Semantic-Aware Remote Estimation of Multiple Markov Sources Under Constraints

Jiping Luo and Nikolaos Pappas
Department of Computer and Information Science, Linköping University, Sweden
Email: {jiping.luo, nikolaos.pappas}@liu.se

***Abstract*—We study the semantic-aware remote estimation of multiple Markov sources over a lossy and rate-constrained channel, where the remote actuator has different tolerances for estimation errors of different states. We show the existence and structure of the optimal policy and introduce an efficient policy search algorithm. Numerical results show that continuous transmission is inefficient, and remarkably, our semantic-aware policy attains the optimum by strategically utilizing fewer transmissions by exploiting the timing of important information.**

I. Introduction

Remote estimation of stochastic processes is a fundamental problem in networked control systems (NCSs) [1], [2]. However, such applications confront critical scalability challenges due to the scarcity of communication resources and the immensity of sensory data [3]. Therefore, there is a need to reduce the amount of generated data communicated between sensors, plants, and actuators. This is often done through various metrics that capture the *value* (semantics) of information and determine whether a new data packet has a positive impact on the control system [4], [5]. Despite significant research efforts, most existing communication protocols are context-agnostic, leading to inefficient communications and degraded performance [5], [6].

The primary objective has been to minimize the estimation error [1], [7]–[10], indicating that information is valuable when it is *accurate*. However, in reality, this is not always the case, as *the actuator may have different tolerance for estimation errors of different states or in different situations*. The Age of Information (AoI), that is, the time elapsed since the last successful reception [11], has recently been employed in NCSs [12]–[15]. It implies that status updates are more valuable when it is *fresher*. However, AoI ignores the source evolution and is thus inefficient in tracking Markovian sources.

Several metrics have been introduced to address the shortcomings of AoI [16]–[21]. Age of Incorrect Information (AoII) [16] captures the cost of not having a correct estimate at the destination for some time. The authors in [18], [19] defined state-aware AoI variables to account for the significance of the source states. However, these metrics treat different estimation errors equally. A new metric, namely cost of actuation error (CAE), was first introduced in [2] to account for the significance of different estimation errors. Specifically, the CAE prioritizes information flow according to the potential control risks or costs incurred without correctly estimating these states. The problem of remote estimation of a single discrete-state Markov source was further studied in [22]–[25]. However, these works only consider a single-source scenario, and theoretical results regarding the existence and structure of the constrained optimal policy have not been developed.

This paper considers a more general case where an agent schedules the update time of multiple Markov sources over a lossy and rate-constrained channel. The system consists of slowly and rapidly evolving sources. Unlike the single-source scenario, where the sampling decision is solely based on state importance, the multi-source system has a two-level significance: state importance and source significance, which is relevant in multi-modal scenarios. We formulate the problem as a constrained Markov Decision Process (CMDP) and show the existence and structure of the constrained optimal policy. We also propose an efficient policy search algorithm, termed *intersection search*, to compute the optimal policy. Numerical results highlight the effectiveness of exploiting data significance in such systems.

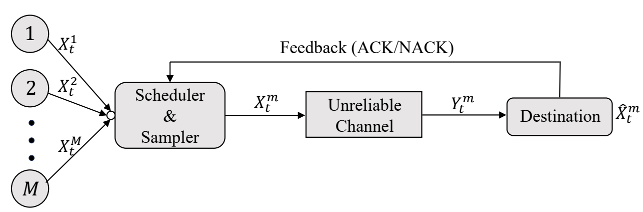


Fig. 1. Remote estimation of multiple Markov sources with feedback.

II. System Model and Problem Formulation

A. System Description

We consider a remote estimation system shown in Fig. 1, where the destination remotely tracks the states of multiple stochastic processes for achieving a specific control task. Denote $\mathcal{M} = \{1, 2, \ldots, M\}$ as the index set of the sources. Each source $m \in \mathcal{M}$ is modeled by a N_m-state discrete-time Markov chain (DTMC), $\{X_t^m\}_{t\geq 1}$. The value of X_t^m

This work has been supported in part by the Swedish Research Council (VR), ELLIIT, the European Union (ETHER) under Grant 101096526; the European Union's Horizon Europe Research and Innovation Programme under the Marie Skłodowska-Curie Grant under Agreement 101131481 (SOVEREIGN); by the Horizon Europe/JU SNS Project ROBUST-6G under Grant 101139068, and the Graduate School in Computer Science (CUGS).

ISBN 978-3-903176-65-2 © 2024 IFIP

is chosen from a finite set $\mathcal{X}^m = \{1, 2, \ldots, N_m\}$. The state transition probabilities of each source m are represented by a stochastic matrix $\mathbf{Q}^m$, where $\mathbf{Q}^m_{i,j} = \Pr[X^m_{t+1} = j | X^m_t = i]$ is the transition probability from state i to state j between two consecutive time slots. A DTMC is said to be *slowly evolving* if it tends to stay in the same state across two consecutive time slots, i.e., $\mathbf{Q}$ is diagonally dominant. Otherwise, we call it *rapidly evolving*. We consider that the system consists of slowly and rapidly evolving sources.

An agent observes the sources' states and decides at the beginning of each time slot which source to sample. After generating a status update of some source, the transmitter sends the packet to the destination over a wireless channel. We assume that only one packet can be sent at a time. We denote the sampling decision at time t by A_t, where $A_t = 0$ means no source is selected and the transmitter remains salient, while $A_t = m, m \neq 0$ means taking a sample from source m and transmitting it immediately. The transmitter can only transmit packets intermittently due to resource constraints and network contention. So, we impose a constraint on the transmission frequency that is no larger than $F_{\text{max}} \in (0, 1]$.

