Goal-Oriented Tensor: Beyond Age of Information Towards Semantics-Empowered Goal-Oriented Communications

Aimin Li, *Graduate Student Member, IEEE,* Shaohua Wu, *Member, IEEE,*
Sumei Sun, *Fellow, IEEE*, and Jie Cao, *Member, IEEE*

***Abstract*—Optimizations premised on open-loop metrics such as Age of Information (AoI) indirectly enhance the system's decision-making *utility*. We therefore propose a novel closed-loop metric named Goal-oriented Tensor (GoT) to directly quantify the impact of semantic mismatches on goal-oriented decision-making *utility*. Leveraging the GoT, we consider a *sampler & decision-maker* pair that works collaboratively and distributively to achieve a shared goal of communications. We formulate a two-agent infinite-horizon Decentralized Partially Observable Markov Decision Process (Dec-POMDP) to conjointly deduce the optimal deterministic sampling policy and decision-making policy. To circumvent the *curse of dimensionality* in obtaining an optimal deterministic joint policy through Brute-Force-Search, a sub-optimal yet computationally efficient algorithm is developed. This algorithm is predicated on the search for a Nash Equilibrium between the sampler and the decision-maker. Simulation results reveal that the proposed *sampler & decision-maker* co-design surpasses the current literature on AoI and its variants in terms of both goal achievement *utility* and sparse sampling rate, signifying progress in the semantics-conscious, goal-driven sparse sampling design.**

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

I. Introduction

The recent advancement of the emerging 5G and beyond has spawned the proliferation of both theoretical development and application instances for Internet of Things (IoT) networks. In such networks, timely status updates are becoming increasingly crucial for enabling real-time monitoring and actuation across a plethora of applications. To address the inadequacies of traditional throughput and delay metrics in such contexts, the *Age of Information* (AoI) has emerged as an innovative metric to capture the data freshness perceived by the receiver [2], defined as $\mathrm{AoI}(t) = t - U(t)$, where $U(t)$ denotes the generation time of the latest received packet before time t. Since its inception, AoI has garnered significant research attention and has been extensively analyzed in the queuing systems [3]–[10], physical-layer communications [11]–[17], MAC-layer communications [18]–[22], industrial IoT [23], [24], energy harvesting systems [25]–[28], and etc. (see [29] and the references therein for more comprehensive review). These research efforts are driven by the consensus that a freshly received message typically contains more valuable information, thereby enhancing the *utility* of decision making.

Though AoI has been proven efficient in many freshness-critical applications, it 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 and ignores the effect of End-to-End (E2E) information mismatch on the decision-making process.

To address shortcoming (a), a typical approach is to impose a non-linear penalty on AoI [30]–[33]. In [30], the authors map the AoI to a non-linear and non-decreasing function $f(\mathrm{AoI}(t))$ to evaluate the degree of "*discontent*" resulting from *stale* information. Subsequently, the optimal sampling policy is deduced for an arbitrary non-decreasing penalty function. The authors in [31] introduce two penalty functions, namely the exponential penalty function $a^{\mathrm{AoI}(t)}-1$ and the logarithmic penalty function $\log_a(\mathrm{AoI}(t+1))$, for evaluating the *Cost of Update Delay* (CoUD). In [32], the binary indicator function $\mathbb{1}_{\{\mathrm{AoI(t)>d}\}}$ is applied to evaluate whether the most recently received message is up-to-date. Specifically, the penalty assumes a value of 1 when the instantaneous AoI surpasses a predetermined threshold d; otherwise, the penalty reverts to 0. The freshness of web crawling is evaluated through this AoI-threshold binary indicator function. In [33], an analogous binary indicator approach is implemented in caching systems to evaluate the freshness of information.

The above works tend to tailor a particular non-linear penalty function to evaluate the information value. However, the selection of penalty functions in the above research relies exclusively on empirical configurations, devoid of rigorous derivations. To this end, several information-theoretic techniques strive to explicitly derive the non-linear penalty function in terms of AoI [34]–[37]. One such quintessential work is the auto-correlation function $\mathbb{E}\left[X_t^* X_{t-\mathrm{AoI}(t)}\right]$, which

Part of this work was presented by IEEE WCNC 2024 [1].

This work has been supported in part by the National Key Research and Development Program of China under Grant no. 2020YFB1806403, and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant no. 2022B1515120002, and 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, and in part by the Major Key Project of PCL under Grant no. PCL2024A01, and in part by the Shenzhen Science and Technology Program under Grant no. ZDSYS20210623091808025. (Corresponding author: Shaohua Wu.)

