Goal-oriented Tensor: Beyond AoI Towards Semantics-Empowered Goal-Oriented Communications

Aimin Li*‡, Shaohua Wu*†, and Sumei Sun‡
* Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology,
Harbin Institute of Technology (Shenzhen), China
† The Department of Boradband Communication, Peng Cheng Laboratory, China
‡Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore
E-mail: liaimin@stu.hit.edu.cn; hitwush@hit.edu.cn; sunsm@i2r.a-star.edu.sg

2024 IEEE Wireless Communications and Networking Conference (WCNC) | 979-8-3503-0358-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/WCNC57260.2024.10570877

***Abstract*—The intricate interplay of source dynamics, unreliable channels, and *staleness* of information has long been recognized as a significant impediment for the receiver to achieve accurate, timely, and goal-oriented decision making. Thus, a plethora of promising metrics, such as *Age of Information* and *Value of Information* have emerged to quantify these adverse factors. Optimizing these metrics indirectly improves the goal-oriented *utility* of decision making. Nevertheless, no metric has been devised to directly evaluate the *utility*. To this end, this paper investigates a novel tensor-based metric, named Goal-oriented Tensor (GoT), to directly quantify the impact of semantic mismatches on decision making. Leveraging the GoT, we design a *sampler-decision maker* pair that works collaboratively to achieve a shared goal. This sampling-decision making co-design is challenging since the sampler and the decision maker are strongly coupled. To decouple these processes, we formulate the problem as an infinite-horizon Decentralized Partially Observable Markov Decision Process (Dec-POMDP) to conjointly deduce the optimal joint policy. We tested the *sampler-decision maker* co-design in terms of goal achievement *utility* and sampling rate, achieving significant performance advancements over conventional state-of-the-art sampling methodologies.**

***Index Terms*—Goal-oriented Semantic communications, Goal-oriented Tensor, Status updates, Age of Information, Age of Incorrect Information, Value of Information.**

I. Introduction

Age of Information (AoI) has emerged as an important metric to capture the data *freshness* perceived by the receiver [1] and to facilitate the development of *freshness*-critical network design. Since its inception, AoI has garnered significant research attention and has been extensively analyzed and optimized to improve the performance of queuing systems, physical-layer communications, MAC-layer communications, Internet of Things, etc. [2]. Since a fresh message typically contains critical and valuable information to improve the precision and timeliness of decision-making processes, optimizing AoI has been an attractive research interest.

However, AoI exhibits several critical shortcomings. Specifically, (a) AoI does not provide a direct measure of information value; (b) AoI does not consider the content dynamics of source data; (c) AoI ignores the effect of End-to-End (E2E) information mismatch on the decision-making process. To address these limitations, numerous AoI variants have been extensively investigated. One typical approach in this research avenue is to impose a non-linear penalty on AoI [3]–[5]. This non-linear penalty is called *Value of Information* (VoI), which mitigates the shortcoming (a). Other research attempt to address the shortcoming (b) [6]–[8]. In [6], *Age of Changed Information* (AoCI) is proposed to address the ignorance of content dynamics of AoI. In this regard, unchanged statuses do not necessarily provide new information and thus are not prioritized for transmission. In [7], the authors proposed a new metric to integrate both similarity and age of information to facilitate task-oriented semantics-aware UAV control. In [8], the authors propose an age penalty named *Age of Synchronization* (AoS) to address the shortcoming (b). *Mean Square Error* (MSE) and its variants, like *Urgency of Information* (UoI) [9] and *Age of Incorrect Information* (AoII) [10] are introduced to address the shortcoming (c).

Though the above metrics have addressed the shortcomings of AoI, they do not explicitly reveal *how these metrics affect the utility of decision making*. To answer this question, [11]–[14] introduce a novel metric termed *Cost of Actuation Error* to delve into the cost resulting from the error actuation due to imprecise real-time estimations. Specifically, the *Cost of Actuation Error* $C_{X_t,\hat{X}_t}$ represents the cost under the E2E mismatch $(X_t,\hat{X}_t)_{X_t\neq\hat{X}_t}$, where X_t is the semantic status at the source at time t and $\hat{X}_t$ represents the estimated semantic status at the receiver at time t. The metric unveils that the *utility* of decision making is dependent on the E2E semantic mismatch category, instead of the mismatch duration (AoII) or mismatch degree (MSE) only. For example, an E2E semantic mismatch that a fire is estimated as no fire will result in higher cost; while the opposite scenario will result in lower cost. The cost per unit of time is directly dependent on the distortion category, rather than the mismatch duration.

The metric *Cost of Actuation Error* maps E2E semantic mismatch to real-world actuation error cost. However, this metric is still limited in the following aspects: $i)$ the method

This work has been supported in part by the National Key Research and Development Program of China under Grant no. 2020YFB1806403, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant no. 2022B1515120002, in part by the China Scholarship Council under Grant no. 202306120310, and in part by the Fundamental Research Funds for the Central Universities under Grant no. HIT.DZJJ.2023116.

to obtain a *Cost of Error Actuation* implicitly necessitates a pre-established fixed decision-making policy, which remains unclear; ii) it does not consider the context issue, which also affects the *significance* of information; iii) it holds that *if* $X_t = \hat{X}_t$, *then* $C_{X_t,\hat{X}_t} = 0$, thereby signifying that errorless actuation necessitates no energy expenditure. To address these issues, the present authors have recently proposed a new metric named Goal-oriented Tensor (GoT) in [15], which, compared to *Cost of Error Actuation*, introduces a 3-dimension tensor-based approach to describe the true *utility* of decision making. In this paper, we further technically extend the dimension of GoT and exploit its potential to achieve a goal-oriented E2E *sampler-decision maker* co-design, the contributions of this work are summarized as follows:

