哈爾濱工業大學(深圳)
HARBIN INSTITUTE OF TECHNOLOGY, SHENZHEN

Agency for Science, Technology and Research
SINGAPORE

Sampling to Achieve the Goal: An Age-Aware Remote Markov Decision Process

Authors: Aimin Li[1,2], Shaohua Wu[1,3], Gary Lee[2], Xiaomeng Chen[1], and Sumei Sun[2]

1 Harbin Institute of Technology (Shenzhen), China
2 Institute for Inforcomm Research, Agency for Science, Technology, and Research, Singapore
3 The Department of Boradband Communication, Peng Cheng Laboratory, China

Presentor : Aimin Li
Date: 2024/11/25

IEEE Information Theory Society

ITW 2024
SHEN ZHEN

Outline

Scan me to access the PDF!

Scan me to access the Code!

1920 HIT

Agency for Science, Technology and Research
SINGAPORE

Agency for Science, Technology and Research
SINGAPORE

I Introduction

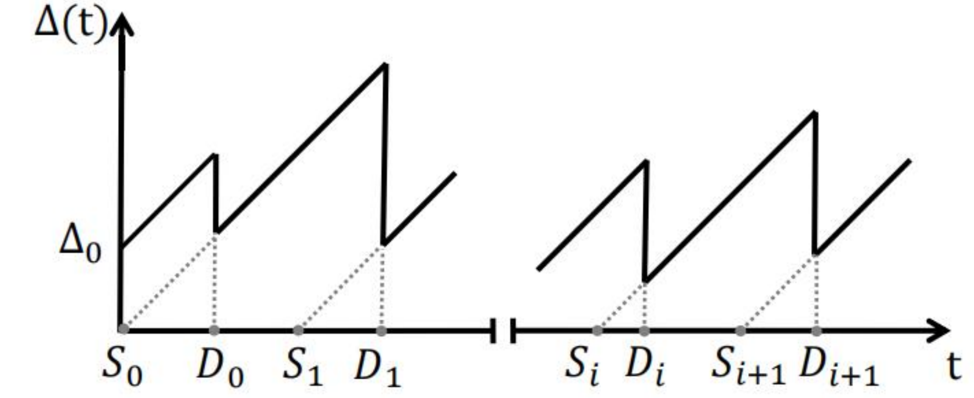
Scan me to access the PDF!

Scan me to access the Code!

Introduction

- Age of Information has becom an **important indicator** in 6G xURLLC.

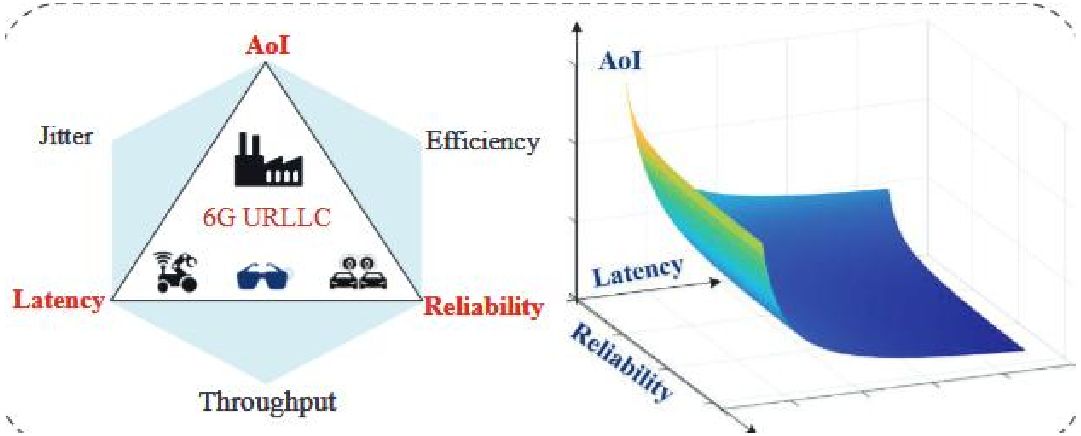
Age of Information (2011,Sanjit Kaul)



Fresher Information, Better Decision Making

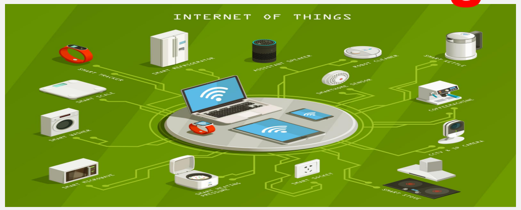
AoI
- Remote Monitoring
- Remote Inference
- Remote Control