We consider an unreliable channel with *i.i.d.* packet-drops. The channel state H_t is 1 if the packet is successfully decoded at the receiver and 0 otherwise. We define the success transmission probability as $p_s = \mathbb{E}[H_t]$. The receiver sends an acknowledgment (ACK)/NACK signal to the transmitter after each transmission to indicate the transmission success or failure. The receiver reconstructs the states of the sources using the latest received packet, i.e.,

$$\hat{X}^m_t = \begin{cases} X^m_t, & \text{if } A_t = m, H_t = 1 \\ \hat{X}^m_{t-1}, & \text{otherwise.} \end{cases} \tag{1}$$

We denote the system state (i.e., the latest information at the agent) at the beginning of each time slot t by $S_t = (X_t, \hat{X}_{t-1})$. Similarily, the state of subsystem m at time t is $S^m_t = (X^m_t, \hat{X}^m_{t-1})$. The discrepancy between the source and reconstructed states can cause actuation errors. In practice, some source states are more important than others; thus, some actuation errors may have a larger impact than others. Let $\delta^m : \mathcal{X}^m \times \mathcal{X}^m \to \mathbb{R}_{\geq 0}$ be a non-commutative function capturing *state-dependent significance* of estimation errors [2]. Then, the *cost of actuation error* (CAE) of taking an action in a certain system state is defined as the weighted sum of all subsystems' control costs

$$c(S_t, A_t, S_{t+1}) = \sum_{m \in \mathcal{M}} \omega_m \delta^m(X^m_t, \hat{X}^m_t), \tag{2}$$

where $\omega_m \in \mathbb{R}^+$ represents the relative *significance* of source m to the overall system.

Remark 1. *Although the cost function c considered here is decomposable, the approach proposed in this paper can deal with coupled collective costs. Moreover, the cost function depends on the subsequent system state, so the agent can only evaluate the effectiveness of sampling actions after receiving feedback packets. It implies that the agent should know a priori the expected CAE of taking an action in a particular state.*

Remark 2. *The system has two levels of significance: the source significance and the state importance. The significance of information depends on several factors, such as the weight (ω_m), the channel condition (p_s), the source evolution pattern ($\mathbf{Q}^m$), and the resource constraint (F_{max}).*

B. Performance Measures

At each time t, the system incurs two costs: an actuation cost defined in (2), and a transmission cost given by

$$f(S_t, A_t) = \mathbb{1}(A_t \neq 0). \tag{3}$$

We define the average costs of a policy $\pi = (A_1, A_2, \ldots)$ by

$$C_s(\pi) \triangleq \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}^\pi \left[c(S_t, A_t, S_{t+1}) \middle| S_1 = s \right], \tag{4}$$

$$F_s(\pi) \triangleq \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}^\pi \left[f(S_t, A_t) \middle| S_1 = s \right], \tag{5}$$

where $\mathbb{E}^\pi$ represents the conditional expectation, given that policy π is employed with an initial state $s \in \mathcal{S}$.

C. System Dynamics

We denote $P : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ the system transition function, where $P(s'|s, a) = \Pr[S_{t+1} = s' | S_t = s, A_t = a]$ is the probability that the system state becomes s' at time $t + 1$ when the agent chooses action a in state s at time t. For a given sampling decision, each subsystem $m \in \mathcal{M}$ evolves independently. Thus, we have

$$P(s'|s, a) = \prod_{m \in \mathcal{M}} P^m(s'(m)|s(m), a), \tag{6}$$

where P^m is the transition function of subsystem m, given by

$$P^m(s'(m)|s(m), a) = \begin{cases} \mathbf{Q}^m_{i,k} p_s, & \text{if } a = m, s(m) = (i, j), s'(m) = (k, i) \\ \mathbf{Q}^m_{i,k} p_f, & \text{if } a = m, s(m) = (i, j), s'(m) = (k, j) \\ \mathbf{Q}^m_{i,k}, & \text{if } a = m, s(m) = (i, i), s'(m) = (k, i) \\ \mathbf{Q}^m_{i,k}, & \text{if } a \neq m, s(m) = (i, j), s'(m) = (k, j) \\ \mathbf{Q}^m_{i,k}, & \text{if } a \neq m, s(m) = (i, i), s'(m) = (k, i) \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $p_f = 1 - p_s$, $\forall i, j, k \in \mathcal{X}^m, i \neq j$.

D. Problem Formulation

We aim to find a policy π to minimize the average CAE while satisfying a transmission frequency constraint. We can formulate this as an optimization problem

$$\mathscr{P}_{\text{CMDP}} : \inf_{\pi \in \Pi} C_s(\pi), \text{ s.t. } F_s(\pi) \leq F_{\text{max}}. \tag{8}$$

The above problem is essentially an *average-cost constrained Markov Decision Process* (CMDP), which is, however, challenging to solve since it imposes a *global constraint that involves the entire decision-making process*.