- We introduce a 4-dimension GoT as a novel metric to directly describe the decision *utility* issue. This metric is an extension of the 3-dimension GoT in [15]. A new dimension, the decision policy π_A is introduced in this paper to determine the value in GoT. This extension is crucial as it facilitates the co-design of sampling and decision-making processes.
- Three aspects of decision-making *utility* are fully considered in this paper: i) the future evolution of the source; ii) the instant cost at the source; iii) the energy and resources consumed by actuation. Compared to the state-of-the-art results in [11]–[14], utility i) and iii) are further considered.
- We accomplish the goal-oriented *sampler-decision maker* co-design. To the best of our knowledge, this represents the first effort that co-design the semantics-aware sampling and decision making. To jointly determine the sampling policy and decision-making policy, we introduce the *team decision theory* and formulate this problem as a two-agent infinite-horizon Dec-POMDP problem, with one agent embodying the sampler and the other representing the decision maker. We designed the Brute-Force-Search-RVI algorithm to solve this problem, which is proven optimal and outperforms prevailing state-of-the-art goal-irrelevant sampling methodologies.

II. System Model

We consider a time-slotted perception-actuation loop as shown in Fig. 1. The semantics $X_t \in \mathcal{S} = \{s_1, \cdots, s_{|\mathcal{S}|}\}$ and context $\Phi_t \in \mathcal{V} = \{v_1, \cdots, v_{|\mathcal{V}|}\}$ are extracted from the observed source, which are then input into a sampler to determine the *significance* of X_t and decide if it warrants transmission via an unreliable channel. The estimated semantic status denoted by $\hat{X}_t \in \mathcal{S}$ represents the most recently received semantic status. We consider a perfect and delay-free feedback channel [11]–[14], with ACK representing a successful transmission and NACK representing the otherwise. In this setup, the sampler could perfectly capture the estimated status at the receiver $\hat{X}_t$, which implies that the sampler is fully observable. The binary indicator $a_S(t) = \pi_S(X_t, \Phi_t, \hat{X}_t) \in \{0, 1\}$ signifies the sampling and transmission action at time slot t, with value 1 representing sampling and transmission, and value 0 the otherwise. $\pi_S : \mathcal{S} \times \mathcal{V} \times \mathcal{S} \rightarrow \{0, 1\}$ represents the sampling policy. The decision maker at the receiver will make decisions $a_A(t) = \pi_A(\hat{X}(t)) \in \mathcal{A}_A = \{a_1, \cdots, a_{|\mathcal{A}_A|}\}$ based on the estimated status $\hat{X}_t$, where $\pi_A : \mathcal{S} \rightarrow \mathcal{A}_A$ represents the decision making policy.

In this system, our objective is to facilitate collaboration between the semantics-aware sampler and the decision maker. This interaction involves a balance: frequent sampling typically yields precise estimations but incurs substantial sampling costs. Conversely, infrequent sampling conserves energy at the source yet necessitates the decision maker to allocate additional resources to meet the objective.

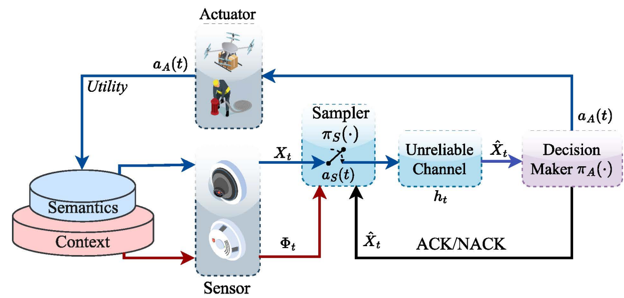


Fig. 1. Illustration of our considered system where transmitted semantic status arrives at a receiver for decision making to achieve a certain goal.

A. Semantics and Context Dynamics

We consider a controlled Discrete Markov source:

$$\Pr\left(X_{t+1} = s_u \left| X_t = s_i, a_A(t) = a_m, \Phi_t = v_k \right.\right) = p_{i,u}^{(k,m)}. \quad (1)$$

Here the dynamics of semantics $X(t)$ is dependent on both the decision making $a_A(t)$ and context Φ_t. Furthermore, we consider that the context Φ_t is a Markov chain, given as:

$$\Pr\left(\Phi_{t+1} = v_r \left| \Phi_t = v_k \right.\right) = p_{k,r}. \quad (2)$$

In this paper, the dynamics of semantics and context are independent of each other.