Support

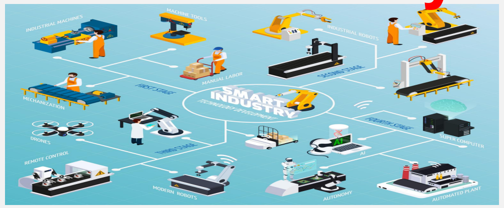


Applications

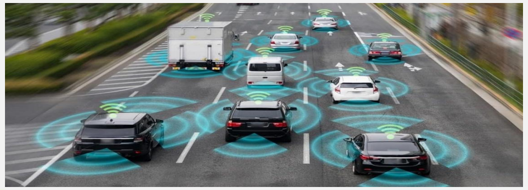
Internet of Things



Smart Industry 4.0



Internet of Vehicle

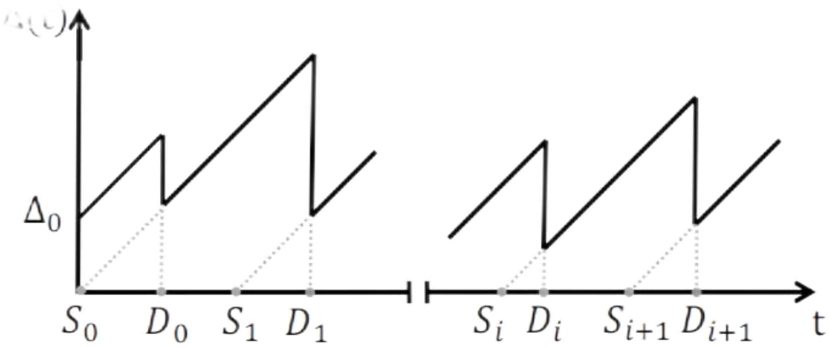


Smart City

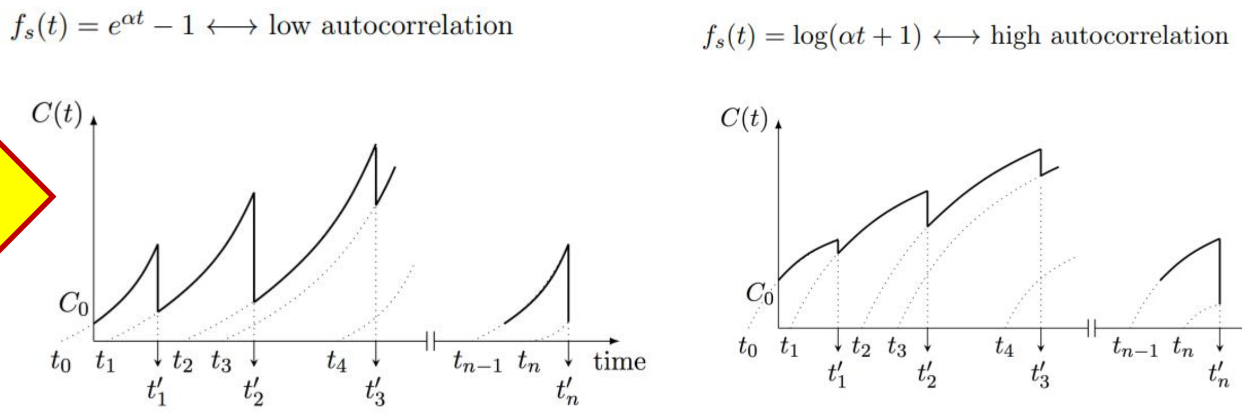
Introduction

- Relationship between Age and Value of Information

Age of Information: (2011,Sanjit Kaul)



Value of Information: Non-linear Age



Fresher Information, Better Decision Making

AoI
- Remote Monitoring
- Remote Inference
- Remote Control

Variants of AoI

- AoII
- UoI
- MSE
- AoCI

Open Questions

For **Goal-oriented Network**, how can we measure the value of information towards achieving a specific goal?

Introduction

- Relationship between Age and Value of Information

Open Questions

For **Goal-oriented Network**, how can we measure the value of information towards achieving a specific goal?

Idea 1

To achieve a goal, we need to **make decisions and take actions** based on potentially **stale ovservations** at the receiver.

Idea 2

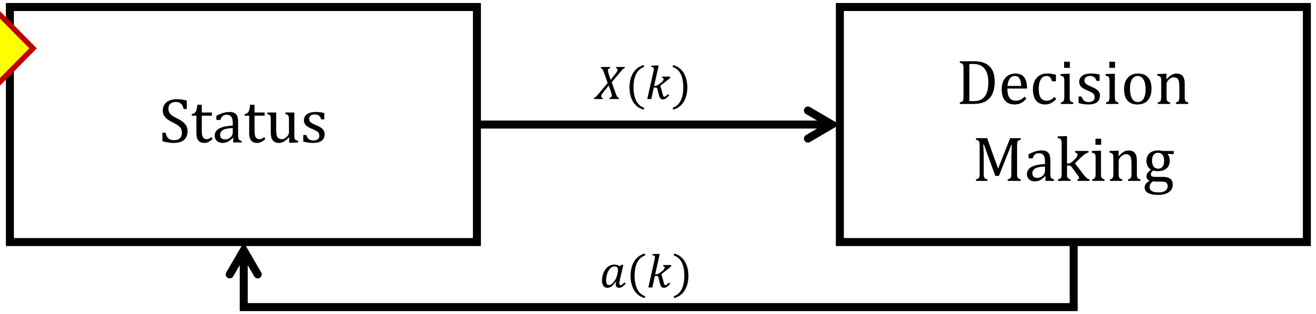
Can we directly evaluate the **utility of the decision making process**?

