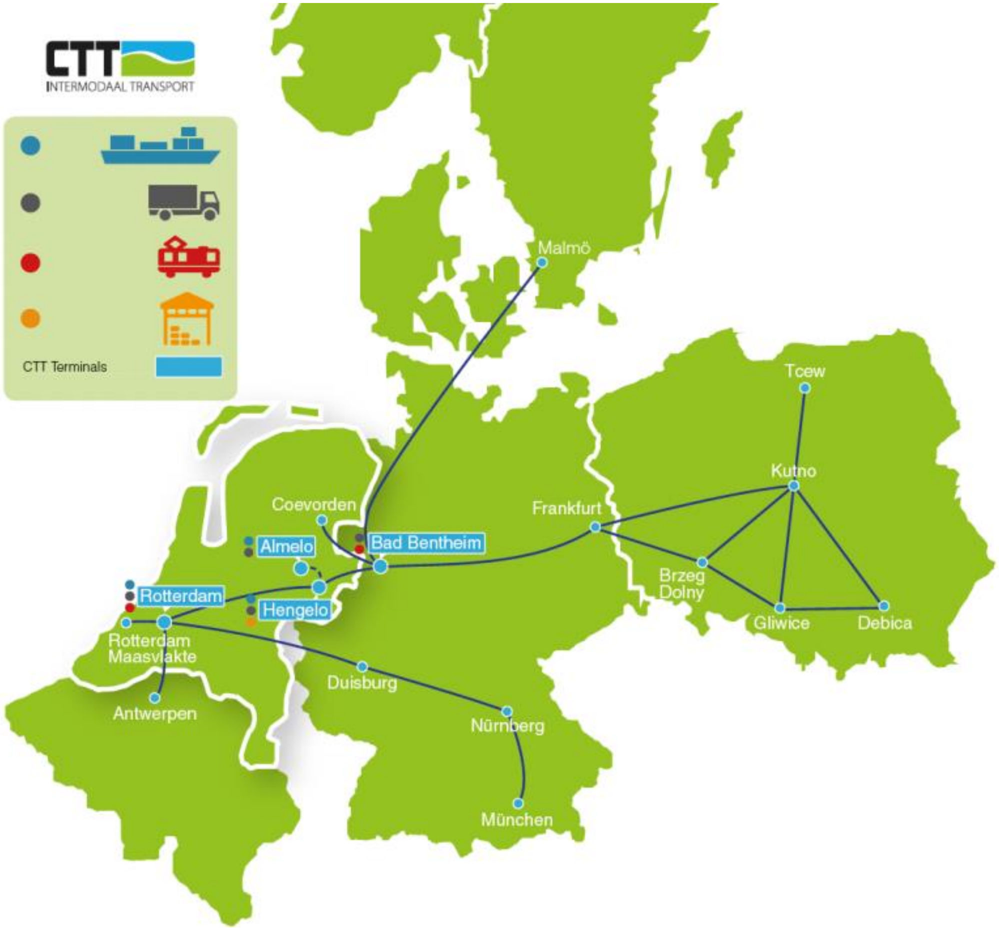
SYNCHROMODAL TRANSPORT
MAIN CHARACTERISTICS

- ***Mode-free booking*** for all freights.
- ***Network-wise scheduling*** at any point in time.
- ***Real-time information*** about the state of the network.
- ***Overall performance*** in both network and time.