A. Li is with the Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China, and also with the Institute for Infocomm Research, Agency for Science, Technology and Research, 138632, Singapore (e-mail: liaimin@stu.hit.edu.cn; Li_Aimin_from.tp@i2r.a-star.edu.sg).

S. Wu is with the Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China, and also with the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: hitwush@hit.edu.cn).

S. Sun is with the Institute for Infocomm Research, Agency for Science, Technology and Research, 138632, Singapore (e-mail: sunsm@i2r.a-star.edu.sg).

J. Cao is with the Department of Electronic Information Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China (e-mail: caojhitsz@ieee.org).

proves to be a non-linear function of AoI when the source is stationary [34]. Another methodology worth noting is the mutual information metric between the present source state X_t and the vector consisting of an ensemble of successfully received updates $\mathbf{W}_t$ [35], [36]. In the context of a Markovian source, this metric can be reduced to $I(X_t; X_{t-\mathrm{AoI}(t)})$, which demonstrates a non-linear dependency on AoI under both the Gaussian Markov source and the Binary Markov source [35]. In [36], an analogous approach is utilized to characterize the *value of information* (VoI) for the Ornstein-Uhlenbeck (OU) process, which likewise demonstrates a non-linear dependency on AoI. In [37], the conditional entropy $H(X_t|\mathbf{W}_t)$ is further investigated to measure the uncertainty of the source for a remote estimator given the history received updates $\mathbf{W}_t$. When applied to a Markov Source, this conditional entropy simplifies to $H(X_t|X_{t-\mathrm{AoI}(t)})$, thus exemplifying a non-linear penalty associated with AoI.

To address shortcoming (b), substantial research efforts have been invested in the optimization of the *Mean Squared Error* (MSE), denoted by $(X_t - \hat{X}_t)^2$, with an ultimate objective of constructing a real-time reconstruction remote estimation system [38]–[41]. In [38], a metric termed *effective age* is proposed to minimize the MSE for the remote estimation of a Markov source. In [39] and [40], two Markov sources of interest, the Wiener process and the OU process are investigated to deduce the MSE-optimal sampling policy. Intriguingly, these policies were found to be threshold-based, activating sampling only when the instantaneous MSE exceeds a predefined threshold, otherwise maintaining a state of idleness. The authors in [41] explored the trade-off between MSE and quantization over a noisy channel, and derived the MSE-optimal sampling strategy for the OU process.

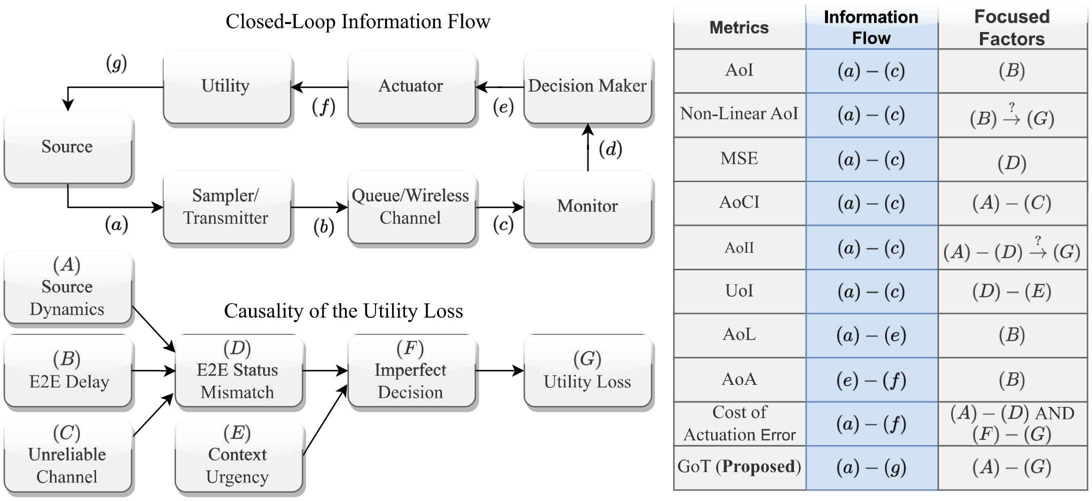
Complementary to the above MSE-centered research, variants of AoI that address shortcoming (b) have also been conceptualized [42]–[45]. In [42], *Age of Changed Information* (AoCI) is proposed to address the ignorance of content dynamics of AoI. In this regard, unchanged statuses do not necessarily provide new information and thus are not prioritized for transmission. In [43], the authors introduce the context-aware weighting coefficient to propose the *Urgency of Information* (UoI), a metric capable of measuring the weighted MSE in diverse urgency contexts. In [44], the authors propose a novel age penalty named *Age of Synchronization* (AoS), a novel metric quantifying the time since the most recent end-to-end synchronization. Moreover, considering that an E2E status mismatch may exert a detrimental effect on the overall system's performance over time, the authors of [45] propose a novel metric called *Age of Incorrect Information* (AoII). This metric quantifies the adverse effect stemming from the duration of the E2E mismatch, revealing that both the degree and duration of E2E semantic mismatches lead to a *utility* reduction for subsequent decision-making.