Fresher Information, Better Decision Making

AoI
- Remote Monitoring
- Remote Inference
- Remote Control

Closed-Loop Control/Decision Making

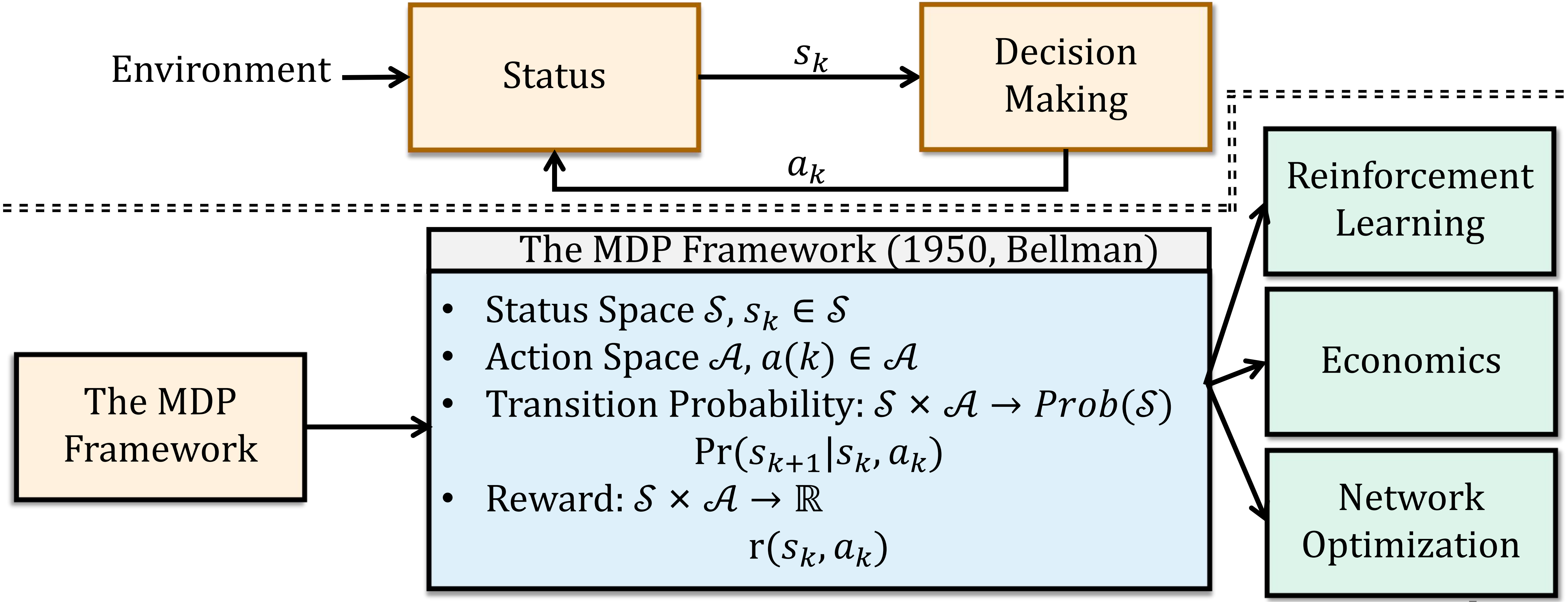
Without Delay



Introduction

- Standard **Markov Decision Process (MDP)** **Without Delay** (1950, Bellman)

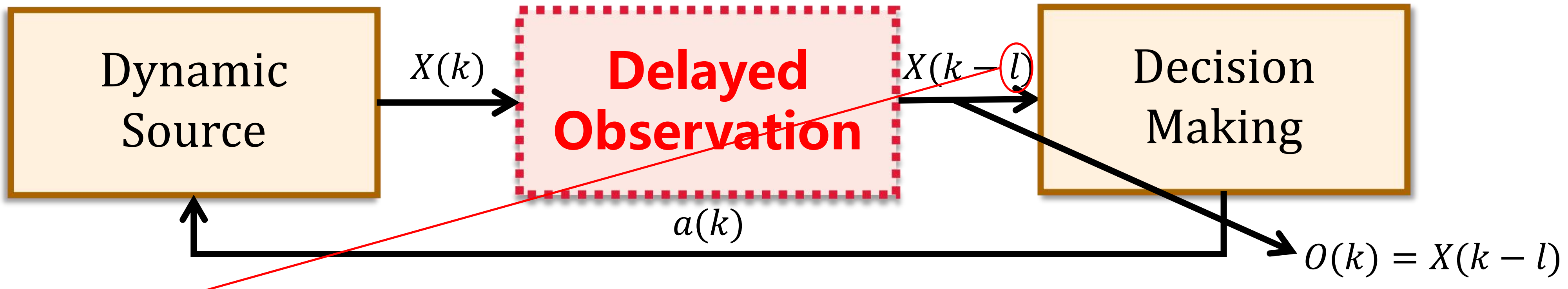
Decision Making Without Delay



Introduction

- Standard MDP With **Constant Observation Delay** (1992 E. Altman and P. Nain)

Closed-Loop Control with **Observation Delay**



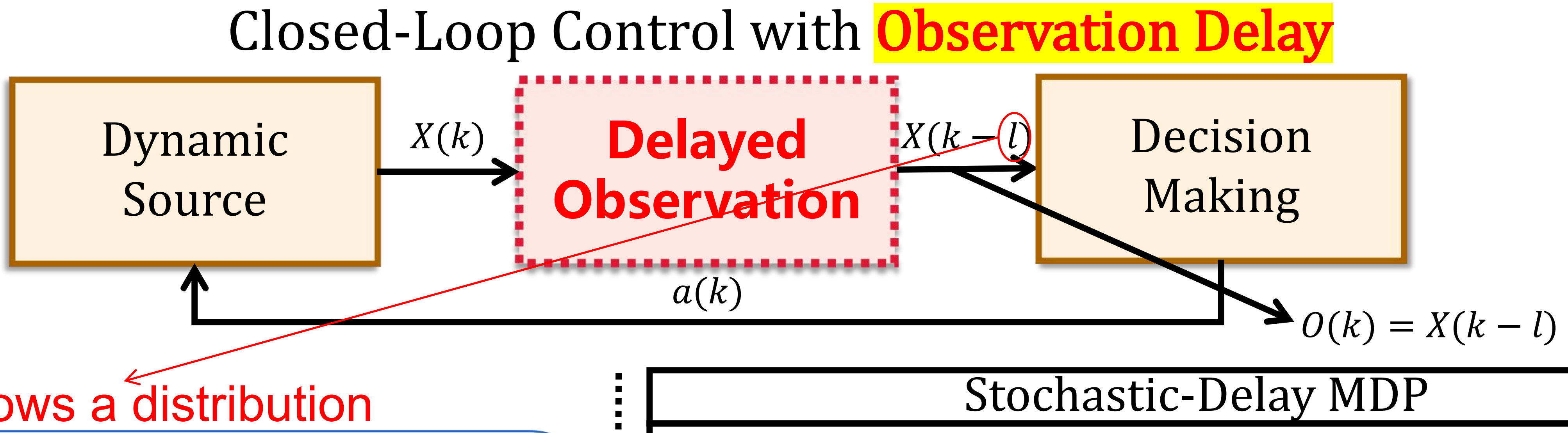
l is a constant

- Constant Delay
  - Through **status augmentation**, a constant-delay problem can be transformed into a **standard MDP without delay**.(1992 E. Altman and P. Nain)

Constant-Delay MDP
• Status $I(k) = (X(k-l), a(k-l), \ldots, a(k-1))$ • Action Space $\mathcal{A}$, $a(k) \in \mathcal{A}$ • Transition Probability: $\mathcal{I} \times \mathcal{A} \rightarrow Prob(\mathcal{I})$ $\Pr(I_{k+1} \mid I_k, a_k)$ • Reward: $\mathcal{I} \times \mathcal{A} \rightarrow \mathbb{R}$ $r'(I_k, a_k) = \mathbb{E}_{X(k)}[r(X(k), a(k)) \mid I_k]$

Introduction

- Standard MDP With **Stochastic Observation Delay** (2003 K. V. Katsikopoulos)



- Stochastic Delay
  - Through status augmentation, a stochestic-delay problem can also be transformed into a standard MDP. (2003 K. V. Katsikopoulos and S. E. Engelbrecht)