**Source of artwork: European Container Terminals (ECT) – The future of freight transport (2011).*

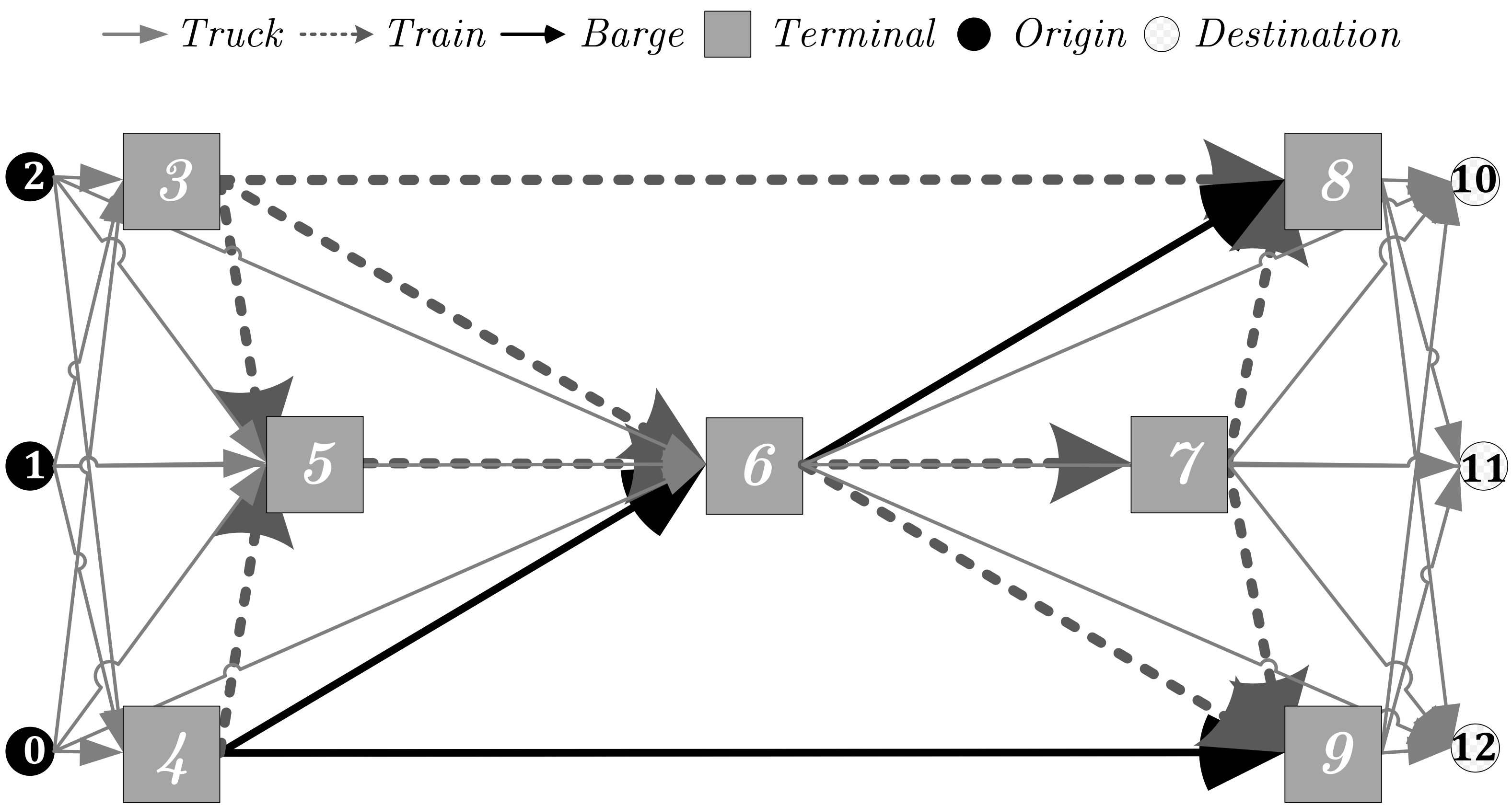
SYNCHROMODAL TRANSPORT
CASE: MOVING CONTAINERS TO/FROM THE HINTERLAND



*Source of artwork: Combi Terminal Twente (CTT) www.ctt-twente.nl

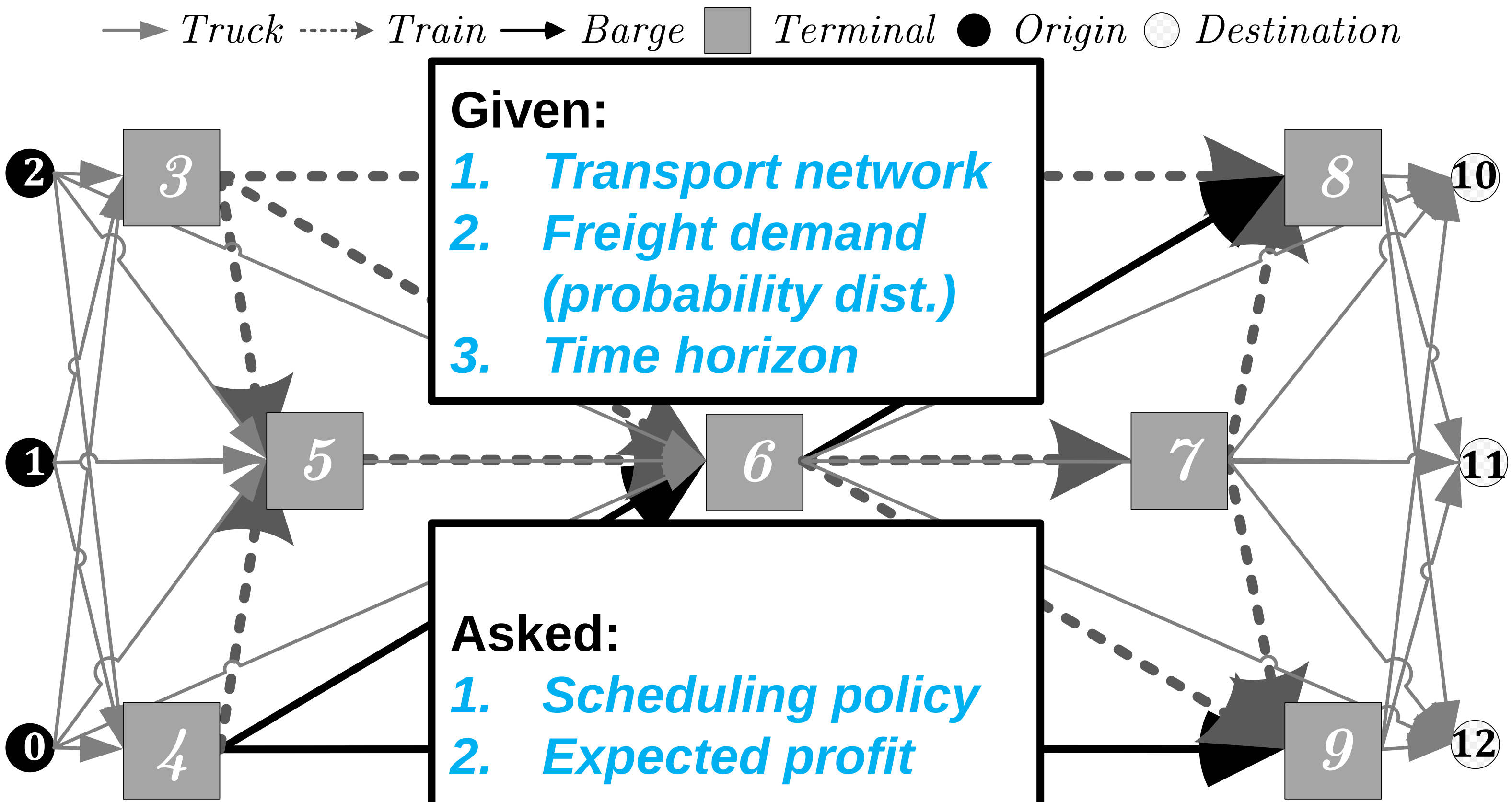
ANTICIPATORY SCHEDULING IN SYNCHROMODALITY

THE OPTIMIZATION PROBLEM



ANTICIPATORY SCHEDULING IN SYNCHROMODALITY

THE OPTIMIZATION PROBLEM



Given:

1. *Transport network*
2. *Freight demand (probability dist.)*
3. *Time horizon*

Asked:

1. *Scheduling policy*
2. *Expected profit*

MARKOV DECISION PROCESS (MDP) MODEL

OPTIMIZATION OF SEQUENTIAL DECISIONS UNDER UNCERTAINTY

State →

$$S_t = [F_{i,d,r,k,t}]_{\forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, r \in \mathcal{R}_t', k \in \mathcal{K}_t} \quad (1)$$

Decision

$$x_t = [x_{i,j,d,k,t}]_{\forall (i,j) \in \mathcal{A}_t, d \in \mathcal{N}_t^{\mathrm{D}}, k \in}$$

s.t.

$$\sum_{j \in \mathcal{N}_t^{\mathrm{I}} \cup \{d\}} x_{i,j,d,k,t} \le F_{i,d,0,k,t}, \quad \forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}$$

$$x_{i,d,d,L_{i,d,t}^{A},t} \ge F_{i,d,0,L_{i,d,t}^{A},t}, \quad \forall (i,d) \in$$

$$x_{i,j,d,k,t} = 0, \quad \forall (i,j) \in \mathcal{A}_t, d \in \mathcal{N}_t^{\mathrm{D}}, k \in \mathcal{K}_t | k$$

$$\sum_{d \in \mathcal{N}_t^{\mathrm{D}}} \sum_{k \in \mathcal{K}_t} x_{i,j,d,k,t} \le Q_{i,j,t}, \quad \forall (i,j$$

$$x_{i,j,d,k,t} \in \mathbb{N} \cup \{0\}, \quad \forall (i,j) \in \mathcal{A}_t, d \in$$

Exogenous Information

State Transition

$$S_t = S^M (S_{t-1}, x_{t-1}, W_t) \quad (4a)$$

s.t.

$$F_{i,d,0,k,t} = F_{i,d,0,k+1,t-1} - \sum_{j \in \mathcal{A}_t} x_{i,j,d,k+1,t-1} + F_{i,d,1,k,t-1} + \widetilde{F}_{i,d,0,k,t}, \quad (4b)$$

$$\forall i \in \mathcal{N}_t^{\mathrm{O}}, d \in \mathcal{N}_t^{\mathrm{D}}, k+1 \in \mathcal{K}_t$$

$$F_{i,d,0,K_t^{\max},t} = F_{i,d,1,K_t^{\max},t-1} + \widetilde{F}_{i,d,0,K_t^{\max},t}, \quad (4c)$$

$$\forall i \in \mathcal{N}_t^{\mathrm{O}}, d \in \mathcal{N}_t^{\mathrm{D}}$$

$$F_{i,d,0,k,t} = F_{i,d,0,k+1,t-1} - \sum_{j \in \mathcal{A}_t} x_{i,j,d,k+1,t-1} + F_{i,d,1,k,t-1}$$

$$+ \sum_{\mathcal{A}_t | M_{j,i,t}=1} x_{j,i,d,k+M_{j,i,t},t-1}, \quad (4d)$$

$$\forall i \in \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, k+1 \in \mathcal{K}_t$$

$$_{d,r,k,t} = F_{i,d,r+1,k,t-1} + \widetilde{F}_{i,d,r,k,t}, \quad (4e)$$

$$_t^{\mathrm{O}}, d \in \mathcal{N}_t^{\mathrm{D}}, r+1 \in \mathcal{R}_t | r \ge 1, k \in \mathcal{K}_t$$

$$F_{i,d,R_t^{\max},k,t} = \widetilde{F}_{i,d,R_t^{\max},k,t}, \quad (4f)$$

$$\forall i \in \mathcal{N}_t^{\mathrm{O}}, d \in \mathcal{N}_t^{\mathrm{D}}, k \in \mathcal{K}_t$$

$$_{,r+1,k,t-1} + \sum_{j \in \mathcal{A}_t | M_{j,i,t}=r+1} x_{j,i,d,k+M_{j,i,t},t-1}, \quad (4g)$$

$$_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, r+1 \in \mathcal{R}_t' | r \ge 1, k \in \mathcal{K}_t$$

$$_{,t} = \sum_{j \in \mathcal{A}_t | M_{j,i,t}=|\mathcal{R}_t'|+1} x_{j,i,d,k+M_{j,i,t},t-1}, \quad (4h)$$

$$\forall i \in \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, k \in \mathcal{K}_t$$