Definition 1. *(Constrained optimal policy) A policy is called constrained optimal if it solves $\mathscr{P}_{CMDP}$ for all $s \in \mathcal{S}$.*

III. STRUCTURE OF THE OPTIMAL POLICY

A. The Lagrangian MDP

Now we derive the structure of constrained optimal policies based on the Lagrangian technique [26]–[28]. We start with rewriting $\mathscr{P}_{\text{CMDP}}$ in its (unconstrained) Lagrangian form. The *Lagrangian average cost* of a policy π is defined as

$$\begin{aligned}\mathcal{L}_s^\lambda(\pi) &\triangleq \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}^\pi\left[l^\lambda(S_t, A_t, S_{t+1})|S_1 = s\right] \\ &= C_s(\pi) + \lambda F_s(\pi), \end{aligned} \tag{9}$$

where $\lambda \geq 0$ is the Lagrangian multiplier, $l^\lambda(s, a, s')$ is the *Lagrangian instantaneous cost* given by

$$l^\lambda(s, a, s') = c(s, a, s') + \lambda f(s, a). \tag{10}$$

Given any $\lambda \geq 0$, define the unconstrained *Lagrangian MDP*

$$\mathscr{P}_{\text{L-MDP}}^\lambda : \inf_{\pi\in\Pi} \mathcal{L}_s^\lambda(\pi). \tag{11}$$

The above problem is a standard MDP with an average cost criterion. The Lagrangian approach thereafter solves the following unconstrained "sup-inf" problem [28]

$$\sup_{\lambda\geq 0} \inf_{\pi\in\Pi} \mathcal{L}_s^\lambda(\pi) - \lambda F_{\max}. \tag{12}$$

The inner problem solves a standard MDP with a fixed multiplier λ, and the outer problem searches for the largest multiplier that produces a feasible policy. Note that in the remaining of this paper, we will refer to $\mathscr{P}_{\text{L-MDP}}^\lambda$ as the inner problem since $\lambda F_{\max}$ is constant.

The next result shows the existence of an optimal policy for the inner problem that is stationary and deterministic.

Definition 2. *(λ-optimal policy) A policy π is called λ-optimal if it solves $\mathscr{P}_{L\text{-}MDP}^\lambda$ and the average costs $\mathcal{L}^\lambda, C^\lambda$, and F^λ are independent of initial states.*

Theorem 1. *(Existence of the λ-optimal policy) For any given $\lambda \geq 0$, there exists a vector $\boldsymbol{h}^\lambda$ that satisfies the following Bellman's optimality equation*

$$\mathcal{L}^\lambda + \boldsymbol{h}^\lambda(s) = \min_{a\in\mathcal{A}} \sum_{s'\in\mathcal{S}} P(s'|s, a)\big(l^\lambda(s, a, s') + \boldsymbol{h}^\lambda(s')\big), \tag{13}$$

for all $s \in \mathcal{S}$. The quantities $\mathcal{L}^\lambda, F^\lambda$, and C^λ are independent of initial states. Furthermore, the λ-optimal is obtained by

$$\pi_\lambda^*(s) = \arg\min_{a\in\mathcal{A}} \sum_{s'\in\mathcal{S}} P(s'|s, a)\big(l^\lambda(s, a, s') + \boldsymbol{h}^\lambda(s')\big). \tag{14}$$

Proof. Please see Theorem 1 in [29]. ☐

Then, we can use the *relative value iteration* (RVI) to solve $\mathscr{P}_{\text{L-MDP}}^\lambda$, as shown in Algorithm 1. Let $s_{\text{ref}} \in \mathcal{S}$ denote an arbitrarily chosen reference state, $\boldsymbol{v}^k$ and $\tilde{\boldsymbol{v}}^k$ denote the value function and the relative value function for iteration k, respectively. In the beginning, we set $\boldsymbol{v}^0 = \tilde{\boldsymbol{v}}^0 = \mathbf{0}$. Then, for each $k \geq 1$, the RVI updates the value functions as follows

Algorithm 1: RVI for solving $\mathscr{P}_{\text{L-MDP}}^\lambda$.

Input: λ, c, f, P, ϵ
Output: $\pi_\lambda^*, F^\lambda, C^\lambda$, $\mathcal{L}^\lambda$

$s_{\text{ref}} = 0, \boldsymbol{v}^0 = \tilde{\boldsymbol{v}}^0 = \mathbf{0}$ for all $s \in \mathcal{S}$
for $k = 1, 2, \ldots$ **do**
  for *each* $s \in \mathcal{S}$ **do**
    UPDATE $\boldsymbol{v}^k, \tilde{\boldsymbol{v}}^k$ using (15)-(16).
  end
  if $\|\boldsymbol{v}^k - \boldsymbol{v}^{k-1}\|_\infty < \epsilon$ **then**
    for *each* $s \in \mathcal{S}$ **do**
      COMPUTE π_λ^* using (14) with $\boldsymbol{h}^\lambda = \tilde{\boldsymbol{v}}^k$.
    end
    COMPUTE average costs using (17)-(19).
    return $\pi_\lambda^*, F^\lambda, C^\lambda$, $\mathcal{L}^\lambda$
  end
end

$$\boldsymbol{v}^{k+1}(s) = \min_{a\in\mathcal{A}} \sum_{s'\in\mathcal{S}} P(s'|s, a)\big(l^\lambda(s, a, s') + \tilde{\boldsymbol{v}}^k(s')\big), \tag{15}$$

$$\tilde{\boldsymbol{v}}^{k+1}(s) = \boldsymbol{v}^{k+1}(s) - \boldsymbol{v}^{k+1}(s_{\text{ref}}). \tag{16}$$

As stated in Proposition 1 below, the sequences $\{\boldsymbol{v}^k\}_{k\geq 0}$ and $\{\tilde{\boldsymbol{v}}^k\}_{k\geq 0}$ converge to $\boldsymbol{v}^*$ and $\boldsymbol{h}^\lambda$, respectively. Then, the λ-optimal policy can be obtained by solving (14), and the average costs can be given by

$$C^\lambda = \sum_{s\in\mathcal{S}} \boldsymbol{\nu}^{\pi_\lambda^*}(s)\boldsymbol{c}^{\pi_\lambda^*}(s), \tag{17}$$

$$F^\lambda = \sum_{s\in\mathcal{S}} \boldsymbol{\nu}^{\pi_\lambda^*}(s)\boldsymbol{f}^{\pi_\lambda^*}(s), \tag{18}$$

$$\mathcal{L}^\lambda = C^\lambda + \lambda F^\lambda, \tag{19}$$

where $\boldsymbol{\nu}^{\pi_\lambda^*}$ is the stationary distribution induced by π_λ^*.