B. Unreliable Channel and Estimate Transition

We assume that the channel realizations exhibit independence and identical distribution (i.i.d.) across time slots, following a Bernoulli distribution. Particularly, the channel realization h_t satisfies that $\Pr(h_t = 1) = p_S$ and the failure probability as $\Pr(h_t = 0) = 1 - p_S$, with p_S representing the successful transmission probability. To characterize the dynamics of $\hat{X}_t$, we consider two cases as described below:

- $a_S(t) = 0$. In this case, the sampler and transmitter remain idle, manifesting that there is no new knowledge given to the receiver, *i.e.*, $\hat{X}_{t+1} = \hat{X}_t$:

$$\Pr\left(\hat{X}_{t+1} = x \middle| \hat{X}_t = s_j, a_S(t) = 0\right) = \mathbb{1}_{\{x = s_j\}}. \quad (3)$$

- $a_S(t) = 1$. In this case, the sampler and transmitter transmit the current semantic status X_t through an unreliable channel. As the channel is unreliable, we differentiate between two distinct situations: $h_t = 1$ and $h_t = 0$:

(a) $h_t = 1$. In this case, the transmission is successful. As such, the estimate at the receiver $\hat{X}_{t+1}$ is nothing but $X(t)$, and the transition probability is

$$\Pr\left(\hat{X}_{t+1}=x\,\middle|\,\hat{X}_t=s_j,X_t=s_i,a_S(t)=1,h_t=1\right)\\=\mathbb{1}_{\{x=s_i\}}. \tag{4}$$

(b) $h_t=0$. In this case, the transmission is not successfully decoded by the receiver. As such, the estimate at the receiver $\hat{X}_{t+1}$ remains $\hat{X}(t)$. In this way, the transition probability is

$$\Pr\left(\hat{X}_{t+1}=x\,\middle|\,\hat{X}_t=s_j,X_t=s_i,a_S(t)=1,h_t=0\right)=\mathbb{1}_{\{x=s_j\}}. \tag{5}$$

As the channel realization h_t is independent with the process of X_t, $\hat{X}_t$, and $a_S(t)$, we have that

$$\begin{aligned}&\Pr\left(\hat{X}_{t+1}=x\,\middle|\,\hat{X}_t=s_j,X_t=s_i,a_S(t)=1\right)\\&=\sum_{h_t}p\left(h_t\right)\Pr\left(\hat{X}_{t+1}=x\,\middle|\,\hat{X}_t=s_j,X_t=s_i,a_S(t)=1,h_t\right)\\&=p_S\cdot\mathbb{1}_{\{x=s_i\}}+(1-p_S)\cdot\mathbb{1}_{\{x=s_j\}}.\end{aligned} \tag{6}$$

Combining (3) with (6) yields the dynamics of the estimate.

C. Goal-oriented Decision Making and Actuating

We note that the previous works focus on the sampling process only, regardless the pragmatics of sampled information, which is achieved through decision-making and actuating. To this end, our paper integrates the decision-making and actuation processes into the sampling policy design. The decision-making and actuation enable the conversion of sampled information into ultimate effectiveness. Here the decision making at time slot t is denoted by $a_A(t)=\pi_A(\hat{X}_t)$, with π_A representing the deterministic decision-making policy.

D. Metric: Goal Characterization Through GoT

A three-dimension GoT could be defined by a mapping:

$$(X_t,\Phi_t,\hat{X}_t)\in\mathcal{S}\times\mathcal{V}\times\mathcal{S}\xrightarrow{\mathcal{L}}\mathrm{GoT}(t)\in\mathbb{R}. \tag{7}$$

In this regard, the GoT, denoted by $\mathcal{L}(X_t,\Phi_t,\hat{X}_t)$ or $\mathrm{GoT}(t)$, indicates the instant cost of the system at time slot t, with the knowledge of $(X_t,\Phi_t,\hat{X}_t)$. From [15], we have shown that a GoT, given a specific triple-tuple $(X_t,\hat{X}_t,\Phi_t)$ and a decision-making policy π_A, could be calculated by

$$\begin{aligned}&\mathrm{GoT}(t)=\mathcal{L}(X_t,\Phi_t,\hat{X}_t|\pi_A)\\&=\Big[C_1(X_t,\Phi_t)-C_2(\pi_A(\hat{X}_t))\Big]^+ + C_3(\pi_A(\hat{X}_t)),\end{aligned} \tag{8}$$

where the status inherent cost $C_1(X_t,\Phi_t)$ quantifies the inherent cost under different semantics-context pairs (X_t,Φ_t) in the absence of external influences; the actuation gain cost $C_2(\pi_A(\hat{X}_t))$ quantifies the prospective reduction in severity resulting from the actuation $\pi_A(\hat{X}(t))$; the actuation inherent cost $C_3(\pi_A(\hat{X}_t))$ reflects the resources consumed by a particular actuation $\pi_A(\hat{X}(t))$. The ramp function $[\cdot]^+$ ensures that any actuation $\pi_A(\hat{X}_t)$ reduces the cost to a maximum of 0. A visualization of a specific 3-dimension GoT is shown in Fig. 2, which is obtained through the following definition:

$$\begin{aligned}&C_1(X_t,\Phi_t)=\left(\begin{array}{c|ccc} & 0 & 1 & 2\\\hline 0 & 0 & 1 & 3\\ 1 & 0 & 2 & 5\end{array}\right), \quad \pi_A(\hat{X}_t)=[0,1,2],\\&C_2(\pi_A(\hat{X}_t))=2\pi_A(\hat{X}_t), \qquad C_3(\pi_A(\hat{X}_t))=\pi_A(\hat{X}_t).\end{aligned} \tag{9}$$

In this paper, the 3-dimension GoT $\mathrm{GoT}(t)$ is extended to the 4-dimensions GoT $\mathrm{GoT}^{\pi_A}(t)$ by integrating the actuation policies π_A as an additional dimension. Under different π_A, we could obtain different 3-dimension GoT.