Reward Function

$$R_t(x_t) = \sum_{(i,j) \in \mathcal{A}_t} \sum_{d \in \mathcal{N}_t^{\mathrm{D}}} \left((A_{i,j,d,t} - C_{i,j,d,t}) \sum_{k \in \mathcal{K}_t} x_{i,j,d,k,t} \right) - \sum_{(i,j) \in \mathcal{A}_t} (B_{i,j,t} \cdot y_{i,j,t}) \quad (5a)$$

where

$$y_{i,j,t} = \begin{cases} 1, & \text{if } \sum_{d \in \mathcal{N}_t^{\mathrm{D}}} \sum_{k \in \mathcal{K}_t} (x_{i,j,d,k,t}) > 0 \\ 0, & \text{otherwise} \end{cases}, \quad \forall (i,j) \in \mathcal{A}_t \quad (5b)$$

Objective →

$$\max_{\pi \in \Pi} \mathbb{E} \left[\sum_{t \in \mathcal{T}} \gamma_t R_t(x_{t,\pi}) \middle| S_0 \right] \quad (6)$$

Bellman's Recursion

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} (R_t(x_t) + \gamma_t \mathbb{E}[V_{t+1}(S_{t+1})]), \quad \forall S_t \in \mathcal{S} \quad (7)$$

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} (R_t(x_t) + \gamma_t \mathbb{E}[V_{t+1}(S^M(S_t, x_t, W_{t+1}))]), \quad \forall S_t \in \mathcal{S} \quad (8)$$

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \left(R_t(x_t) + \gamma_t \sum_{\omega \in \Omega_{t+1}} p_\omega^{\Omega_{t+1}} (V_{t+1}(S^M(S_t, x_t, \omega))) \right), \quad \forall S_t \in \mathcal{S} \quad (9)$$

MDP MODEL – NETWORK EVOLUTION

A VIRTUAL TIME-WINDOW FOR FREIGHT TYPES

- Freight ***release-day r*** is relative to the current day *t*.
- Freight ***time-window length k*** is relative to the release-day *r*.
- Consider $F_{i,d,r,k,t}$ freights with $k=4$ sent from terminal *i* to terminal *j* using a service that lasts 2 days:

	t=7	t=8	t=9	t=10	t=11
	Monday	Tuesday	Wednesday	Thursday	Friday
i	$F_{i,d,0,4,7}$				
j		$F_{j,d,1,2,8}$	$F_{j,d,0,2,9}$		
d					$F_{d,d,0,0,11}$

MARKOV DECISION PROCESS (MDP) MODEL

OPTIMIZATION OF SEQUENTIAL DECISIONS UNDER UNCERTAINTY

State →

$$S_t = [F_{i,d,r,k,t}]_{\forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, r \in \mathcal{R}_t', k \in \mathcal{K}_t} \quad (1)$$

Decision

$$x_t = [x_{i,j,d,k,t}]_{\forall (i,j) \in \mathcal{A}_t, d \in \mathcal{N}_t^{\mathrm{D}}, k \in}$$

s.t.

$$\sum_{j \in \mathcal{N}_t^{\mathrm{I}} \cup \{d\}} x_{i,j,d,k,t} \le F_{i,d,0,k,t}, \quad \forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}$$

$$x_{i,d,d,L^A_{i,d,t},t} \ge F_{i,d,0,L^A_{i,d,t},t}, \quad \forall (i,d) \in$$

$$\begin{cases} 0, & \text{otherwise} \end{cases}$$

Exogenous Information

$$S_t = S^M(S_{t-1}, x_{t-1}, W_t) \quad (4a)$$

s.t.

$$F_{i,d,0,k,t} = F_{i,d,0,k+1,t-1} - \sum_{j \in \mathcal{A}_t} x_{i,j,d,k+1,t-1} + F_{i,d,1,k,t-1} + \widetilde{F}_{i,d,0,k,t}, \quad (4b)$$

$$\forall i \in \mathcal{N}_t^{\mathrm{O}}, d \in \mathcal{N}_t^{\mathrm{D}}, k+1 \in \mathcal{K}_t$$

$$K_t^{\max}, t, \quad (4c)$$

$$+ F_{i,d,1,k,t-1}$$

(4d)

(4e)

(4f)

$$\sum_{j \in \mathcal{A}_t | M_{j,i,t} = r+1} x_{j,i,d,k+M_{j,i,t},t-1}, \quad (4g)$$

$$\mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, r+1 \in \mathcal{R}_t' | r \ge 1, k \in \mathcal{K}_t$$

$$= \sum_{j \in \mathcal{A}_t | M_{j,i,t} = |\mathcal{R}_t'|+1} x_{j,i,d,k+M_{j,i,t},t-1}, \quad (4h)$$

$$\forall i \in \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, k \in \mathcal{K}_t$$

The three curses of dimensionality in our MDP:

1. ***All states***
2. ***All feasible decisions***
3. ***All realizations of uncertain demand***

Objective →

$$\max_{\pi \in \Pi} \mathbb{E} \left[\sum_{t \in \mathcal{T}} \gamma_t R_t(x_{t,\pi}) \middle| S_0 \right] \quad (6)$$

Bellman's Recursion

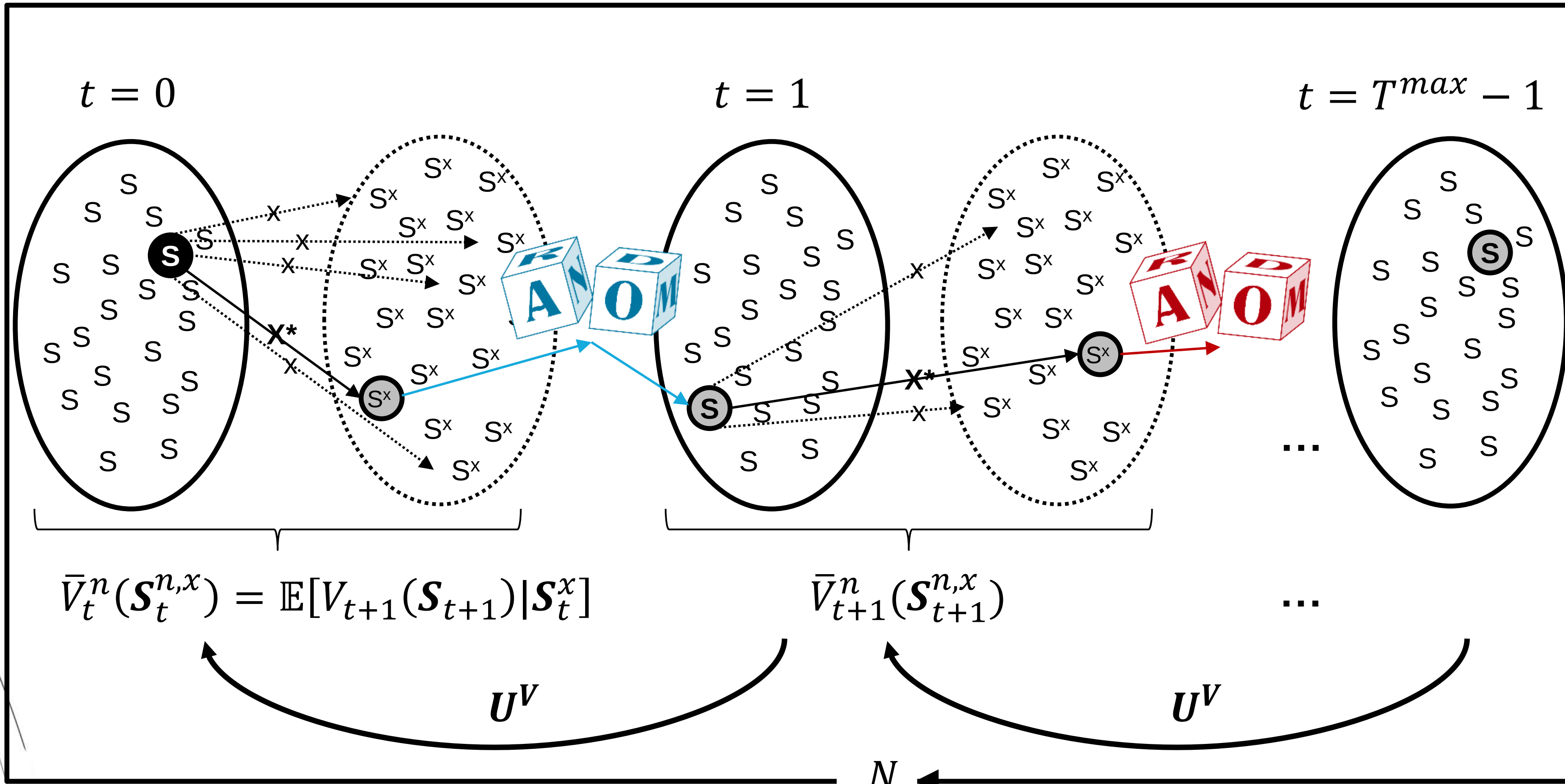
$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \left(R_t(x_t) + \gamma_t \mathbb{E}\left[V_{t+1}(S_{t+1})\right] \right), \quad \forall S_t \in \mathcal{S} \quad (7)$$