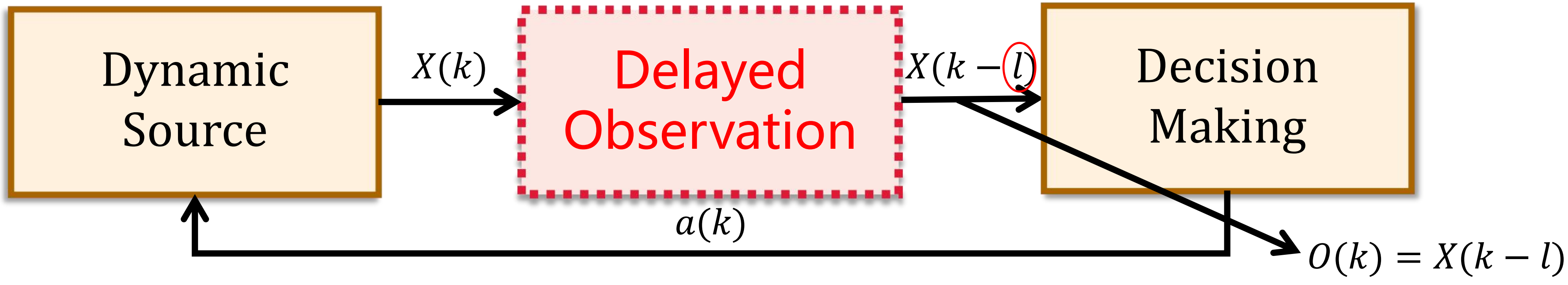
Stochastic-Delay MDP

- Status
  $I(k) = (X(k-l), k-l, a(k-l), \ldots, a(k-1))$
- Action Space $\mathcal{A}$, $a(k) \in \mathcal{A}$
- Transition Probability: $\mathcal{I} \times \mathcal{A} \rightarrow Prob(\mathcal{I})$
  $\Pr(I_{k+1} | I_k, a_k)$
- Reward: $\mathcal{I} \times \mathcal{A} \rightarrow \mathbb{R}$
  $r'(I_k, a_k) = \mathbb{E}_{X(k)}[r(X(k), a(k)) | I_k]$

Introduction

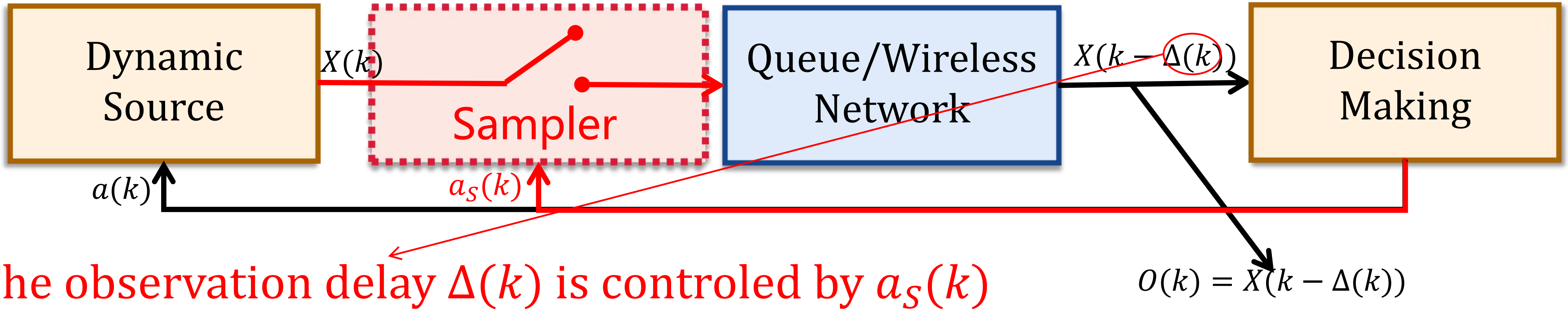
Previous Work

Closed-Loop Control with **Observation Delay**



Proposed Model

AoI-aware Closed-Loop Control and Sampling



The observation delay $\Delta(k)$ is controled by $a_S(k)$

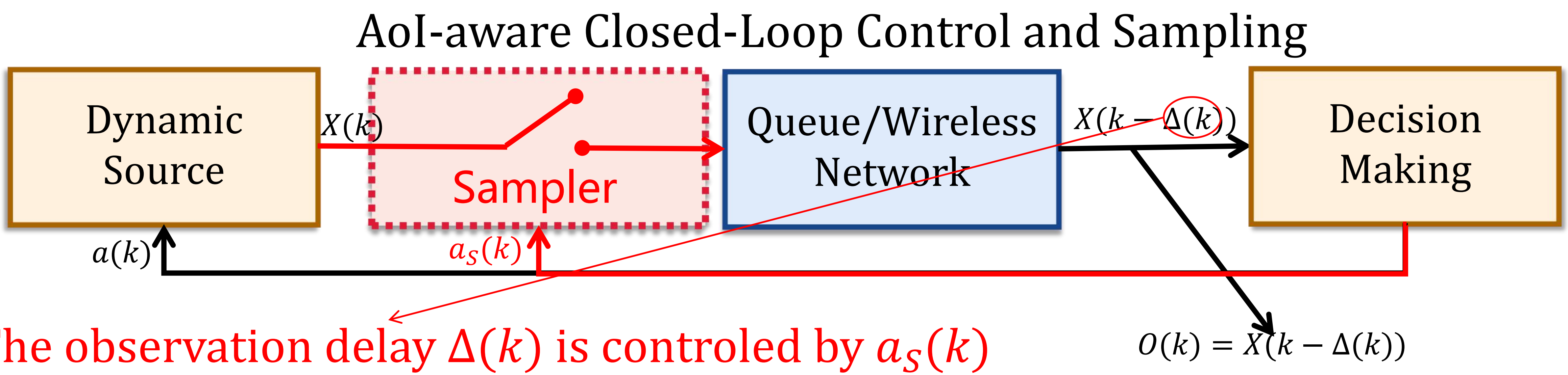
Agency for Science, Technology and Research
SINGAPORE

II System Model

Scan me to access the PDF!

Scan me to access the Code!