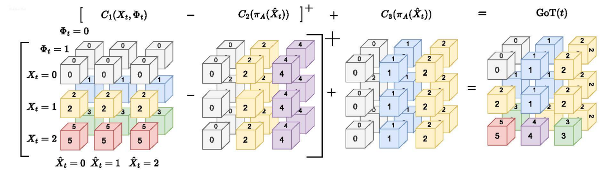
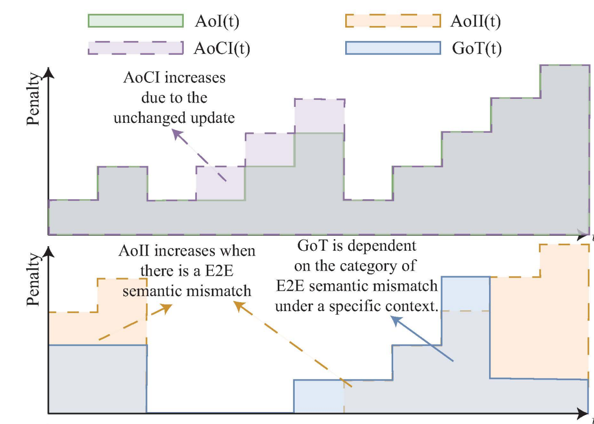


Fig. 2. A visualized example for characterizing the GoT through (8) and (9).



t	0	1	2	3	4	5	6	7	8	9	10
$a_S(t)$	0	1	1	0	0	1	0	1	1	1	0
h_t	\	1	1	\	\	1	\	0	0	0	\
X_t	0	0	0	0	0	1	1	2	2	0	0
$\hat{X}_t$	2	2	0	0	0	0	1	1	1	1	1
Φ_t	0	1	1	1	0	0	0	0	1	1	0

Fig. 3. An illustration of AoI, AoCI, AoII, and GoT in a time-slotted status update system. Here, the value of GoT is obtained from the tensor obtained on the right-hand side of Fig. 2.

To illustrate how GoT characterizes the goal, Fig. 3 exhibits an instantaneous progression of AoI, AoCI, AoII, and GoT. From the time slots $t=0,1,9,10$ in Fig. 3, the inherent limitation of AoII emerges conspicuously, as a duration of mismatch may not necessarily culminate in a cost increase. Instead, the category of E2E semantic mismatch will make sense to the true instant cost.

III. Problem Formulation and Solution

Conventionally, the sampling policy and the decision-making policy are designed in a two-stage manner: first, they determine the optimal sampling policy based on AoI or its variants; and second, they accomplish decision making given the sampling policy. This two-stage separation arises from the inherent limitation of existing metrics that they fail to capture the closed-loop decision *utility*. However, the four-dimension GoT empowers us to achieve the co-design of sampling and decision making. Intuitively, a co-design approach holds the potential to outperform designs that treat components separately.

However, achieving the co-design is a complex work since the sampling and decision-making processes are strongly coupled. To this end, we introduce the *team decision theory* to decouple them. Two agents, one embodying the sampler and the other the decision maker, collaborate to achieve a shared goal. We aim at determining a joint deterministic policy $\boldsymbol{\pi}_C = (\pi_S, \pi_A)$ that minimizes the long-term average cost of the system. It is considered that the sampling and transmission of an update also consumes energy, incurring a C_s cost. In this case, the instant cost of the system could be clarified by $\mathrm{GoT}^{\pi_A}(t) + C_s \cdot a_S(t)$, and the problem is characterized as:

$$\mathcal{P}1: \quad \min_{\boldsymbol{\pi}_C \in \Upsilon} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^{\boldsymbol{\pi}_C}\left(\sum_{t=0}^{T-1} \mathrm{GoT}^{\pi_A}(t) + C_s \cdot a_S(t)\right), \tag{10}$$

where $\boldsymbol{\pi}_C = (\pi_S, \pi_A)$ denotes the joint sampling and decision policy, comprising $\pi_S = (a_S(0), a_S(1), \cdots)$ and $\pi_A = (a_A(0), a_A(1), \cdots)$, which correspond to the sampling action sequence and actuation sequence, respectively.

A. Dec-POMDP Formulation

We formulate problem $\mathcal{P}1$ as a *Decentralized Partially Observable Markov Decision Processes* (DEC-POMDP) problem, which is initially introduced in [16] to solve the cooperative sequential decision issues for distributed multi-agents. Within a Dec-POMDP framework, a team of agents cooperates to achieve a shared goal, relying solely on their localized knowledge. A Dec-POMDP could be denoted by a tuple $\mathscr{M}_{DEC-POMDP} \triangleq \langle n, \mathcal{I}, \mathcal{A}, \mathcal{T}, \Omega, \mathcal{O}, \mathcal{R}\rangle$:

- n denotes the number of agents. In problem $\mathcal{P}1$ we have $n = 2$, signifying the presence of two agents: one agent $\mathcal{A}gent_S$ embodies the sampler, while the other represents the decision maker, denoted by $\mathcal{A}gent_A$.
- $\mathcal{I}$ is the set of the global system status. In $\mathcal{P}1$, this is characterized by $(X_t, \hat{X}_t, \Phi_t) \in \mathcal{S} \times \mathcal{S} \times \mathcal{V}$. For the sake of brevity, we henceforth denote $\mathbf{W}_t = (X_t, \hat{X}_t, \Phi_t)$ in the sequel.
- $\mathcal{T}$ is the transition function defined by

$$\mathcal{T}(\mathbf{w}, \mathbf{a}, \mathbf{w}') \triangleq \Pr(\mathbf{W}_{t+1} = \mathbf{w}' | \mathbf{W}_t = \mathbf{w}, \mathbf{a}_t = \mathbf{a}), \tag{11}$$

which is defined by the transition probability from global status $\mathbf{W}_t = \mathbf{w}$ to status $\mathbf{W}_{t+1} = \mathbf{w}'$, after the agents in the system taking a joint action $\mathbf{a}_t = \mathbf{a} = (a_S(t), a_A(t))$. For the sake of concise notation, we let $p(\mathbf{w}'|\mathbf{w}, \mathbf{a})$ symbolize $\mathcal{T}(\mathbf{w}, \mathbf{a}, \mathbf{w}')$ in the subsequent discourse.