The above studies focused on the open-loop generation-to-reception process within a transmitter-receiver pair, neglecting the closed-loop perception-actuation timeliness. A notable development addressing this issue is the extension from Up/Down-Link (UL/DL) AoI to a closed-loop AoI metric, referred to as the *Age of Loop* (AoL) [46]. Unlike the traditional open-loop AoI, which diminishes upon successful reception of a unidirectional update, the AoL decreases solely when both the UL status and the DL command are successfully received. Another advanced metric in [47], called Age of Actuation (AoA), also encapsulates the actuation timeliness, which proves pertinent when the receiver employs the received update to execute timely actuation.

Notwithstanding the above advancements, the question on *how the E2E mismatch affects the closed-loop utility of decision-making has yet to be addressed*. To address this issue, [48]–[51] introduce a metric termed CoAE to delve deeper into the cost resulting from the error actuation due to imprecise real-time estimations. Specifically, the CoAE is denoted by an asymmetric zero diagonal matrix $\mathbf{C}$, with each value $C_{X_t,\hat{X}_t} > 0$ representing the instant cost under the E2E mismatch status $(X_t, \hat{X}_t)_{X_t \neq \hat{X}_t}$. In this regard, the authors reveal that the *utility* of decision-making bears a close relation to the E2E semantic mismatch category, as opposed to the mismatch duration (AoII) or mismatch degree (MSE). For example, an E2E semantic mismatch category that "Fire" is detected as "No Fire" will result in high cost; while in the opposite scenario, the cost is low. Nonetheless, we notice that i) the method to obtain a CoAE remains unclear, which implicitly necessitates a pre-established decision-making policy; ii) CoAE does not consider the context-varying factors, which may also affect the decision-making *utility*; iii) the zero diagonal property of the matrix implies the supposition that if $X_t = \hat{X}_t$, then $C_{X_t,\hat{X}_t} = 0$, thereby signifying that error-less actuation necessitates no energy expenditure. Fig. 1 provides an overview of the existing metrics.

Against this background, the present authors have recently proposed a new metric referred to as GoT in [52], which, compared to CoAE CoAE, introduces new dimensions of the context Φ_t and the decision-making policy π_A to describe the true *utility* of decision-making. Remarkably, we find that GoT offers a versatile degeneration to established metrics such as AoI, MSE, UoI, AoII, and CoAE. Additionally, it provides a balanced evaluation of the cost trade-off between the sampling and decision-making. The contributions of this work could be summarized as follows:

- We focus on the decision *utility* issue directly by employing the GoT. A controlled Markov source is observed, wherein the transition of the source is dependent on both the decision-making at the receiver and the contextual situation it is situated. In this case, the decision-making will lead to three aspects in *utility*: i) the future evolution of the source; ii) the instant cost at the source; iii) the energy and resources consumed by actuation.
- We accomplish the goal-oriented *sampler & decision-maker* co-design, which, to the best of our knowledge, represents the first work that addresses the trade-off between semantics-aware sampling and goal-oriented decision-making. Specifically, we formulate this problem as a two-agent infinite-horizon Decentralized Partially Observable Markov Decision Process (Dec-POMDP) problem, with one agent embodying the semantics and context-aware sampler, and the other representing the



Metrics	Information Flow	Focused Factors
AoI	$(a)-(c)$	(B)
Non-Linear AoI	$(a)-(c)$	$(B) \overset{?}{\rightarrow} (G)$
MSE	$(a)-(c)$	(D)
AoCI	$(a)-(c)$	$(A)-(C)$
AoII	$(a)-(c)$	$(A)-(D) \overset{?}{\rightarrow} (G)$
UoI	$(a)-(c)$	$(D)-(E)$
AoL	$(a)-(e)$	(B)
AoA	$(e)-(f)$	(B)
Cost of Actuation Error	$(a)-(f)$	$(A)-(D)$ AND $(F)-(G)$
GoT (**Proposed**)	$(a)-(g)$	$(A)-(G)$

Fig. 1. Interconnections of Age of Information and Beyond in the literature.

goal-oriented decision-maker. Note that the optimal solution of even a finite-horizon Dec-POMDP is known to be NEXP-hard [53], we develop the RVI-Brute-Force-Search Algorithm. This algorithm seeks to derive optimal deterministic joint policies for both sampling and decision-making. A thorough discussion on the optimality of our algorithm is also presented.