Proposition 1. *(Convergence of RVI) For any $\lambda \geq 0$, the sequences $\{\boldsymbol{v}^k\}_{k\geq 0}$ generated by Eqs.* (15)-(16) *convergences, i.e., $\lim_{k\to\infty} \boldsymbol{v}^k \to \boldsymbol{v}^*$. Moreover, $\mathcal{L}^\lambda = \boldsymbol{v}^*(s_{ref})$ and $\boldsymbol{h}^\lambda = \boldsymbol{v}^* - \boldsymbol{v}^*(s_{ref})$ is a solution to* (13).

Proof. Please see Proposition 2 in [29]. ☐

B. Structure of the Constrained Optimal Policy

In what follows we show that there exists a constrained optimal policy that is in no case more complex than a randomized mixture of two Markovian deterministic policies. Before stating that, we first present some important properties of the cost functions $\mathcal{L}^\lambda$, C^λ, and F^λ. These properties provide the basis for solving the outer problem in (12).

Proposition 2. *(Properties of the optimal cost functions) $\mathcal{L}^\lambda$ is a piecewise linear, concave, and monotonically increasing function of λ. $F^\lambda(C^\lambda)$ is piece-wise constant, monotonically non-increasing (non-decreasing).*

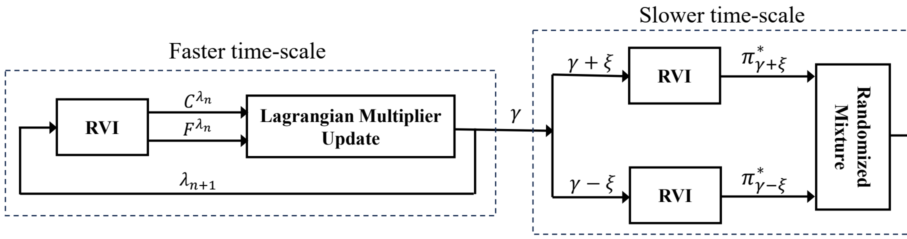


Fig. 2. Flow diagram of finding the constrained optimal policy of CMDPs. The RVI algorithm is applied to solve Lagrangian MDPs with a fixed λ. The Lagrangian multiplier is updated until the γ value is found. Then, the constrained optimal policy is constructed as a mixture of two deterministic policies.

Proof. Please see Proposition 3 in [29]. □

Definition 3. *(γ value) By the monotonicity of F^λ and $\mathcal{L}^\lambda$, the solution to the outer problem in* (12) *can be defined by*

$$\gamma \triangleq \inf\{\lambda : \lambda \geq 0, F^\lambda \leq F_{max}\}. \tag{20}$$

Proposition 3. *(Feasibility of the Lagrangian problem) For any $F_{max} \in (0,1]$, the unconstrained problem in* (12) *always returns a solution, that is $0 \leq \gamma < \infty$.*

Proof. Please see Proposition 4 in [29]. □

Although the above propositions ensure the existence of an optimal solution to problem (12), it may not be optimal to the original CMDP since the duality gap can be non-zero. In [26, Theorem 4.3] and [30, Page 5] are provided the sufficient conditions under which an optimal policy π^* for a Lagrangian MDP is also optimal for the original constrained problem $\mathscr{P}_{\text{CMDP}}$, i.e., zero duality gap. Specifically, the policy π^* must solve $\mathscr{P}^\lambda_{\text{L-MDP}}$ for some $\lambda > 0$, and in addition, the average costs satisfy $F^\lambda = F_s(\pi^*) = F_{\max}, C^\lambda = C_s(\pi^*)$ for all $s \in \mathcal{S}$. Notice that π^* needs not to be deterministic. We verify that these conditions are valid and establish the structure of the optimal policy in the following theorem.

Theorem 2. *(Structure of the optimal policy) The optimal policy π^* is deterministic if the following condition holds*

$$\gamma(F^\gamma - F_{max}) = 0. \tag{21}$$

Otherwise, the optimal policy is a randomized mixture of two deterministic policies, i.e.,

$$\pi^* = q\pi^*_{\gamma^-} + (1-q)\pi^*_{\gamma^+}, \tag{22}$$

where $q = (F_{max} - F^{\gamma^+})/(F^{\gamma^-} - F^{\gamma^+})$, F^{γ^-} and F^{γ^+} are the left-hand limit and right-hand limit at point γ. Notice that $\pi^*_{\gamma^-}$ is infeasible while $\pi^*_{\gamma^+}$ is feasible.

Proof. Please see Theorem 2 in [29]. □

IV. An Efficient Search Method for CMDP

A. Bisection Search Method

Fig. 2 gives a schematic representation of the flow diagram of the two-timescale solution approach [31], [32]. Specifically, on the faster time scale, finding a λ-optimal policy for any given λ, while on the slower timescale, λ is updated until γ is found. However, finding γ and the corresponding optimal policies is computationally demanding [30].