Lemma 1. *The transition functions of the Dec-POMDP:*

$$\begin{aligned} p\left((s_u, x, v_r)\,|(s_i, s_j, v_k), (1, a_m)\right) &= p_{i,u}^{(k,m)} \cdot p_{k,r} \\ &\cdot \left(p_S \cdot \mathbb{1}_{\{x=s_i\}} + (1-p_S) \cdot \mathbb{1}_{\{x=s_j\}}\right) \end{aligned}, \tag{12}$$

$$p\left((s_u, x, v_r)\,|(s_i, s_j, v_k), (0, a_m)\right) = p_{i,u}^{(k,m)} \cdot p_{k,r} \cdot \mathbb{1}_{\{x=s_j\}}, \tag{13}$$

for any $x \in \mathcal{S}$ and indexes i, j, $u \in \{1, 2, \cdots, |\mathcal{S}|\}$, k, $r \in \{1, 2, \cdots, |\mathcal{V}|\}$, and $m \in \{1, 2, \cdots, |\mathcal{A}_A|\}$.

Proof. Refer to [17, Appendix A] for the proof. □

- $\mathcal{A} = \mathcal{A}_S \times \mathcal{A}_A$, with $\mathcal{A}_S \triangleq \{0, 1\}$ representing the action set of the sampler, and $\mathcal{A}_A \triangleq \{a_0, \cdots, a_{M-1}\}$ representing the action set of the decision maker.
- $\Omega = \Omega_S \times \Omega_A$, with Ω_S signifies the sampler's observation domain. In this instance, the sampler $\mathcal{A}gent_S$ is entirely observable, with Ω_S encompassing the comprehensive system state $o_S^{(t)} = \mathbf{W}_t$. Ω_A signifies the actuator's observation domain. In this case, the decision-maker $\mathcal{A}gent_A$ is partially observable, with Ω_A comprising $o_A^{(t)} = \hat{X}(t)$. The joint observation at time instant t is denoted by $\mathbf{o}_t = (o_S^{(t)}, o_A^{(t)})$.
- $\mathcal{O} = \mathcal{O}_S \times \mathcal{O}_A$ represents the observation function, where $\mathcal{O}_S$ and $\mathcal{O}_A$ denotes the observation function of the sampler $\mathcal{A}gent_S$ and the actuator $\mathcal{A}gent_A$, respectively, defined as:

$$\mathcal{O}(\mathbf{w}, \mathbf{o}) \triangleq \Pr(\mathbf{o}_t = \mathbf{o} | \mathbf{W}_t = \mathbf{w}), \tag{14}$$

$$\mathcal{O}_S(\mathbf{w}, o_S) \triangleq \Pr(o_S^{(t)} = o_S | \mathbf{W}_t = \mathbf{w}), \tag{15}$$

$$\mathcal{O}_A(\mathbf{w}, o_A) \triangleq \Pr(o_A^{(t)} = o_A | \mathbf{W}_t = \mathbf{w}). \tag{16}$$

The observation function of an agent $\mathcal{A}gent_i$ signifies the conditional probability of agent $\mathcal{A}gent_i$ perceiving o_i, contingent upon the prevailing global system state as $\mathbf{W}_t = \mathbf{w}$. For the sake of brevity, we henceforth let $p_A(o_A|\mathbf{w})$ represent $\mathcal{O}_A(\mathbf{w}, o_A)$ and $p_S(o_S|\mathbf{w})$ represent $\mathcal{O}_S(\mathbf{w}, o_A)$ in the subsequent discourse. In our considered model, the observation functions are deterministic, characterized by lemma 2.

Lemma 2. *The observation functions of the Dec-POMDP:*

$$p_S\left((s_u, s_r, v_q)|(s_i, s_j, v_k)\right) = \mathbb{1}_{\{(s_u, s_r, v_q) = (s_i, s_j, v_k)\}}, \tag{17}$$

$$p_A\left(s_z|(s_i, s_j, v_k)\right) = \mathbb{1}_{\{s_z = s_j\}}. \tag{18}$$

for indexes z, i, j, u, $r \in \{1, 2, \cdots |\mathcal{S}|\}$, and k, $q \in \{1, 2, \cdots |\mathcal{V}|\}$.

- $\mathcal{R}$ is the reward function, characterized by a mapping $\mathcal{I} \times \mathcal{A} \to \mathbb{R}$. In the long-term average reward maximizing setup, resolving a Dec-POMDP is equivalent to addressing the following problem:

$$\min_{\boldsymbol{\pi}_C \in \Upsilon} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^{\boldsymbol{\pi}_C}\left(-\sum_{t=0}^{T-1} r(t)\right). \tag{19}$$

Subsequently, to establish the congruence with the problem $\mathcal{P}1$, the reward function is correspondingly defined as:

$$r(t) = \mathcal{R}^{\pi_A}(\mathbf{w}, a_S) = -\mathrm{GoT}^{\pi_A}(t) - C_s \cdot a_S(t). \tag{20}$$

B. Solutions to the Infinite-Horizon Dec-POMDP

Solving a Dec-POMDP is known to be NEXP-complete for the finite-horizon setup [16]. For an infinite-horizon Dec-POMDP problem, finding an optimal policy for a Dec-POMDP problem is usually undecidable. However, both the sampling and decision-making policies are deterministic within our considered model, given as $a_S(t) = \pi_S(\mathbf{w})$ and $a_A(t) = \pi_A(o_A)$.