- To further mitigate the "*curse of dimensionality*" intricately linked with the execution of the optimal RVI-Brute-Force-Search, we introduce a low-complexity yet efficient algorithm to solve the problem. The algorithm is designed by decoupling the problem into two distinct components: a Markov Decision Process (MDP) problem and a Partially Observable Markov Decision Process (POMDP) issue. Following this separation, the algorithm endeavors to search for the joint Nash Equilibrium between the sampler and the decision-maker, providing a sub-optimal solution to this goal-oriented communication-decision-making co-design.

II. Goal-oriented Tensor

A. Specific Examples of Goal-Oriented Communications

Consider a time-slotted communication system. Let $X_t \in \mathcal{S}$ represent the perceived status of the source at time slot t, and $\hat{X}_t \in \mathcal{S}$ denote the observed status at the receiver end at time slot t. A notable subset of goal-oriented communications is *real-time reconstruction-oriented* communications, which is dedicated to achieving real-time and accurate status reconstruction. This goal can be represented as:

$$\min \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}(X_t-\hat{X}_t)^2\right]. \qquad (1)$$

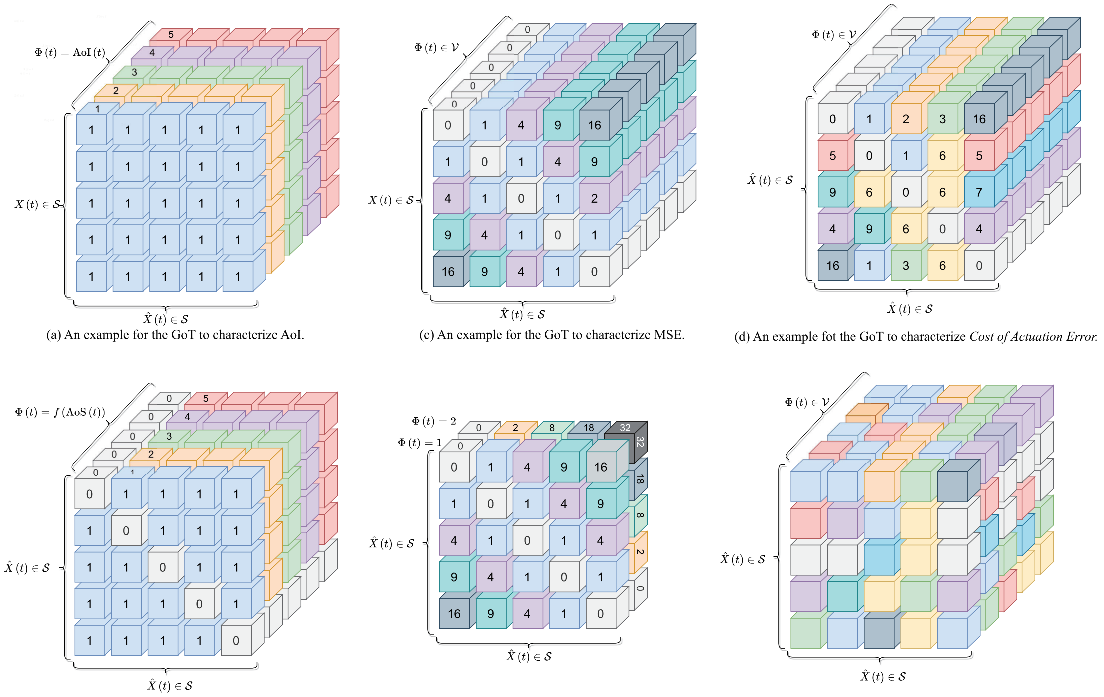
Although *real-time reconstruction* is of significant value, it doesn't encompass the broader purpose of transmission—primarily, to enhance the accuracy of decision-making processes. To this end, the metric CoAE is proposed in [48]–[51] to describe the significance of the error at the actuation point. CoAE, represented by the cost function $C_{X_t,\hat{X}_t}$, quantifies the instantaneous cost associated with decision errors due to mismatch between X_t and $\hat{X}_t$. By minimizing the long-term average CoAE, the effects of mismatches that lead to significant decision-making errors can be mitigated. For instance, in fire monitoring systems, the error of misdetecting a 'Fire' situation as 'No Fire' could have far more severe repercussions than the converse. This asymmetry necessitates a CoAE-optimized policy that prioritizes the minimization of more severe errors, ensuring that the system is particularly sensitive to delivering 'Fire' message. The CoAE-oriented problem is given as:

$$\min \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}C_{X_t,\hat{X}_t}\right], \qquad (2)$$