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \left(R_t(x_t) + \gamma_t \mathbb{E}\left[V_{t+1}\left(S^M(S_t, x_t, W_{t+1})\right)\right] \right), \quad \forall S_t \in \mathcal{S} \quad (8)$$

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \left(R_t(x_t) + \gamma_t \sum_{\omega \in \Omega_{t+1}} p_\omega^{\Omega_{t+1}} \left(V_{t+1}\left(S^M(S_t, x_t, \omega)\right)\right) \right), \quad \forall S_t \in \mathcal{S} \quad (9)$$

APPROXIMATE DYNAMIC PROGRAMMING (ADP)

HEURISTIC FRAMEWORK FOR SOLVING LARGE MARKOV MODELS.[1]



1. *For a comprehensive explanation see Powell (2010) Approximate Dynamic Programming.*

ADP – THE REDUCED DECISION SPACE $\mathcal{X}_t^{\mathrm{R}}$

RESTRICTED POLICIES (RP) 1 AND 2

RP 1:
Aggregated time-windows at each terminal.
Aggregated time-windows, destinations, and origins, at each origin.

$$x_t \in \mathcal{X}_t \tag{8a}$$

$$x_{i,d,d,k,t} = 0, \quad \forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, k \in \mathcal{K}_t | k > L_{i,d,t}^{A} \tag{8b}$$

$$x_{i,j,d,k,t} \geq \left(F_{i,d,0,k,t}\right)\left(x_{i,j,d,t}^{\mathrm{RG}} - M_{i,j,d,k,t}^{\mathrm{R}}\right), \quad \forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}, j \in \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, k \in \mathcal{K}_t \tag{8c}$$

$$\sum_{d \in \mathcal{N}_t^{\mathrm{D}}} \sum_{i \in \mathcal{N}_t^{\mathrm{O}}} x_{i,j,d,k,t} = \left|\mathcal{N}_t^{\mathrm{D}}\right| \left|\mathcal{N}_t^{\mathrm{O}}\right| x_{j,t}^{\mathrm{RO}}, \quad \forall j \in \mathcal{N}_t^{\mathrm{I}} | \exists_{\forall i \in \mathcal{N}_t^{\mathrm{O}}} (i,j) \in \mathcal{A}_t^{\mathrm{I}} \tag{8d}$$

RP 2:
Aggregated time-windows at terminals.
Aggregated time-windows and origins, at each origin.

$$x_t \in \mathcal{X}_t \tag{8a}$$

$$x_{i,d,d,k,t} = 0, \quad \forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, k \in \mathcal{K}_t | k > L_{i,d,t}^{A} \tag{8b}$$

$$x_{i,j,d,k,t} \geq \left(F_{i,d,0,k,t}\right)\left(x_{i,j,d,t}^{\mathrm{RG}} - M_{i,j,d,k,t}^{\mathrm{R}}\right), \quad \forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}, j \in \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}}, k \in \mathcal{K}_t \tag{8c}$$

$$\sum_{i \in \mathcal{N}_t^{\mathrm{O}}} x_{i,j,d,k,t} = \left|\mathcal{N}_t^{\mathrm{O}}\right| x_{j,d,t}^{\mathrm{RO}}, \quad \forall j \in \mathcal{N}_t^{\mathrm{I}} | \exists_{\forall i \in \mathcal{N}_t^{\mathrm{O}}} (i,j) \in \mathcal{A}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}} \tag{24}$$

ADP – THE REDUCED DECISION SPACE $\mathcal{X}_t^{\mathrm{R}}$

RESTRICTED POLICIES (RP) 1 AND 2

RP 1:

$$x_t \in \mathcal{X}_t \quad (8a)$$

Restricted Policy 1:

1. ***All released freights at a terminal, to a given destination, are scheduled together***
2. ***All released freights at all origins are scheduled together***

RP 2:

$$x_t \in \mathcal{X}_t \quad (8a)$$

Restricted Policy 2:

1. ***All released freights at a terminal, to a given destination, are scheduled together***
2. ***All released freights at all origins, to a given destination, are scheduled together***

ADP – THE VALUE FUNCTION APPROXIMATION (VFA)

PARAMETRIC APPROXIMATION OF DOWNSTREAM REWARDS

VFA →

$$\overline{V}_t^n\left(S_t^{x,n}\right) = \sum_{b \in \mathcal{B}} \theta_{b,t}^n \phi_{b,t}\left(S_t^{x,n}\right) = \phi_t\left(S_t^{x,n}\right)^T \theta_t^n \tag{11}$$

Basis functions

$$\phi_{b(i,d)}\left(S_t^{x,n}\right) = \sum_{k \in \mathcal{K}_t | k < \Psi} \sum_{r \in \mathcal{R}'_t} F_{i,d,r,k,t'}^{x,n}, \quad \forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}} \tag{12a}$$

$$\phi_{b'(i,d)}\left(S_t^{x,n}\right) = \sum_{k \in \mathcal{K}_t | k \geq \Psi} \sum_{r \in \mathcal{R}'_t} F_{i,d,r,k,t'}^{x,n}, \quad \forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}} \tag{12b}$$

$$\phi_{b''(d)}\left(S_t^{x,n}\right) = \sum_{i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}} \sum_{k \in \mathcal{K}_t | k \geq \Psi} \sum_{r \in \mathcal{R}'_t} F_{i,d,r,k,t'}^{x,n}, \quad \forall d \in \mathcal{N}_t^{\mathrm{D}} \tag{12c}$$

$$\phi_{|\mathcal{B}|}\left(S_t^{x,n}\right) = 1 \tag{12d}$$

Recursive least square method for updating the VFA

$$\overline{V}_t^n(S_t^{n,x*}) := U_t^n(\overline{V}_t^{n-1}(S_t^{n,x*}), S_t^{n,x*}, [\hat{v}_t^n]_{\forall t \in \mathcal{T}}) \tag{13a}$$

$$\text{s.t.}$$

$$\theta_t^n = \theta_t^{n-1} - \left(H_t^n\right)^{-1} \phi_t\left(S_t^{x,n}\right) \left(\overline{V}_{t-1}^{n-1}\left(S_{t-1}^{x,n}\right) - \sum_t^{T^{\max}-1} \hat{v}_t^n\right) \tag{13b}$$

$$H_t^n = \lambda^n H_t^{n-1} + \phi_t\left(S_t^{x,n}\right) \phi_t\left(S_t^{x,n}\right)^T \tag{13c}$$

$$\lambda^n = 1 - \frac{\lambda}{n} \tag{13d}$$

ADP – THE VALUE FUNCTION APPROXIMATION (VFA)

PARAMETRIC APPROXIMATION OF DOWNSTREAM REWARDS

VFA →

$$\overline{V}_t^n\left(S_t^{x,n}\right) = \sum_{b\in\mathcal{B}} \theta_{b,t}^n \phi_{b,t}\left(S_t^{x,n}\right) = \phi_t\left(S_t^{x,n}\right)^T \theta_t^n \tag{11}$$

$$\phi_{b(i,d)}\left(S_t^{x,n}\right) = \sum_{k\in\mathcal{K}_t|k<\Psi} \sum_{r\in\mathcal{R}'_t} F_{i,d,r,k,t'}^{x,n}, \quad \forall i \in \mathcal{N}_t^{\mathrm{O}} \cup \mathcal{N}_t^{\mathrm{I}}, d \in \mathcal{N}_t^{\mathrm{D}} \tag{12a}$$

The features of a post-decision state:

1. ***Intermodal-path freights*** per location, per destination.
2. ***Trucking freights*** per location, per destination.
3. ***Total freights*** per destination.
4. ***Constant***.