In this case, it is feasible to determine a joint optimal policy via Brute Force Search across the decision-making policy space.

The idea is inspired by that given a deterministic decision policy π_A, the sampling problem can be formulated as a standard MDP problem:

Proposition 1. *Given a deterministic decision-making policy π_A, the optimal sampling problem could be formulated by a standard MDP problem $\mathscr{M}_{\mathrm{MDP}}^{\pi_A} \triangleq \langle \mathcal{I}, \mathcal{A}_S, \mathcal{T}_{\mathrm{MDP}}^{\pi_A}, \mathcal{R} \rangle$, where the elements are given as follows:*
- *$\mathcal{I}$: the same as the pre-defined Dec-POMDP tuple.*
- *$\mathcal{A}_S = \{0, 1\}$: the sampling and transmission action set.*
- *$\mathcal{T}^{\pi_A}$: the transition function given a deterministic decision policy π_A, which is*

$$\begin{aligned} \mathcal{T}^{\pi_A}(\mathbf{w}, a_S, \mathbf{w}') &= p^{\pi_A}\left(\mathbf{w}' | \mathbf{w}, a_S\right) \\ &= \sum_{o_A \in \mathcal{O}_A} p\left(\mathbf{w}' | \mathbf{w}, (a_S, \pi_A(o_A))\right) p_A(o_A | \mathbf{w}), \end{aligned} \tag{21}$$

where $p\left(\mathbf{w}' | \mathbf{w}, (a_S, \pi_A(o_A))\right)$ could be obtained by Lemma 1 and $p(o_A|\mathbf{w})$ could be obtained by Lemma 2.
- *$\mathcal{R}$: the same as the pre-defined Dec-POMDP tuple.*

We now proceed to solve the MDP problem $\mathscr{M}_{\mathrm{MDP}}^{\pi_A}$. To deduce the optimal sampling policy under decision policy π_A, it is imperative to resolve the Bellman equations [18]:

$$\theta_{\pi_A}^* + V_{\pi_A}(\mathbf{w}) = \max_{a_S \in \mathcal{A}_A} \left\{ \mathcal{R}^{\pi_A}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}' | \mathbf{w}, a_S) V_{\pi_A}(\mathbf{w}') \right\}, \tag{22}$$

where $V^{\pi_A}(\mathbf{w})$ is the value function and $\theta_{\pi_A}^*$ is the optimal long-term average reward given the decision policy π_A. We apply the relative value iteration (RVI) algorithm to solve this problem. The details are shown in Algorithm 1:

Algorithm 1: The RVI Algorithm to Solve the MDP Given the decision policy π_A

Input: The MDP tuple $\mathscr{M}_{\mathrm{MDP}}^{\pi_A}$, ϵ, π_A;
Initialization: $\forall \mathbf{w} \in \mathcal{I}$, $\tilde{V}_{\pi_A}^{0}(\mathbf{w}) = 0$, $\tilde{V}_{\pi_A}^{-1}(\mathbf{w}) = \infty$, $k = 0$;
Choose $\mathbf{w}^{ref}$ arbitrarily;
while $||\tilde{V}_{\pi_A}^{k}(\mathbf{w}) - \tilde{V}_{\pi_A}^{k-1}(\mathbf{w})|| \geq \epsilon$ **do**
  $k = k + 1$;
  for $\mathbf{w} \in \mathcal{I} - \mathbf{w}^{ref}$ **do**
    $\tilde{V}_{\pi_A}^{k}(\mathbf{w}) = -g_k + \max_{a_S} \left\{ \mathcal{R}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I} - \mathbf{w}^{ref}} p(\mathbf{w}' | \mathbf{w}, a_S) \tilde{V}_{\pi_A}^{k-1}(\mathbf{w}') \right\}$;
$\theta^*(\pi_A, \pi_S^*) = -\tilde{V}_{\pi_A}^{k}(\mathbf{w}) \max_{a_S \in \mathcal{A}_S} \left\{ \mathcal{R}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}' | \mathbf{w}, a_S) \tilde{V}_{\pi_A}^{k}(\mathbf{w}') \right\}$;
for $\mathbf{w} \in \mathcal{I}$ **do**
  $\pi_S^*(\pi_A, \mathbf{w}) = \arg\max_{a_S} \left\{ \mathcal{R}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}' | \mathbf{w}, a_S) \tilde{V}_{\pi_A}^{k}(\mathbf{w}') \right\}$;
Output: $\pi_S^*(\pi_A)$, $\theta^*(\pi_A, \pi_S^*)$

With Proposition 1 and Algorithm 1 in hand, we could then perform a Brute Force Search across the decision policy space Υ_A, thereby acquiring the joint sampling-decision-making policy. The algorithm, called RVI-Brute-Force-Search Algorithm, is elaborated in Algorithm 2. The optimality property of the algorithm is illustrated in Theorem 1.