A commonly used method in the literature is the *Bisection search plus RVI* (Bisec-RVI) method [17], [33]. It leverages the monotonicity of the function F^λ to determine γ, or more precisely, the "root" of the function $F^\lambda - F_{\max}$ in a predetermined interval $[0, \lambda_{\max}]$. Here, $\lambda_{\max}$ is a reasonably large constant chosen such that $F^{\lambda_{\max}} < F_{\max}$. Notice that the "root" needs not to exist because of the discontinuity of F^λ. However, we can bracket γ within an arbitrarily small interval.

At each iteration, the Bisec-RVI method divides the interval into two halves and selects the subinterval that brackets the root. We denote the interval at the n-th iteration as $I_n = [\lambda^n_-, \lambda^n_+]$ and the middle point of I_n as γ^n, where $I_0 = [0, \lambda_{\max}]$ and $\gamma^n = (\lambda^n_- + \lambda^n_+)/2$. The Bisec-RVI uses the RVI to solve the problem $\mathscr{P}^{\gamma^n}_{\text{L-MDP}}$. If the average transmission cost at point γ^n is larger than $F_{\max}$, it implies that the root γ is located within the subinterval $[\gamma^n, \lambda^n_+]$. Therefore, we can update the interval as $I_{n+1} = [\gamma^n, \lambda^n_+]$. On the other hand, if the average transmission cost at point γ^n is smaller than $F_{\max}$, then we have $I_{n+1} = [\lambda^n_-, \gamma^n]$. This process continues until a desired accuracy is achieved, i.e., $|\lambda^n_- - \lambda^n_+| < \xi$. Then, the optimal policy is given by

$$\pi^* = \mu\pi^*_{\lambda^n_-} + (1-\mu)\pi^*_{\lambda^n_+}, \tag{23}$$

where μ is a randomization factor given by

$$\mu = \frac{F_{\max} - F^{\lambda^n_+}}{F^{\lambda^n_-} - F^{\lambda^n_+}}. \tag{24}$$

The time complexity of the bisection search is $\mathcal{O}(\log_2 \frac{\lambda_{\max}}{\xi})$. So it requires a large number of iterations to reach a desired accuracy. Moreover, the algorithm exhibits a slow convergence rate near the breakpoint γ because the derivative of F^λ is zero.

B. Proposed Intersection Search Method

Next, we propose a novel method, called *Intersection search plus RVI* (Insec-RVI), that efficiently solves $\mathscr{P}_{\text{CMDP}}$ by exploiting the concavity and piecewise linearity of $\mathcal{L}^\lambda$. In this method, γ^n is no longer the middle point of I_n but an intersection point of two tangents to the function $\mathcal{L}^\lambda$, as shown in Fig 3. More specifically, one tangent is formed by a point $(\lambda^n_-, \mathcal{L}^{\lambda^n_-})$ with a slop of $F^{\lambda^n_-}$, and another tangent is formed by a point $(\lambda^n_+, \mathcal{L}^{\lambda^n_+})$ with a slop of $F^{\lambda^n_+}$. Then, the intersection point of these two lines, $(\gamma^n, \tilde{\mathcal{L}}^{\gamma^n})$, is given by

$$\gamma^n = \frac{C^{\lambda_+^n} - C^{\lambda_-^n}}{F^{\lambda_-^n} - F^{\lambda_+^n}}, \tag{25}$$

$$\tilde{\mathcal{L}}^{\gamma^n} = F^{\lambda_-^n}(\gamma^n - \lambda_-^n) + \mathcal{L}^{\lambda_-^n}. \tag{26}$$

Analogous to the Bisec-RVI method, if $F^{\gamma^n} \geq F_{\max}$, then $\lambda_-^{n+1} = \gamma^n$, otherwise $\lambda_+^{n+1} = \gamma^n$.

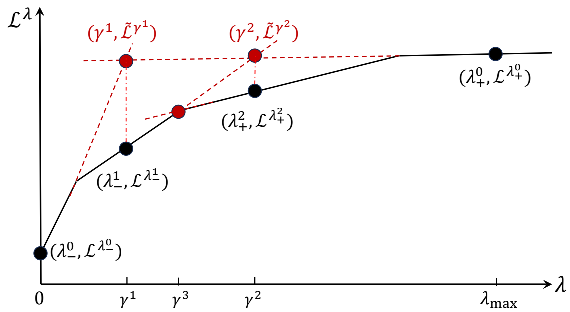


Fig. 3. An illustration of the Insec-RVI method. Consider that function $\mathcal{L}^\lambda$ has 4 segments within the interval $[0, \lambda_{\max}]$ and γ is the second corner of $\mathcal{L}^\lambda$. The red dots represent the intersection points at each iteration, while the black dots are the corresponding projected points onto curve $\mathcal{L}^\lambda$.

Since γ is a corner of function $\mathcal{L}^\lambda$, the algorithm converges when the intersection point is located on $\mathcal{L}^\lambda$, that is $\tilde{\mathcal{L}}^{\gamma^n} = \mathcal{L}^{\gamma^n}$. To compute the optimal policy, we can simply use Eqs. (23)-(24) by setting $\lambda_-^n = \gamma^n - \zeta/4, \lambda_+^n = \gamma^n + \zeta/4$, where ζ is an arbitrarily chosen small perturbation. Recall that the stopping criterion of the Bisec-RVI is $|\lambda_-^n - \lambda_+^n| < \xi$. Therefore, the Insec-RVI can achieve comparable performance to the Bisec-RVI by setting $\zeta = \xi$, without introducing any additional computations. In addition, choosing an appropriate $\lambda_{\max}$ value can be problematic when using Bisec-RVI since it can either be too large (leading to excessive computation time) or too small (rendering it infeasible). Fortunately, in our Insec-RVI algorithm, we can simply initialize the costs as $F^{\lambda_+} \leftarrow 0$ and $C^{\lambda_+} \leftarrow C_{\max}$, where $C_{\max}$ is the average CAE induced by the non-transmission policy. The pseudo-code is summarized in Algorithm 2.