Recursive least square method for updating the VFA

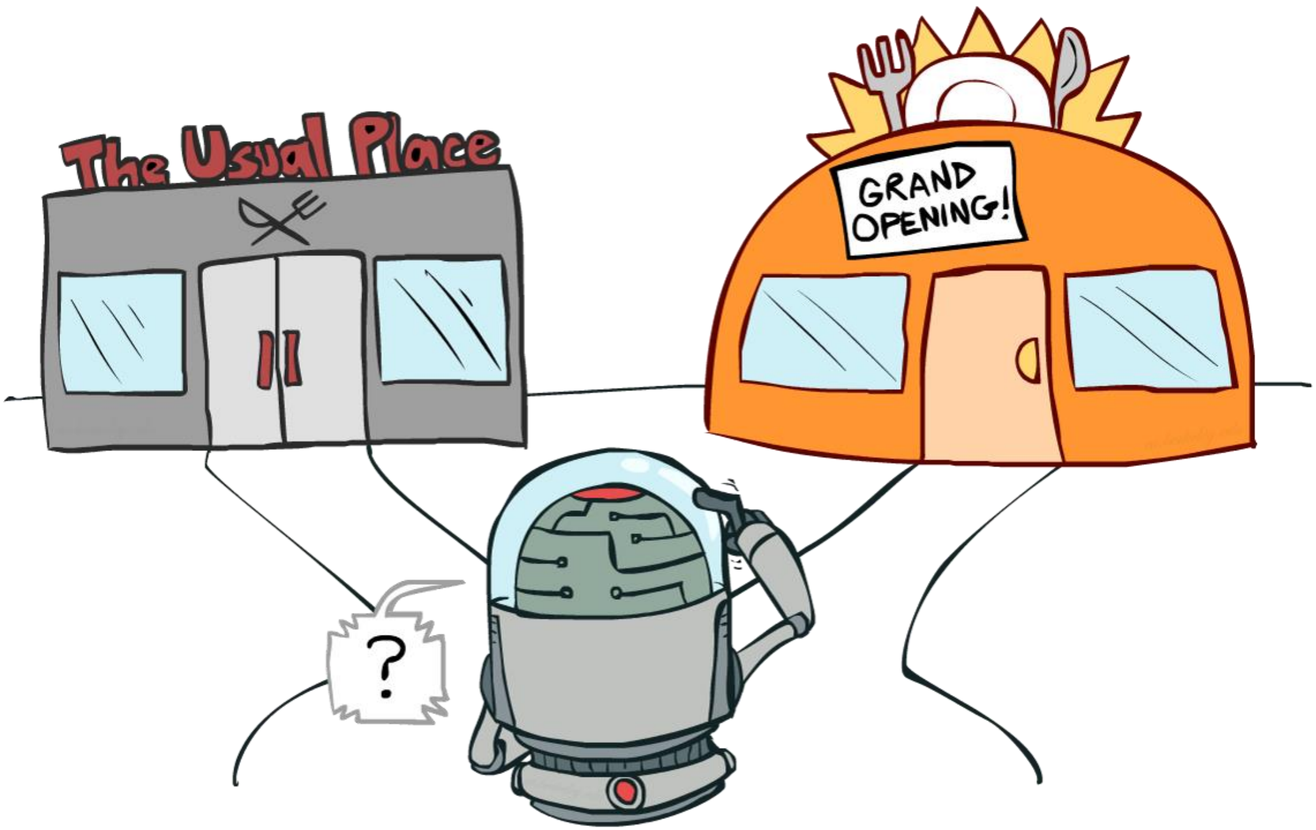
s.t.

$$\theta_t^n = \theta_t^{n-1} - \left(H_t^n\right)^{-1} \phi_t\left(S_t^{x,n}\right)\left(\overline{V}_{t-1}^{n-1}\left(S_{t-1}^{x,n}\right) - \sum_{t}^{T^{\max}-1} \widehat{v}_t^n\right) \tag{13b}$$

$$H_t^n = \lambda^n H_t^{n-1} + \phi_t\left(S_t^{x,n}\right)\phi_t\left(S_t^{x,n}\right)^T \tag{13c}$$

$$\lambda^n = 1 - \frac{\lambda}{n} \tag{13d}$$

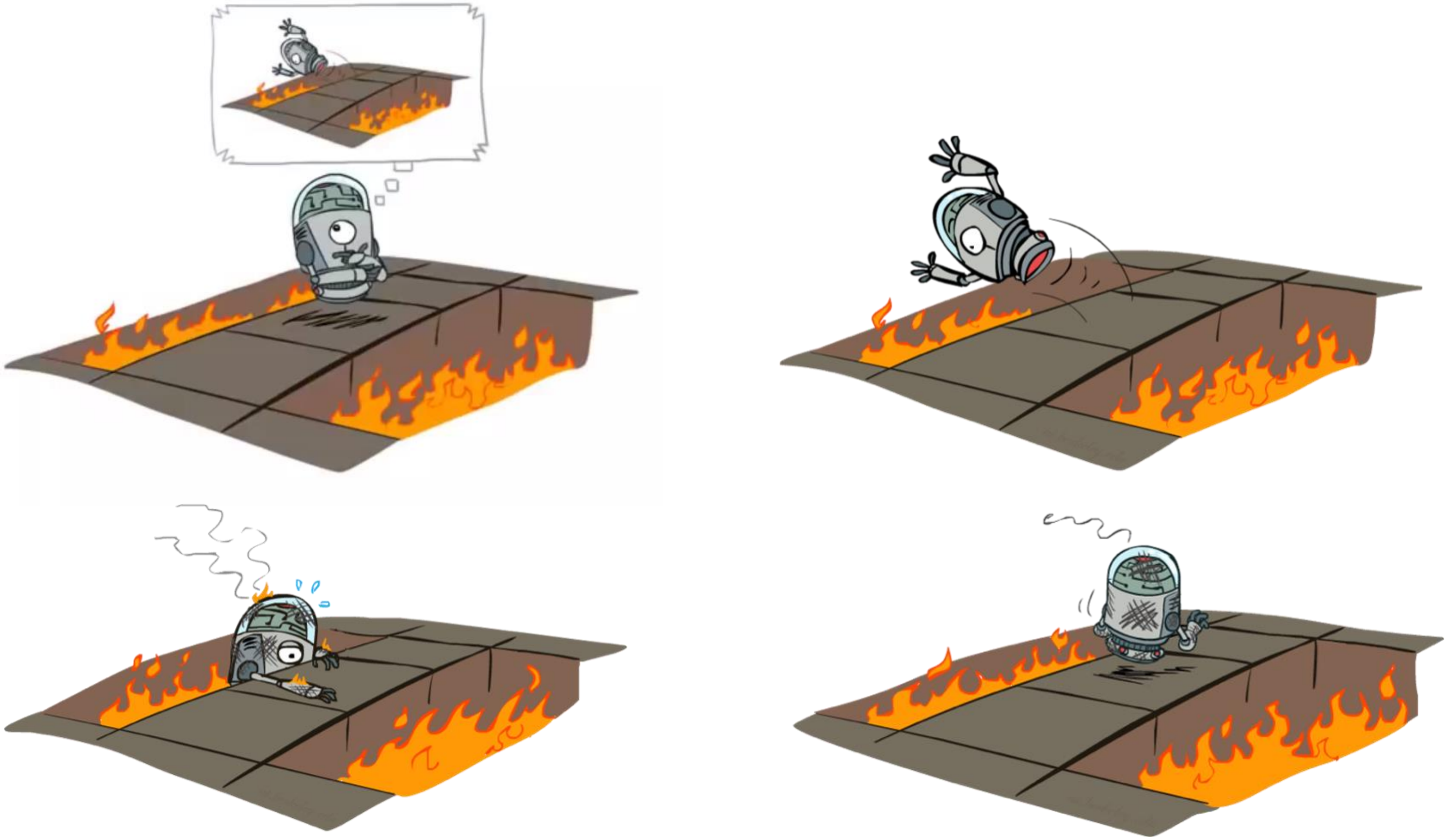
ADP – EXPLORATION VS. EXPLOITATION
ESCAPING LOCAL OPTIMA …



**Source of artwork: Dan Klein and Pieter Abbeel – Reinforcement Learning (2013), University of California at Berkeley*

ADP – EXPLORATION VS. EXPLOITATION

… OR AVOIDING LOCAL NADIR!