Algorithm 2: The RVI-Brute-Force-Search Algorithm

Input: The Dec-POMDP tuple $\mathscr{M}_{DEC-POMDP}$;
for $\pi_A \in \Upsilon$ **do**
  Formulate the MDP problem $\mathscr{M}_{\mathrm{MDP}}^{\pi_A} \triangleq \langle \mathcal{I}, \mathcal{A}_S, \mathcal{T}_{\mathrm{MDP}}^{\pi_A}, \mathcal{R} \rangle$ as given in Proposition 1;
  Run Algorithm 1 to obtain $\pi_S^*(\pi_A)$ and $\theta^*(\pi_A, \pi_S^*)$;
Calculate the optimal joint policy:
$\begin{cases} \pi_A^* = \arg\min_{\pi_A} \theta_{\pi_A}^* \\ \pi_S^* = \pi_S(\pi_A^*) \end{cases}$;
Output: π_S^*, π_A^*

Theorem 1. *The RVI-Brute-Force-Search Algorithm (Algorithm 2) could achieve the optimal joint deterministic policies (π_S^*, π_A^*), given that the transition function $\mathcal{T}^{\pi_A}$ follows a unichain.*

Proof. If the the transition function $\mathcal{T}^{\pi_A}$ follows a unichain, we obtain from [19, Theorem 8.4.5] that for any π_A, we could obtain the optimal deterministic policy π_S^* such that $\theta^*(\pi_A, \pi_S^*) \leq \theta^*(\pi_A, \pi_S)$. Also, Algorithm 2 assures that for any π_A, $\theta^*(\pi_A^*, \pi_S^*) \leq \theta^*(\pi_A, \pi_S^*)$. This leads to the conclusion that for any $\boldsymbol{\pi}_C = (\pi_S, \pi_A) \in \Upsilon$, we have that

$$\theta^*(\pi_A^*, \pi_S^*) \leq \theta^*(\pi_A, \pi_S). \tag{23}$$

IV. Simulation Results

For the simulation setup, we set $\mathcal{A}_A = \{0, \cdots, 10\}$, $\mathcal{S} = \{s_0, s_1, s_2\}$, $\mathcal{V} = \{v_0, v_1, v_2\}$ and the corresponding cost is:

$$\mathbf{C}_1(X_t, \Phi_t) = \begin{pmatrix} 0 & 20 & 50 \\ 0 & 10 & 20 \end{pmatrix}, \tag{24}$$

We assume $C_2(\pi_A(\hat{X}_t))$ and $C_3(\pi_A(\hat{X}_t))$ are both linear to the decision making with $C_2(\pi_A(\hat{X}_t) = C_g \cdot \pi_A(\hat{X}_t)$ and $C_3(\pi_A(\hat{X}_t)) = C_I \cdot \pi_A(\hat{X}_t)$, where $C_g = 8$ and $C_I = 1$.

A. Comparing Benchmarks: Separate Design

The conventional separate designs treat sampling and decision making in a two-stage manner:

(1) For the decision-making process, the decision policy π_A is predetermined by a greedy methodology:

$$\begin{aligned} \pi_A(\hat{X}_t) = \arg\min_{a_A \in \mathcal{S}_A} \mathbb{E}_{\Phi_t} \Big\{ & \left[C_1(\hat{X}_t, \Phi_t) - C_2(\pi_A(\hat{X}_t)) \right]^+ \\ & + C_3(\pi_A(\hat{X}_t)) \Big\}. \end{aligned} \tag{25}$$

This greedy-based approach entails selecting the decision that minimizes the cost in the current step given that the estimate $\hat{X}_t$ is perfect. By calculating (25), we obtain a greedy-based decision-making policy $\pi_A(\hat{X}_t) = [0, 3, 7]$.

(2) For the sampling process, the following state-of-the-art comparing benchmarks are considered:

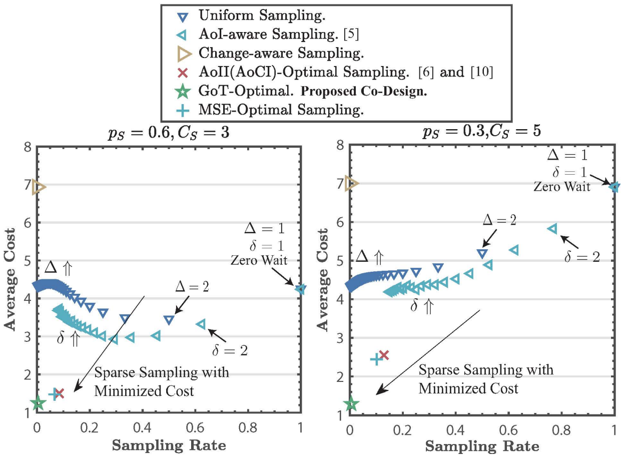


Fig. 4. Average Cost vs. Sampling Rate under different policies and parameters setup. The series of Uniform and AoI-aware policies are obtained through shifting the intervals Δ and δ.

• **Uniform.** Sampling is triggered periodically, *i.e.*, $a_S(t) = \mathbb{1}_{\{t=K*\Delta\}}$, where $K = 0, 1, 2, \cdots$ and $\Delta \in \mathbb{N}^+$. For each Δ, the sampling rate is calculated as $1/\Delta$ and the long-term average cost is obtained through Markov chain simulations.

• **Age-aware.** Sampling is executed when the AoI attains a predetermined threshold, *i.e.*, $a_S(t) = \mathbb{1}_{\{\mathrm{AoI(t)}>\delta\}}$. In Fig. 4, we dynamically shift the threshold δ to explore the balance between sampling rate and *utility*.