System Model



The observation delay $\Delta(k)$ is controled by $a_S(k)$

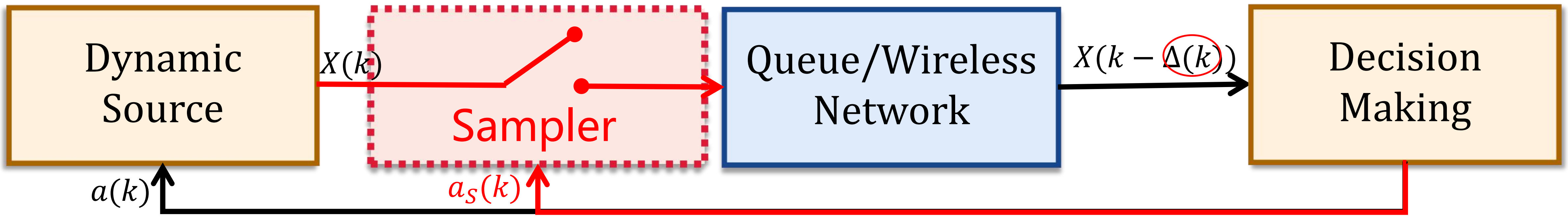
Goal-oriented Sampling Framework:

- ☐ How to make decision at the receiver side based on **stale observation** O(k)?
- ☐ How does the stale information affect this decision making process?

Our Purpose:

- ☐ When to Sample?
- ☐ What to Sample? (**Semantics-aware**)
- ☐ How to control?(**Goal-oriented**)

System Model



AoI-aware Remote MDP

- Status Space $\mathcal{S}$, $X(k) \in \mathcal{S}$
- Action Space $\mathcal{A}$, $a(k) \in \mathcal{A}$
- Dynamics of Source: $\mathcal{S} \times \mathcal{A} \rightarrow Prob(\mathcal{S})$

$$\Pr(X(k+1)|X(k), a(k))$$

- Dynamics of Age: s

$$\Delta(k) = k - S_i, \forall D_i \leq t < D_{i+1}$$
$$S_i = max\{t|t \leq D_i, a_S(t) = 1\}$$

- Random and i.i.d Delay Y_i: Y_i follows a given discerete distribution

$$D_i = S_i + Y_i, \forall i \in \mathbb{N}^+$$

- Reward: $\mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$

$$\mathrm{r}(X(k), a(k))$$

Decision Making Policy

- Available Information for decision making at time slot k, defined as $\mathcal{I}_k$:

$$(X(t - \Delta(t)), \Delta(t), a(t-1), a_S(t-1))_{t \leq k}$$

- Decision making Policy at time slot k:

$$\pi_k: \mathcal{I}_k \rightarrow (a(k), a_S(k))$$

Objective/Goal

$$\max_{\pi_{1:\infty}} \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}\left[\sum_{k=1}^{T} r(X(k), a(k))\right]$$

Agency for Science, Technology and Research
SINGAPORE

III Solution

Scan me to access the PDF!

Scan me to access the Code!

Solution Part I: Sufficient Statistics

Objective/Goal

$$\max_{\pi_{1:\infty}} \lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left[\sum_{k=1}^{T} r(X(k), a(k))\right]$$

$$(a(k), a_S(k)) = \pi_k(\mathcal{I}_k)$$

$$\mathcal{I}_k = (X(t-\Delta(t)), \Delta(t), a(t-1), a_S(t-1))_{t\le k}$$

$$\max_{\pi_{1:\infty}} \lim_{D_n\to\infty} \frac{\mathbb{E}\left[\sum_{k=1}^{D_n} r(X(k), a(k))\right]}{\mathbb{E}[D_n]}$$

$$\max_{\Phi_{1:\infty}} \lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} \mathbb{E}\left[\sum_{k=D_i}^{D_{i+1}-1} r(X(k), a_i)\right]}{\sum_{i=0}^{n-1} \mathbb{E}[D_{i+1} - D_i]}$$

$$(a_i, S_{i+1}) = \Phi_i(\mathcal{G}(i))$$

Challange 1

Curse of History Dimensionality

$$\mathcal{I}_k \in \mathcal{S}^k \times \Lambda^k \times \mathcal{A}^k \times \{0,1\}^k$$

Reformulated Problem

Lemma 1. At time slot D_i, the smallest sufficient statics of $\mathcal{I}_{D_i}$ is

$$\mathcal{G}(i) = (X(D_i - \Delta(D_i)), \Delta(D_i), a(D_i - 1)) = (X(S_i)), Y_i, a_{i-1})$$

$$\mathcal{G}(i) \in \mathcal{S} \times \mathcal{Y} \times \mathcal{A}$$

Solution Part II: Stationry Determinstic Policy

Objective/Goal

$$\max_{\Phi_{1:\infty}} \lim_{n\to\infty} \frac{\sum_{i=0}^{n} \mathbb{E}\left[\sum_{k=\mathrm{D_i}}^{\mathrm{D_{i+1}}-1} r(X(k), a_i)\right]}{\sum_{i=0}^{n} \mathbb{E}[\mathrm{D_{i+1}} - \mathrm{D_i}]}$$

Objective/Goal

$$\max_{\Phi} \lim_{n\to\infty} \frac{\sum_{i=0}^{n} \mathbb{E}\left[\sum_{k=\mathrm{D_i}}^{\mathrm{D_{i+1}}-1} r(X(k), a_i)\right]}{\sum_{i=0}^{n} \mathbb{E}[\mathrm{D_{i+1}} - \mathrm{D_i}]}$$

$$(a_i, \mathrm{S_{i+1}}) = \Phi(\mathcal{G}(\mathrm{i}))$$

Challange 2

Curse of Φ_i $as\ i \in [1, \infty)$

Lemma 2. If $X(k)$ is a unichain, then an optimal stationary policy Φ^* exists in the remote-MDP problem.