**Source of artwork: Dan Klein and Pieter Abbeel – Reinforcement Learning (2013), University of California at Berkeley*

ADP – EPSILON GREEDY EXPLORATION

ESCAPING LOCAL OPTIMA

Algorithm 1 ADP Algorithm

Initialize $\left[\overline{V}_t^0\right]_{\forall t \in \mathcal{T}}$

for $n = 1$ **to** N **do**

  $S_0^n := S_0$

  for $t = 0$ **to** $T^{max} - 1$ **do**

    $x_t^{n*} := \arg\max_{x_t^n \in \mathcal{X}_t^{\mathrm{R}}} \left(R_t\left(x_t^n\right) + \gamma_t \overline{V}_t^{n-1}\left(S^{M,x}\left(S_t^n, x_t^n\right)\right)\right)$

    $S_t^{n,x*} := S^{M,x}\left(S_t^n, x_t^{n*}\right)$

    $\widehat{v}_t^n := \left(R_t\left(x_t^{n*}\right) + \gamma_t \bar{V}_t^{n-1}\left(S_t^{n,x*}\right)\right)$

    $W_{t+1}^n := \mathrm{Random}\left(\Omega\right)$

    $S_{t+1}^n := S^M\left(S_t^n, x_t^{n*}, W_{t+1}^n\right)$

  end for

  for $t = T^{max} - 1$ **to** 0 **do**

    $\overline{V}_t^n(S_t^{n,x*}) := U_t^n(\overline{V}_t^{n-1}(S_t^{n,x*}), S_t^{n,x*}, [\widehat{v}_t^n]_{\forall t \in \mathcal{T}})$

  end for

end for

return $\left[\overline{V}_t^N\right]_{\forall t \in \mathcal{T}}$

ADP – EPSILON GREEDY EXPLORATION

ESCAPING LOCAL OPTIMA

Algorithm 1 ADP Algorithm

Initialize $\left[\overline{V}_t^0\right]_{\forall t \in \mathcal{T}}$
for $n = 1$ **to** N **do**
  $S_0^n := S_0$
  for $t = 0$ **to** $T^{max} - 1$ **d**
    $x_t^{n*} := \underset{x_t^n \in \mathcal{X}_t^{\mathrm{R}}}{\arg\max} \left(\ldots (x \right.$
    $S_t^{n,x*} := S^{M,x}\left(S_t^n, x_t^{n}\right.$
    $\widehat{v}_t^n := \left(R_t\left(x_t^{n*}\right) + \gamma_t \bar{V}_t\right.$
    $W_{t+1}^n := \mathrm{Random}\left(\Omega\right)$
    $S_{t+1}^n := S^M\left(S_t^n, x_t^{n*}, W_{t+1}^n\right)$
  end for
  for $t = T^{max} - 1$ **to** 0 **do**
    $\overline{V}_t^n(S_t^{n,x*}) := U_t^n(\overline{V}_t^{n-1}(S_t^{n,x*}), S_t^{n,x*}, [\widehat{v}_t^n]_{\forall t \in \mathcal{T}})$
  end for
end for
return $\left[\overline{V}_t^N\right]_{\forall t \in \mathcal{T}}$

Algorithm 2 ϵ-greedy strategy for exploration

if Random $[0,1) < \epsilon$ **then**
  $x_t^{n*} := \mathrm{Random}\left(\mathcal{X}_t^{\mathrm{R}}\right)$
else
  $x_t^{n*} := \underset{x_t^n \in \mathcal{X}_t^{\mathrm{R}}}{\arg\max} \left(R_t\left(x_t^n\right) + \gamma_t \bar{V}_t^{n-1}\left(S^{M,x}\left(S_t^n, x_t^n\right)\right)\right)$
end if

ADP – VALUE OF PERFECT INFORMATION (VPI)

EXPLORATION BASED ON A BAYESIAN BELIEF

Exploration decision

$$x_t^{n*} = \underset{x_t^n \in \mathcal{X}_t^{\mathrm{R}}}{\arg\max} \left(v_t^{E,n}(K_t^n, S_t^n x_t^n) \right) \tag{14}$$

Bayesian belief

$$K_t^n = (\overline{V}_t^n, C_t^n) = (\phi_t, \theta_t^n, C_t^n) \tag{15}$$

Value of exploration

$$v_t^{E,n}(K_t^n, S_t^n, x_t^n) = \sqrt{\sigma_t^{2,n}(K_t^n, S_t^{x,n})} f \left(- \frac{\delta(S_t^{x,n})}{\sqrt{\sigma_t^{2,n}(K_t^n, S_t^{x,n})}} \right) \tag{16a}$$

s.t.

$$\delta(S_t^{x,n}) = \left| \overline{V}_t^{x,n} \left(S_t^{x,n}\right) - \max_{y_t^n \in \mathcal{X}_t^{\mathrm{R}} | y \neq x_t^n} \overline{V}_t^{x,n} \left(S_t^{y,n}\right) \right| \tag{16b}$$

$$\sigma_t^{2,n}(K_t^n, S_t^{x,n}) = \phi\left(S_t^{x,n}\right)^T C_t^n \phi\left(S_t^{x,n}\right) \tag{16c}$$

Update VFA and belief

$$\theta_t^n = \theta_t^{n-1} - \frac{\left(\theta_t^{n-1}\right)^T \phi\left(S_t^{x,n}\right) - \sum_t^{T^{\max}-1} \widehat{v}_t^n}{\sigma^{2,\mathrm{E}} + \sigma_t^{2,n-1}\left(S_t^{x,n}\right)} C_t^n \phi\left(S_t^{x,n}\right) \tag{17}$$

$$C_t^n = C_t^{n-1} - \frac{C_t^{n-1} \phi\left(S_t^{x,n}\right) \phi\left(S_t^{x,n}\right)^T C_t^{n-1}}{\sigma^{2,\mathrm{E}} + \sigma_t^{2,n}\left(S_t^{x,n-1}\right)} \tag{18}$$

ADP – VALUE OF PERFECT INFORMATION (VPI)

EXPLORATION BASED ON A BAYESIAN BELIEF

Exploration decision

$x_t^{n*} = \arg$ (14)

(15)

(16a)

(16b)

(16c)

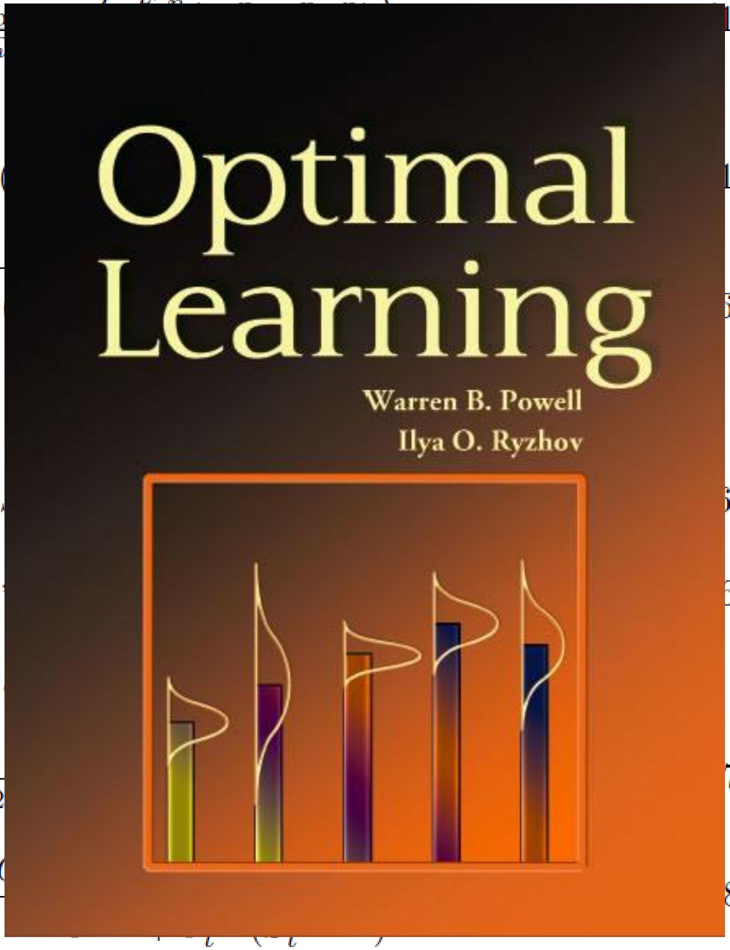
(17)

and belief

$C_t^n = C_t^{n-1} -$ (18)

Dearden et al. (1999) Model based Bayesian exploration:

The value of perfect information is the ***expected improvement in future decision quality*** arising from the information acquired by exploration.