• **Change-aware** Sampling is triggered whenever the source status changes, *i.e.*, $a_S(t) = \mathbb{1}_{\{X_t \neq X_{t-1}\}}$.

• **Optimal AoII (also Optimal AoCI).** The AoII-optimal sampling policy turns out to be $a_S(t) = \mathbb{1}_{\{X_t \neq \hat{X}_t\}}$ [10]. From [6], the AoCI-optimal sampling policy is $a_S(t) = \mathbb{1}_{\{X_t \neq X_{t-\mathrm{AoI}(t)}\}}$. Note that $\hat{X}_t = X_{t-\mathrm{AoI}(t)}$, these two sampling policies are equivalent.

B. Co-Design Through GoT

We observe a close intertwining of sampling and decision-making, indicating that a separate design cannot achieve optimal performance. By introducing the 4-dimensional GoT and reformulating the co-design challenge as a Dec-POMDP problem, we decouple these processes. Utilizing the RVI-Brute-Force-Search Algorithm (referenced as Algorithm 2), we solve problem $\mathcal{P}1$, distributively obtaining optimal semantics-aware sampling and goal-relevant decision-making policies.

From Fig. 4, it is evident that the *sampler-decision maker* co-design yields the optimal long-term average *utility* (as represented by the lowest position of the green star on the Y axis) and the sparsest sampling rate (as represented by the lowest position of the green star on the X axis). This is because only information bearing goal-relevant semantics is sampled and transmitted. By integrating a well-matched decision policy, the proposed goal-oriented semantic-aware sampling method outperforms prevailing state-of-the-art results.

V. Conclusion

In this paper, we have investigated the 4-dimension GoT metric to directly describe the goal-oriented system decision-making *utility*. Employing the GoT, we have formulated an infinite horizon Dec-POMDP problem to accomplish the co-design of sampling and decision making. We have developed the RVI-Brute-Force-Search Algorithm to attain optimal joint policies. Comparative analyses have verified that the proposed goal-oriented *sampler-decision maker* co-design enhances sparse sampling while maximizing the *utility*, signifying a promising step towards a sparse sampler and goal-oriented decision maker co-design.

References

[1] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, “Minimizing age of information in vehicular networks,” in *2011 8th Annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks*, 2011, pp. 350–358.

[2] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, “Age of Information: An Introduction and Survey,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, 2021.

[3] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, “Age and value of information: Non-linear age case,” in *2017 IEEE Proceedings ISIT*, 2017, pp. 326–330.

[4] ——, “The cost of delay in status updates and their value: Non-linear ageing,” *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4905–4918, 2020.

[5] Y. Sun and B. Cyr, “Sampling for data freshness optimization: Non-linear age functions,” *J. Commun. Netw.*, vol. 21, no. 3, pp. 204–219, 2019.

[6] X. Wang, W. Lin, C. Xu, X. Sun, and X. Chen, “Age of changed information: Content-aware status updating in the internet of things,” *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 578–591, 2022.

[7] Y. Xu, H. Zhou, and Y. Deng, “Task-Oriented Semantics-Aware Communication for Wireless UAV Control and Command Transmission,” *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 2232–2236, 2023.

[8] J. Zhong, R. D. Yates, and E. Soljanin, “Two freshness metrics for local cache refresh,” in *2018 IEEE Proceedings ISIT*, 2018, pp. 1924–1928.

[9] X. Zheng, S. Zhou, and Z. Niu, “Urgency of information for context-aware timely status updates in remote control systems,” *IEEE Trans. Wirel. Commun.*, vol. 19, no. 11, pp. 7237–7250, 2020.

[10] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, “The age of incorrect information: A new performance metric for status updates,” *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, pp. 2215–2228, 2020.

[11] M. Kountouris and N. Pappas, “Semantics-Empowered Communication for Networked Intelligent Systems,” *IEEE Commun. Mag.*, vol. 59, pp. 96–102, 2020.

[12] N. Pappas and M. Kountouris, “Goal-Oriented Communication For Real-Time Tracking In Autonomous Systems,” in *2021 Proceeding IEEE ICAS*, 2021, pp. 1–5.

[13] M. Salimnejad, M. Kountouris, and N. Pappas, “Real-time Reconstruction of Markov Sources and Remote Actuation over Wireless Channels,” *arXiv*, 2023. [Online]. Available: https://arxiv.org/abs/2302.13927

[14] E. Fountoulakis, N. Pappas, and M. Kountouris, “Goal-oriented policies for cost of actuation error minimization in wireless autonomous systems,” *arXiv*, 2023. [Online]. Available: https://arxiv.org/abs/2303.04908

[15] A. Li, S. Wu, S. M. R. Lu, S. Sun, and Q. Zhang, “Towards goal-oriented semantic communications: New metrics, open challenges, and future research directions,” *Submitted to IEEE Wirel. Commun.*, 2023.

[16] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, “The complexity of decentralized control of markov decision processes,” *Math. Oper. Res.*, vol. 27, no. 4, pp. 819–840, 2002.

[17] A. Li, S. Wu, S. Sun, and J. Cao, “Goal-oriented tensor: Beyond age of information towards semantics-empowered goal-oriented communications,” *Arxiv Preprint*, 2023. [Online]. Available: https://arxiv.org/abs/2307.00535

[18] D. Bertsekas, *Dynamic Programming and Optimal Control: Volume I.* Athena scientific, 2012, vol. 1.

[19] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.