Remark 3. *To the best of our knowledge, the intersection search method is the first to solve single-constraint CMDPs by leveraging the piece-wise linearity of $\mathcal{L}^\lambda$. Suppose $\mathcal{L}^\lambda$ has W segments in the range $[0, \lambda_{max}]$. The Insec-RVI method requires at most $W-1$ iterations, which is much fewer compared to the Bisec-RVI method as $W - 1 \ll \log_2(\frac{\lambda_{max}}{\xi})$ when $\xi \to 0$.*

Algorithm 2: Insec-RVI for solving $\mathscr{P}_{\text{CMDP}}$.

Input: c, f, P, $\lambda_{\max}, \epsilon, \xi$
Output: γ, π^*

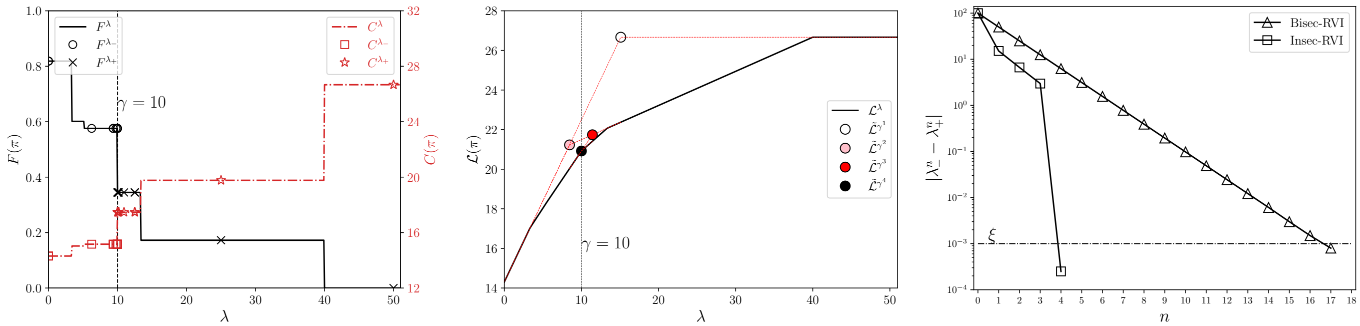
/* check if $\gamma = 0$ */
$\pi^*_{\lambda_-}, F^{\lambda_-}, C^{\lambda_-}, \mathcal{L}^{\lambda_-} \leftarrow \text{RVI}(0, f, d, P, \epsilon)$
if $F^{\lambda_-} \leq F_{max}$ **then**
  $\gamma \leftarrow 0, \pi^* \leftarrow \pi^*_{\lambda_-}$
  return γ, π^*
end
/* find a $\gamma > 0$ with $F^\gamma = F_{\max}$ */
$F^{\lambda_+} \leftarrow 0, C^{\lambda_+} \leftarrow C_{\max}, \mathcal{L}^{\lambda_+} \leftarrow C_{\max}$
while *True* **do**
  // Compute the intersection point
  $\gamma = (C^{\lambda_+} - C^{\lambda_-})/(F^{\lambda_-} - F^{\lambda_+})$
  $\tilde{\mathcal{L}}^\gamma = F^{\lambda_-}(\gamma - \lambda_-) + \mathcal{L}^{\lambda_-}$
  // Check the stopping condition
  $F^\gamma, C^\gamma, \mathcal{L}^\gamma \leftarrow \text{RVI}(\gamma, f, d, P, \epsilon)$
  if $\mathcal{L}^\gamma = \tilde{\mathcal{L}}^\gamma$ **then**
    // Compute the optimal policy
    if $F^\gamma = F_{max}$ **then**
      $\pi^* \leftarrow \pi^*_\gamma$
    else
      $\pi^*_{\lambda_-}, F^{\lambda_-} \leftarrow \text{RVI}(\gamma - \xi/4, f, d, P, \epsilon)$
      $\pi^*_{\lambda_+}, F^{\lambda_+} \leftarrow \text{RVI}(\gamma + \xi/4, f, d, P, \epsilon)$
      $\mu \leftarrow (F_{\max} - F^{\lambda_+})/(F^{\lambda_-} - F^{\lambda_+})$
      $\pi^* \leftarrow \mu\pi^*_{\lambda_-} + (1-\mu)\pi^*_{\lambda_+}$
    end
    return γ, π^*
  else
    // Shrink the interval
    if $F^\gamma \leq F_{max}$ **then**
      $\lambda_+ \leftarrow \gamma$, $C_+^\lambda \leftarrow C^\gamma$, $F_+^\lambda \leftarrow F^\gamma$, $\mathcal{L}_+^\lambda \leftarrow \mathcal{L}^\gamma$
    else
      $\lambda_- \leftarrow \gamma$, $C_-^\lambda \leftarrow C^\gamma$, $F_-^\lambda \leftarrow F^\gamma$, $\mathcal{L}_-^\lambda \leftarrow \mathcal{L}^\gamma$
    end
  end
end