Solution Part III: Stationry Determinstic Policy

Objective/Goal

$$h^* = \max_{\Phi} \limsup_{n\to\infty} \frac{\sum_{i=0}^{n} \mathbb{E}\left[\sum_{k=D_i}^{D_{i+1}-1} r(X(k), a_i)\right]}{\sum_{i=0}^{n} \mathbb{E}[D_{i+1} - D_i]}$$

Nonlinear Fractional Programming[1]

Equivalent Standard MDP Problem

$$R^*(\lambda) = \max_{\Phi} \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} \left\{ \mathbb{E}\left[\sum_{k=D_i}^{D_{i+1}-1} r(X(k), a_i)\right] - \lambda \mathbb{E}[D_{i+1} - D_i] \right\}, \lambda \geq 0$$

Lemma 3. If $R^*(\lambda)=0$, then $h^* = \lambda$.
If $R^*(\lambda) > 0$, then $h^* > \lambda$.
If $R^*(\lambda) < 0$, then $h^* < \lambda$.

Solve by Relative Value Iteration

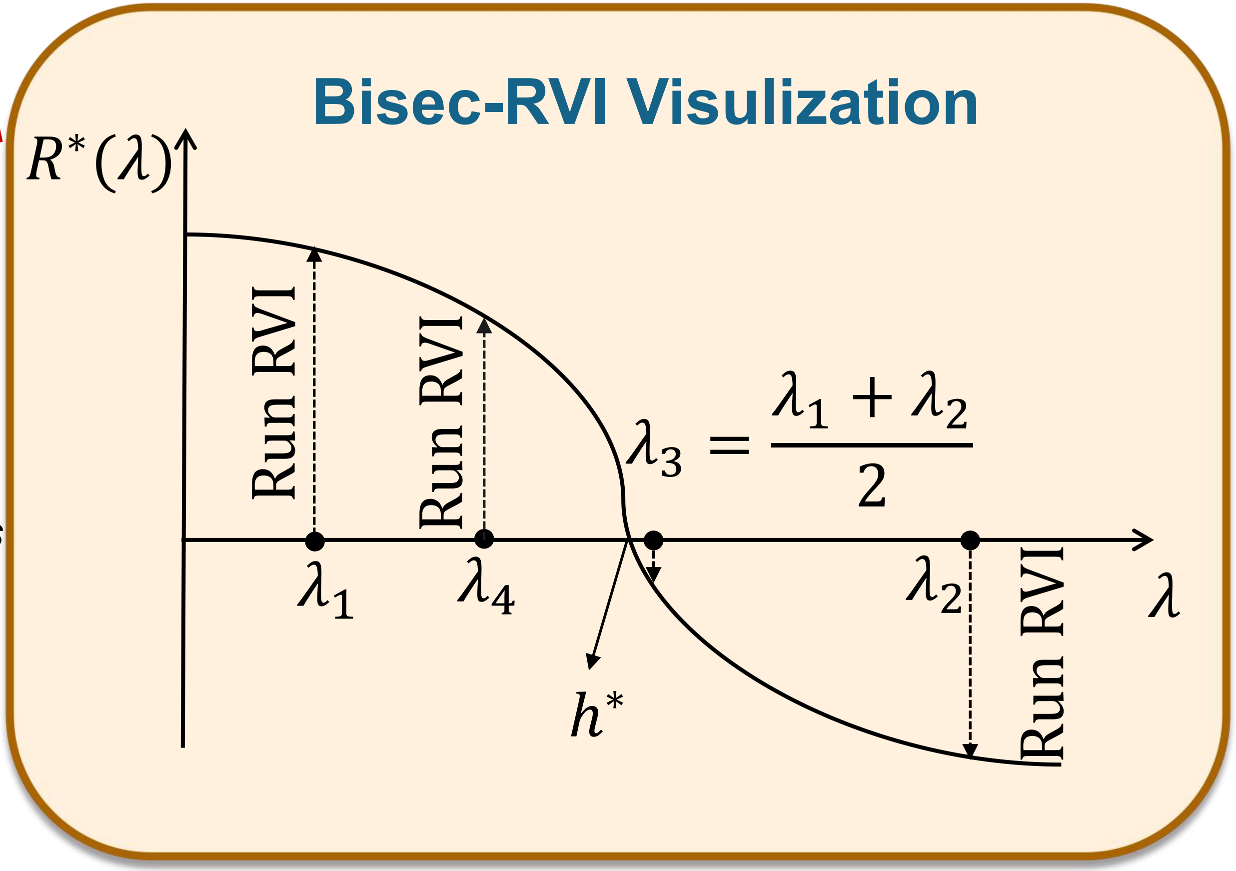
[1] W. Dinkelbach, "On nonlinear fractional programming," Management Science, vol. 13, no. 7, pp. 492–498, 1967.

Solution Part IV: Tow Layer Bisec-RVI

Find h^* such that $R^*(h^*) = 0$

Root Finding of an Un-explicit Function

Bisec-RVI Algorithm [1-4]
Step 1: Run RVI to solve $U(\lambda)$ Step 2: Update λ to find the root Step 3: Repeat Step 1



[1] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," IEEE Trans. Inf. Theory, 2017.

[2] Ahmed M Bedewy, Y. Sun, Sastry Kompella, Ness B Shroff "Optimal sampling and scheduling for timely status updates in multi-source networks," , IEEE Trans. Inf. Theory, 2021.

[3] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the Wiener process for remote estimation over a channel with random delay," IEEE Trans. Inf. Theory, 2019.

[4] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non_x0002_linear age functions," J. Commun. Netw., 2019.

Solution Part V: Fixed Point Iteration

Motivation

Bisec-RVI Algorithm [1-4]

Step 1: Run RVI to solve $U(\lambda)$

Step 2: Update λ to find the root

Step 3: Repeat Step 1

Disadvantage

Run **time-consuming RVI** for **multiple times**!

Improved One-Layer Iteration (Proposed)

Rewrite Bellman

Theorem 2. *Solving Problem $\mathscr{P}_{\text{MDP}}(h^*)$ with $U(h^*) = 0$ is equivalent to solving the following nonlinear equations:*

$$\begin{cases} W^*(\gamma) = \min\limits_{A_i, Z_i} \left\{ g(\gamma, A_i, Z_i; h^*) + \mathbb{E}[W^*(\gamma') \,|\, \gamma, Z_i, A_i] \right\} \\ \text{for } \gamma \in \mathcal{S} \times \mathcal{Y} \times \mathcal{A}, \\ h^* = \min\limits_{A_i, Z_i} \left\{ \dfrac{q(\gamma^{\text{ref}}, A_i, Z_i) + \mathbb{E}[W^*(\gamma') \,|\, \gamma^{\text{ref}}, A_i, Z_i]}{f(Z_i)} \right\}, \end{cases} \tag{17}$$

where $\gamma^{\text{ref}} \in \mathcal{S} \times \mathcal{Y} \times \mathcal{A}$ can be arbitrarily chosen.