ADP – VPI MODIFICATIONS

BE MORE CONSERVATIVE IN EXPLORATION AND UPDATING

1. Exploration decisions that focus on more than just the value of exploration:

$$x_t^{\mathrm{E2}} = \arg\max \left(\overline{V}_t^{x,n}\left(S_t^{x,n}\right) + v_t^{E,n}(S_t^n, K_t^n, x_t) \right)$$

$$x_t^{\mathrm{E3}} = \arg\max \left(R_t\left(S_t^n, x_t\right) + \overline{V}_t^{x,n}\left(S_t^{x,n}\right) + v_t^{E,n}(S_t^{x,n}, K_t^n, x_t) \right)$$

$$x_t^{\mathrm{E4}} = \arg\max \left((1-\alpha^n) \left(R_t\left(S_t^n, x_t\right) + \overline{V}_t^{x,n}\left(S_t^{x,n}\right) \right) + \alpha^n v_t^{E,n}(S_t^{x,n}, K_t^n, x_t) \right)$$

2. Update VFA and belief with stage or post-decision state dependent noise:

$$\sigma_t^{2,\mathrm{E2}} = \frac{T^{\max} - t}{T^{\max}} \eta^{\mathrm{E}}$$

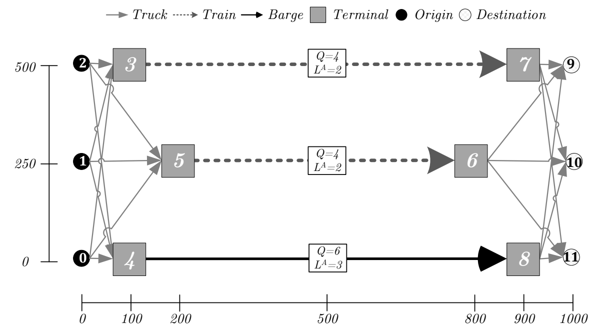
$$\sigma_t^{2,\mathrm{E3}} = \sigma_t^{2,n}(S_t^{x,n})$$

$$\sigma_{t,n}^{2,\mathrm{E4}} = \frac{T^{\max} - t}{T^{\max}} \eta^{\mathrm{E}} + \sigma_t^{2,n}(S_t^{x,n})$$

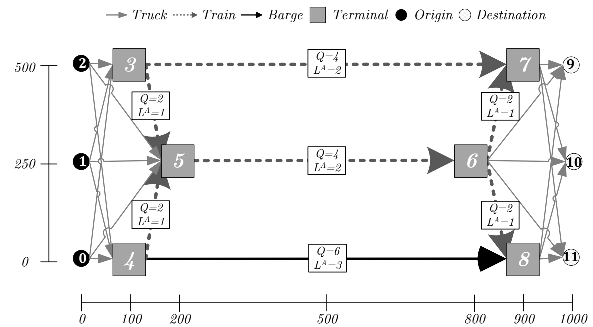
NUMERICAL EXPERIMENTS
EXPERIMENTAL SETUP

i. Tuning experiments:
- Relatively certain freight demand.
- Test all ADP designs, and tune their parameters.
- ***Goal: define the best ADP design.***

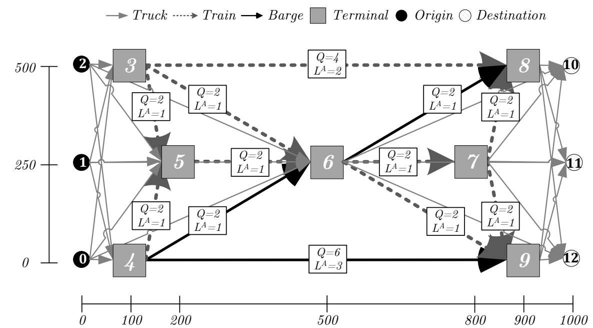
ii. Benchmark experiments:
- Various uncertain demand profiles.
- Compare against two smart benchmark heuristics.
- ***Goal: study the gains (or losses) of using our ADP design.***



Network 1



Network 2



Network 3

NUMERICAL EXPERIMENTS

TUNING EXP. [1/3]: MAXIMUM REALIZED REWARD PER ADP DESIGN

Performance:

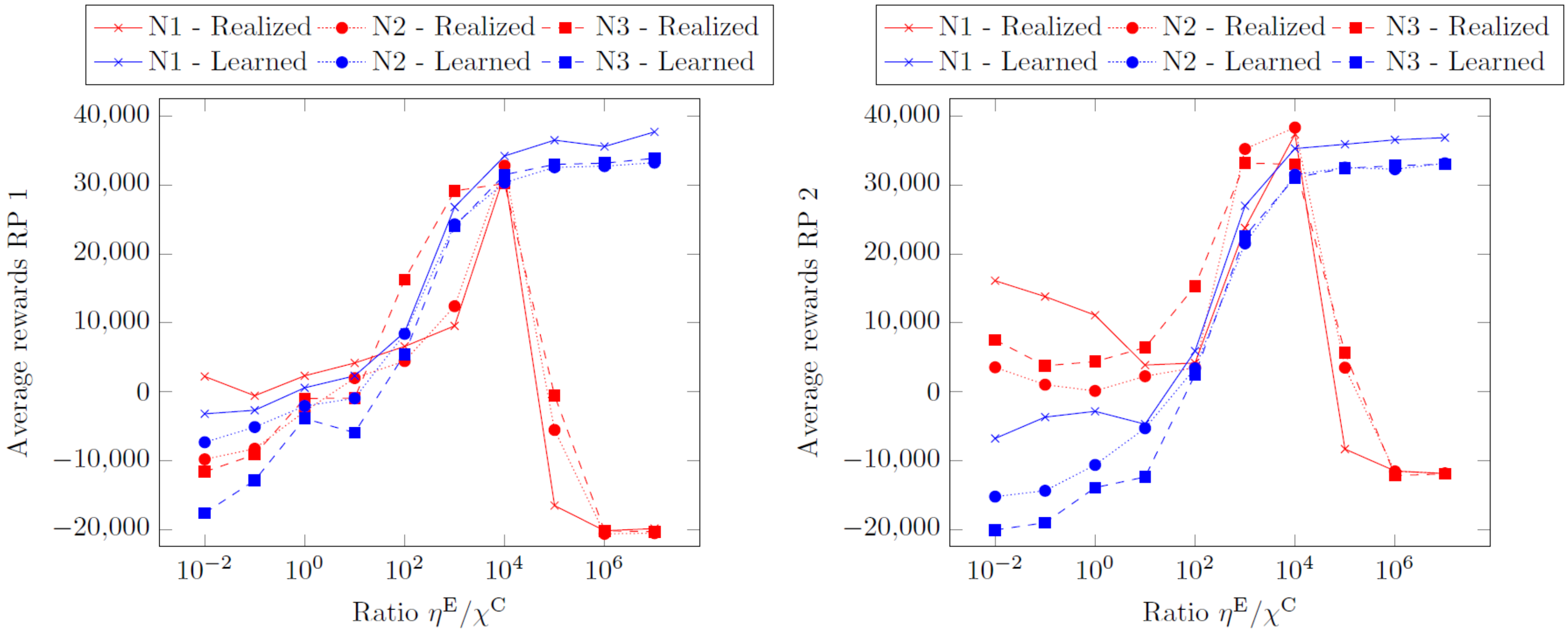
Learned rewards are those ADP thinks the resulting policy will achieve.
Realized rewards are the actual rewards (i.e., profit) achieved in a simulation of the resulting policy.