V. Numerical Results

In this section, we validate the performance of the proposed methods through numerical analysis. For ease of analysis, we consider an NCS with $M = 2$ sources. The two sources are assumed to be equally important, i.e., $\omega_1 = \omega_2 = 1$. Each source has three states and is characterized by a symmetric transition matrix, that is, $\mathbf{Q}_{i,j} = p$ for all $i \neq j$ and $1 - 2p$ otherwise. For comparison purposes, the first source is assumed to be slowly evolving with $p = 0.1$, while the second one is rapidly evolving with $p = 0.4$. Also, they have the same CAE matrix given by

$$\delta^1 = \delta^2 = \begin{bmatrix} 0 & 10 & 30 \\ 30 & 0 & 10 \\ 10 & 30 & 0 \end{bmatrix}.$$

Fig. 4 compares the performance of the Bisec-RVI and the Insec-RVI methods. It can be observed that F^λ is piecewise constant and $\mathcal{L}^\lambda$ is piecewise linear and concave. Moreover, $\gamma = 10$ is a breakpoint of F^λ and a corner of $\mathcal{L}^\lambda$. Interestingly, it can be seen from Fig. 4a that for any $F_{\max} \geq 0.8185$ the solution to the unconstrained problem $\mathscr{P}^0_{\text{L-MDP}}$ is also a solution to the constrained problem $\mathscr{P}_{\text{CMDP}}$. It implies that *continuous transmission is unnecessary and has little marginal performance gain.* Additionally, when $F_{\max} \in [0.8185, 1] \cup$



(a) Bisection search. (b) Intersection search. (c) Convergence performance.

Fig. 4. Comparison of different policy search methods when $F_{\max} = 0.4$ and $p_s = 0.4$.

$\{0.8185, 0.6012, 0.5758, 0.3448, 0.1724, 0\}$, the optimal policy is stationary deterministic. Otherwise, there is no such λ that satisfies $F^{\lambda} = F_{\max}$. By Theorem 2, the optimal policy is a combination of two deterministic policies.

Fig. 4c plots the width of the search interval of the Bisec-RVI and the Insec-RVI methods. The Bisec-RVI method demonstrates a logarithmic decrease in the search interval, converging within 17 iterations in this particular case. However, it requires 13 iterations in the subinterval $[5.2, 13.4]$ to meet the convergence tolerance of $\xi = 10^{-3}$. Also, the number of iterations goes to infinity as $\epsilon \to 0$. On the other hand, the Insec-RVI algorithm requires only 4 iterations and it terminates immediately when reaching this subinterval.

Fig. 5 illustrates the transmission frequency of each source. It is observed that the Insec-RVI policy schedules more frequently the slowly evolving source, and increases the importance of the rapidly evolving source in relatively reliable channel conditions. Now we look into how the Insec-RVI policy caters to state-dependent requirements. Let T^1 and T^2 denote the transmission frequencies of the slowly evolving and rapidly evolving sources obtained by the Insec-RVI policy, respectively. Here, $T^1_{i,j}(T^2_{i,j})$ denotes the transmission frequency when the slowly (rapidly) evolving source is in state i and the reconstructed state is j. When $p_s = 0.4$ and $F_{\max} = 0.8$, the values of T^1 and T^2 are

$$T^1 = \begin{bmatrix} 0 & 0.045 & 0.058 \\ 0.058 & 0 & 0.045 \\ 0.045 & 0.058 & 0 \end{bmatrix}, T^2 = \begin{bmatrix} 0 & 0.066 & 0.098 \\ 0.098 & 0 & 0.066 \\ 0.066 & 0.098 & 0 \end{bmatrix}.$$

In this specific case, we can observe a two-level significance. Firstly, the rapidly evolving source is more important than the slowly evolving source. Secondly, some states are scheduled more frequently, indicating they are more critical than others. On the other hand, the transmission frequencies under the distortion-optimal policy can be obtained by choosing equal cost for all errors, i.e., $\delta^1_{i,j} = \delta^2_{i,j} = 1, \forall i \neq j$. The results are

$$T^1 = \begin{bmatrix} 0 & 0.058 & 0.058 \\ 0.058 & 0 & 0.058 \\ 0.058 & 0.058 & 0 \end{bmatrix}, T^2 = \begin{bmatrix} 0 & 0.076 & 0.076 \\ 0.076 & 0 & 0.076 \\ 0.076 & 0.076 & 0 \end{bmatrix}.$$

Obviously, the distortion-optimal policy is incapable of satisfying state-dependent and context-aware requirements in such systems.

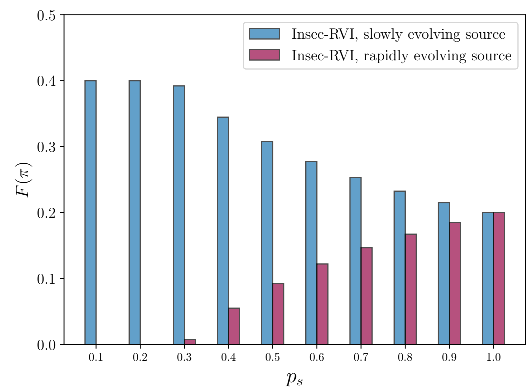


Fig. 5. The transmission frequency vs. p_s.

VI. Conclusions

In this work, we studied the semantic-aware remote estimation of multiple Markov sources over a lossy and rate-constrained channel. We proved the structure of the constrained optimal policy and proposed a novel policy search method that can find the optimal policy using only a few iterations. *An important takeaway message is that we can significantly reduce the number of uninformative transmissions by exploiting the timing of important information.*

References

[1] J. Chakravorty and A. Mahajan, "Fundamental limits of remote estimation of autoregressive markov processes under communication constraints," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, 2017.