Proof. See Appendix G. □

One-Layer Fixed-Point Iterations (FPBI)

The solution to (17) is the fixed point!

$$\boldsymbol{Q}_{k+1} = f(\boldsymbol{Q}_k)$$

Agency for Science, Technology and Research

SINGAPORE

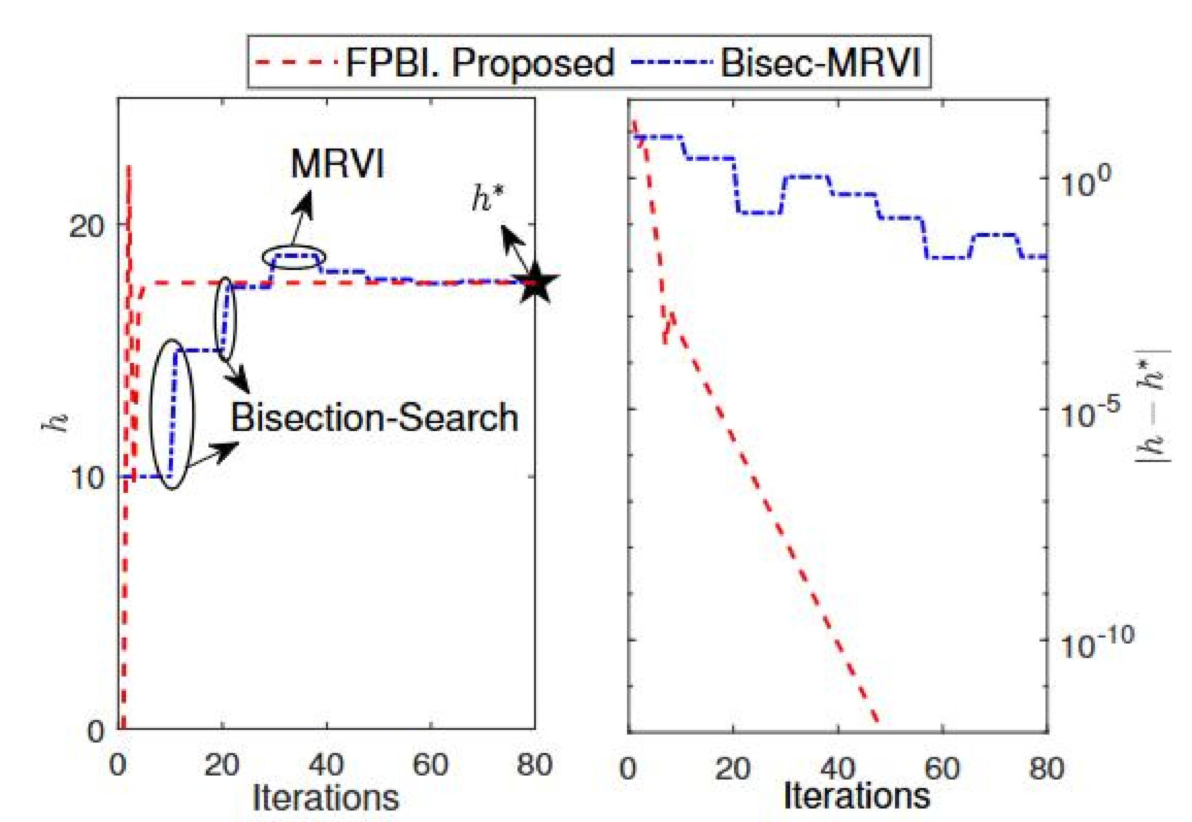
IV Simulation Results

Scan me to access the PDF!

Scan me to access the Code!