ADP Design		Network 1		Network 2		Network 3	
		Realized	Learned	Realized	Learned	Realized	Learned
RP 1	*BF*	*-7,994*	*38,219*	*-11,247*	*33,720*	*-16,548*	*-17,928*
	BF + ε-greedy	*-4,628*	*-6,984*	*-11,485*	*33,228*	*-18,172*	*-18,507*
	BF + VPI	*34,044*	*36,571*	*34,284*	*29,493*	*34,898*	*23,285*
RP 2	*BF*	*-4,912*	*-3,803*	*-11,734*	*34,060*	*-11,949*	*34,495*
	BF + ε-greedy	*880*	*37,386*	*-11,450*	*-12,091*	*-11,949*	*33,356*
	BF + VPI	*40,439*	*35,407*	*40,195*	*31,107*	*38,314*	*30,791*

Suppose that there are two freights at each location in Network 3, then :
RP 1 *has 1.9x10^4 decisions, or* ***0.01 % of decision space!***
RP 2 *has 5.8x10^4 decisions, or* ***0.02 % of decision space!***

NUMERICAL EXPERIMENTS

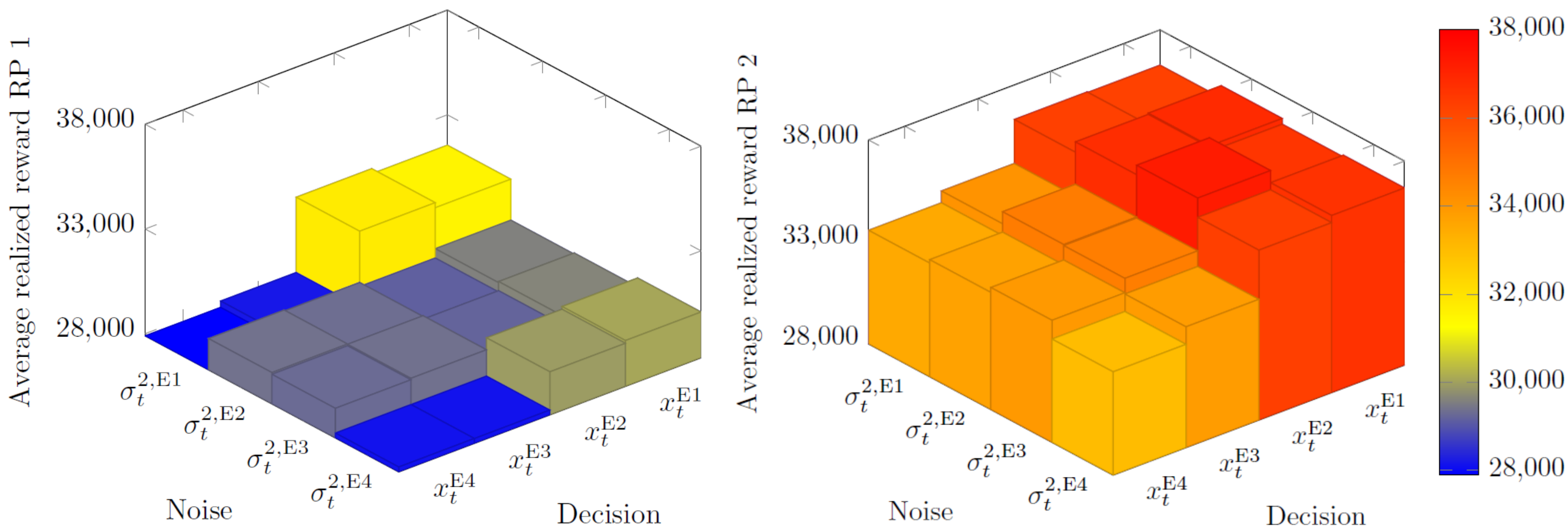
TUNING EXP. [2/3]: NOISE AND COVARIANCE PARAMETERS IN VPI



Similar to other literature of VPI, we observed that there seems to be an ***optimal ratio of the noise parameters*** *in our problem, around 10^4.*

NUMERICAL EXPERIMENTS

TUNING EXP. [3/3]: PROPOSED VPI MODIFICATIONS OVER ALL NETWORKS



*From our proposed modifications, **including the downstream rewards in the exploration decision** and **updating with a noise term equal to the variance of a post-decision state** seems to perform the best.*

NUMERICAL EXPERIMENTS

BENCHMARK EXP. [1/2]: DEMAND PROFILES WITH VARIOUS TIME-WINDOW DIST.

Release-day (RD) : 0, 1, 2 days
Time-window (TW) length: 4, 5, 6 days

Average realized rewards (over ten replications)

RD distribution	TW distribution	Network 1 BH	Network 1 ADP	Network 2 BH	Network 2 ADP	Network 3 BH	Network 3 ADP
Short	Short	*17,862*	*12,131*	*12,339*	*11,289*	*22,191*	*(19)*
	Medium	*25,286*	*22,775*	*18,232*	*23,486*	*26,634*	*15,001*
	Long	*33,007*	*35,111*	*25,805*	*32,524*	*30,680*	*29,745*
Medium	Short	*17,812*	*10,938*	*12,160*	*9,877*	*22,141*	*209*
	Medium	*25,302*	*22,267*	*18,052*	*23,015*	*26,573*	*14,612*
	Long	*32,805*	*34,508*	*25,420*	*31,806*	*30,473*	*29,502*
Long	Short	*17,773*	*10,724*	*12,062*	*9,281*	*22,256*	*(44)*
	Medium	*25,276*	*21,956*	*17,951*	*23,422*	*26,568*	*14,167*
	Long	*32,876*	*34,511*	*25,401*	*32,462*	*30,467*	*29,274*

NUMERICAL EXPERIMENTS

BENCHMARK EXP. [2/2]: DEMAND PROFILES WITH VARIOUS TIME-WINDOW DIST.

Release-day (RD) : 0, 1, 2 days
Time-window (TW) length: 4, 5, 6 days

Average realized rewards (over ten replications)

RD distribution	TW distribution	Network 1 BH+RP	Network 1 ADP	Network 2 BH+RP	Network 2 ADP	Network 3 BH+RP	Network 3 ADP
Short	Short	*9,273*	*12,131*	*9,014*	*11,289*	*9,374*	*(19)*
	Medium	*9,677*	*22,775*	*9,244*	*23,486*	*9,537*	*15,001*
	Long	*10,151*	*35,111*	*9,601*	*32,524*	*9,728*	*29,745*
Medium	Short	*9,322*	*10,938*	*9,003*	*9,877*	*9,494*	*209*
	Medium	*9,814*	*22,267*	*9,338*	*23,015*	*9,728*	*14,612*
	Long	*10,341*	*34,508*	*9,719*	*31,806*	*9,881*	*29,502*
Long	Short	*9,438*	*10,724*	*9,074*	*9,281*	*9,601*	*(44)*
	Medium	*10,037*	*21,956*	*9,485*	*23,422*	*9,890*	*14,167*
	Long	*10,643*	*34,511*	*9,944*	*32,462*	*10,150*	*29,274*

WHAT TO REMEMBER

- In scheduling freight in synchromodal transport using ADP, ***VPI exploration significantly improves the policy, and learned values, of traditional ADP designs.***

- To apply VPI in a finite-horizon ADP with basis functions, ***exploring and updating should be slightly more conservative*** than in conventional infinite-horizon VPI.

- For larger networks, further research in the ***reduction of the decision space*** is necessary for ADP to achieve the largest gains over competing policies in synchromodal transport.

UNIVERSITY OF TWENTE.

DINALOG
Dutch Institute
for Advanced Logistics

THANKS FOR YOUR ATTENTION!

ARTURO E. PÉREZ RIVERA

PhD Candidate

Department of Industrial Engineering and Business Information Systems

University of Twente, The Netherlands

https://www.utwente.nl/bms/iebis/staff/perezrivera/

a.e.perezrivera@utwente.nl

Beta Symposium 2017 - Tuesday, September 26th
Eindhoven, The Netherlands