[2] N. Pappas and M. Kountouris, "Goal-oriented communication for real-time tracking in autonomous systems," in *IEEE ICAS*, 2021.

[3] L. Zhang, H. Gao, and O. Kaynak, "Network-induced constraints in networked control systems—a survey," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, 2013.

[4] J. Gielis, A. Shankar, and A. Prorok, "A critical review of communications in multi-robot systems," *Current robotics reports*, vol. 3, no. 4, 2022.

[5] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, vol. 59, no. 6, 2021.

[6] P. Popovski *et al.*, "A perspective on time toward wireless 6g," *Proceedings of the IEEE*, vol. 110, no. 8, 2022.

[7] J. Wu, Q.-S. Jia, K. H. Johansson, and L. Shi, "Event-based sensor data scheduling: Trade-off between communication rate and estimation quality," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, 2013.
[8] A. S. Leong, A. Ramaswamy, D. E. Quevedo, H. Karl, and L. Shi, "Deep reinforcement learning for wireless sensor scheduling in cyber–physical systems," *Automatica*, vol. 113, 2020.
[9] M. Pezzutto, L. Schenato, and S. Dey, "Transmission power allocation for remote estimation with multi-packet reception capabilities," *Automatica*, vol. 140, 2022.
[10] G. Cocco, A. Munari, and G. Liva, "Remote monitoring of two-state markov sources via random access channels: an information freshness vs. state estimation entropy perspective," *IEEE Journal on Selected Areas in Information Theory*, vol. 4, 2023.
[11] N. Pappas, M. A. Abd-Elmagid, B. Zhou, W. Saad, and H. S. Dhillon, *Age of Information: Foundations and Applications*. Cambridge University Press, 2023.
[12] V. Tripathi *et al.*, "Wiswarm: Age-of-information-based wireless networking for collaborative teams of uavs," in *IEEE INFOCOM*, 2023.
[13] P. Kutsevol, O. Ayan, N. Pappas, and W. Kellerer, "Experimental study of transport layer protocols for wireless networked control systems," in *IEEE SECON*, 2023.
[14] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, "Age-of-information vs. value-of-information scheduling for cellular networked control systems," in *Proceedings of the 10th ACM/IEEE ICCPS*, 2019.
[15] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the wiener process for remote estimation over a channel with random delay," *IEEE Transactions on Information Theory*, vol. 66, no. 2, 2019.
[16] A. Maatouk, M. Assaad, and A. Ephremides, "The age of incorrect information: An enabler of semantics-empowered communication," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, 2023.
[17] Y. Chen and A. Ephremides, "Minimizing age of incorrect information for unreliable channel with power constraint," in *IEEE GLOBECOM*, 2021.
[18] G. Stamatakis, N. Pappas, and A. Traganitis, "Control of status updates for energy harvesting devices that monitor processes with alarms," in *IEEE Globecom Workshops*, 2019.
[19] S. Jayanth, N. Pappas, and R. V. Bhat, "Distortion minimization with age of information and cost constraints," in *IEEE WiOpt*, 2023.
[20] A. Nikkhah, A. Ephremides, and N. Pappas, "Age of actuation in a wireless power transfer system," in *IEEE INFOCOM WKSHPS*, 2023.
[21] X. Zheng, S. Zhou, and Z. Niu, "Urgency of information for context-aware timely status updates in remote control systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, 2020.
[22] M. Salimnejad, M. Kountouris, and N. Pappas, "Real-time reconstruction of markov sources and remote actuation over wireless channels," *IEEE Transactions on Communications*, 2024.
[23] ——, "State-aware real-time tracking and remote reconstruction of a markov source," *Journal of Communications and Networks*, vol. 25, no. 5, 2023.
[24] E. Fountoulakis, N. Pappas, and M. Kountouris, "Goal-oriented policies for cost of actuation error minimization in wireless autonomous systems," *IEEE Communications Letters*, vol. 27, no. 9, 2023.
[25] A. Zakeri, M. Moltafet, and M. Codreanu, "Semantic-aware sampling and transmission in real-time tracking systems: A pomdp approach," *techrxiv*, 2024.
[26] F. J. Beutler and K. W. Ross, "Optimal policies for controlled markov chains with a constraint," *Journal of mathematical analysis and applications*, vol. 112, no. 1, 1985.
[27] L. I. Sennott, "Constrained average cost markov decision chains," *Probability in the Engineering and Informational Sciences*, vol. 7, no. 1, 1993.
[28] E. Altman, *Constrained Markov decision processes*. CRC Press, 1999.
[29] J. Luo and N. Pappas, "Semantic-aware remote estimation of multiple markov sources under constraints," *arXiv*, 2024.
[30] D.-j. Ma, A. M. Makowski, and A. Shwartz, "Estimation and optimal control for constrained markov chains," in *IEEE CDC*, 1986.
[31] C. Tessler, D. J. Mankowitz, and S. Mannor, "Reward constrained policy optimization," in *International Conference on Learning Representations (ICLR)*, 2019.
[32] D. V. Djonin and V. Krishnamurthy, "Q-learning algorithms for constrained markov decision processes with randomized monotone policies: Application to mimo transmission control," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, 2007.
[33] M. Ono, M. Pavone, Y. Kuwata, and J. Balaram, "Chance-constrained dynamic programming with application to risk-aware robotic space exploration," *Autonomous Robots*, vol. 39, 2015.