Simulation Results

Bisec-RVI vs. Proposed FPBI

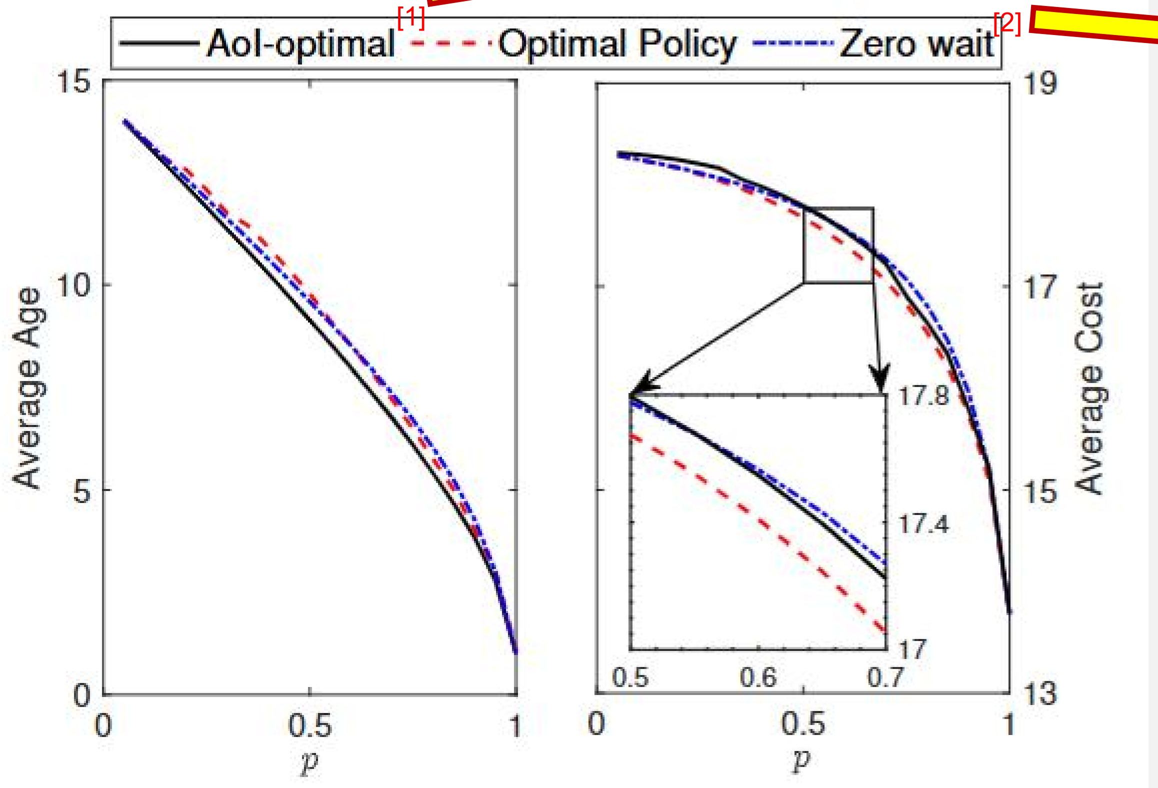


Insight

- The proposed one-layer FPBI algorithm **converges faster** than the Bisec-RVI.
- The improvement in the convergence rate is achieved by **eliminating the need to update λ in the outer layer**.

Simulation Results

Sampling Policies Comparsion



Benchmark 1: AoI-optimal

Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," IEEE Trans. Inf. Theory, 2017.

Benchmark 2: Zero-waiting

Sampling is activated upon the successful delivery of a new packet at the receiver. This policy can maximize the throughput.

Insights

- ◆ Age-optimal policy does not necessarily result in optimal decision making.
- ◆ Sampling-Decision Making Co-design achieves the best utility of decision-making performance.

Agency for Science, Technology and Research
SINGAPORE

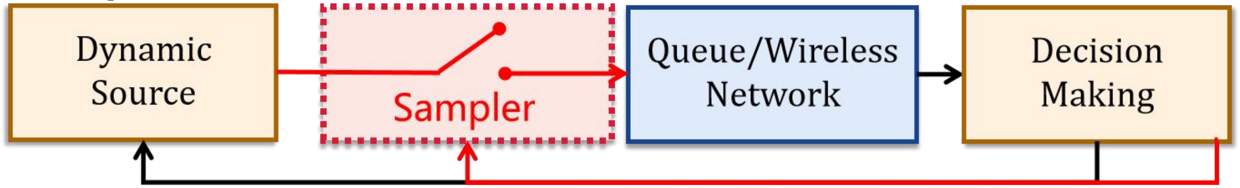
V Conclusion

Scan me to access the PDF!

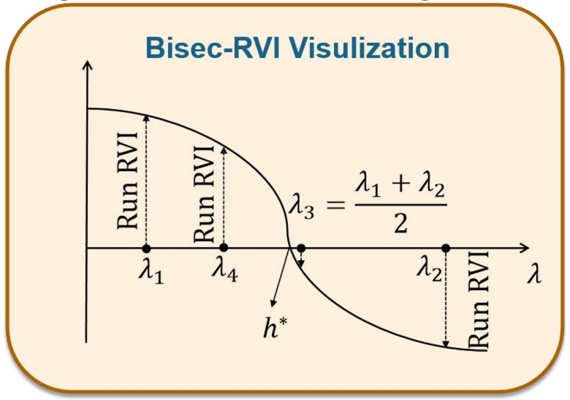
Scan me to access the Code!

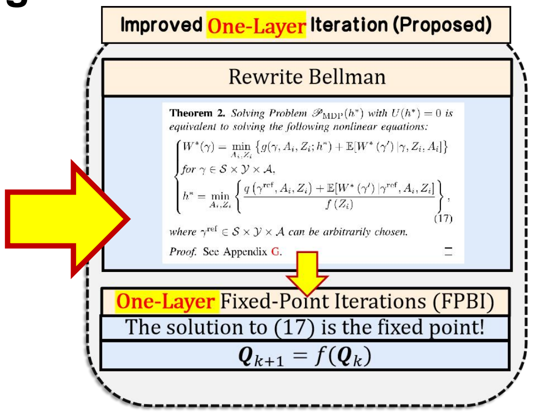
Conclusion

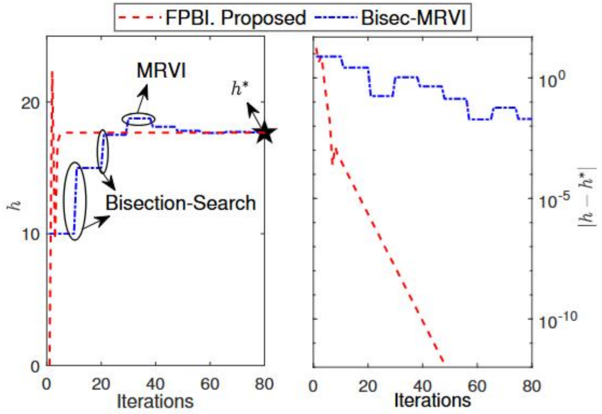
1. We propose a new model named age-aware remote MDP. Only the most relevent packet towards decision-making is sampled and transmitted.



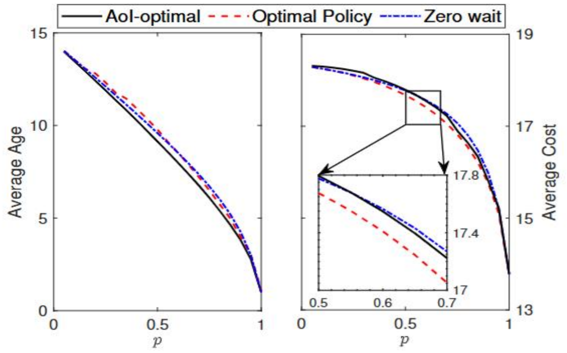
2. We develop Bisec-RVI to solve this problem. To address the complexity of the Bisec-RVI, we further develop a one-layer low-complexity FPBI algorithm.







3. We show that Age-optimal policy does not necessarily result in optimal decision making. Instead, we can optimize the sampling and the decision process simutaneously to achieve the goal.



Agency for Science, Technology and Research
SINGAPORE

Thanks for Listening!

Presentor ： Aimin Li
E-mail: liaimin@stu.hit.edu.cn

Scan me to access the PDF!

Scan me to access the Code!