Following the avenue of CoAE, we notice that the matrix-based metric could be further augmented to tensors to realize more flexible goal characterizations. For example, drawing parallels with the concept of *Urgency of Information* [43], we can introduce a context element Φ_t to incorporate context-aware attributes into this metric.[1] Accordingly, we can define a three-dimensional GoT through a specified mapping: $\mathcal{L}: \mathcal{S}\times\mathcal{V}\times\mathcal{S}\to\mathbb{R}$. This mapping assigns a distinct non-negative real cost to each element within the three-dimension space $\mathcal{S}\times\mathcal{V}\times\mathcal{S}$. In this way, the GoT is defined as a mapping $\mathrm{GoT}(t) \triangleq \mathcal{L}(X_t,\Phi_t,\hat{X}_t)$. The function $\mathcal{L}$ can be graphically represented as a tensor, as illustrated in Fig. 2, hence the term Goal-oriented Tensor is applied. In this regard, the GoT indicates the instantaneous cost of the system at time slot t given the tuple $(X_t,\Phi_t,\hat{X}_t)$. Consequently, the overarching goal of this system is to minimize the long-term average cost:

$$\min \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1}\mathrm{GoT}(t)\right]. \qquad (3)$$

[1] It is important to note that the GoT could be expanded into higher dimensions by integrating additional components, including actuation policies, task-specific attributes, and other factors.


$\Phi(t) = \mathrm{AoI}(t)$
1
2
3
4
5
$X(t) \in \mathcal{S}$
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
$\hat{X}(t) \in \mathcal{S}$
$\Phi(t) \in \mathcal{V}$
0
0
0
0
0
$X(t) \in \mathcal{S}$
0 1 4 9 16
1 0 1 4 9
4 1 0 1 2
9 4 1 0 1
16 9 4 1 0
$\hat{X}(t) \in \mathcal{S}$
$\Phi(t) \in \mathcal{V}$
$\hat{X}(t) \in \mathcal{S}$
0 1 2 3 16
5 0 1 6 5
9 6 0 6 7
4 9 6 0 4
16 1 3 6 0
$\hat{X}(t) \in \mathcal{S}$
$\Phi(t) = f(\mathrm{AoS}(t))$
0 5
0 4
0 3
0 2
0 1
$\hat{X}(t) \in \mathcal{S}$
0 1 1 1 1
1 0 1 1 1
1 1 0 1 1
1 1 1 0 1
1 1 1 1 0
$\hat{X}(t) \in \mathcal{S}$
$\Phi(t) = 2$
$\Phi(t) = 1$
0 2 8 18 32
0
32 18 8 2 0
$\hat{X}(t) \in \mathcal{S}$
0 1 4 9 16
1 0 1 4 9
4 1 0 1 4
9 4 1 0 1
16 9 4 1 0
$\hat{X}(t) \in \mathcal{S}$
$\Phi(t) \in \mathcal{V}$
$\hat{X}(t) \in \mathcal{S}$
$\hat{X}(t) \in \mathcal{S}$

(a) An example for the GoT to characterize AoI.

(c) An example for the GoT to characterize MSE.

(d) An example fot the GoT to characterize *Cost of Actuation Error.*

(b) An example for the GoT to characterize AoII.

(e) An example for the GoT to characterize UoI.

(f) The general GoT.

Fig. 2. GoT visualizations, where distinct colored square signify varying levels of costs or penalties per time slot give a specific tuple $(X_t, \Phi_t, \hat{X}_t)$.

B. *Flexibility of GoT*

In this subsection, we demonstrate, through visualized examples and mathematical formulations, that a three-dimension GoT can degenerate to existing metrics. Fig. 2 showcases a variety of instances of the GoT metric.

- **Degeneration to AoI**: AoI is generally defined as $\mathrm{AoI}(t) \triangleq t - \max\{G_i : D_i < t\}$, where G_i is the generated time stamp of the i-th status update, D_i represents the corresponding deliver time slot. Since AoI is semantics-agnostic [48], [54], the tensor values of the corresponding GoT only depend on the freshness context $\Phi_t \triangleq \mathrm{AoI}(t)$. The GoT can be reduced to

$$\mathrm{GoT}(t) = \mathcal{L}(X_t, \Phi_t, \hat{X}_t) \overset{(a)}{=} \Phi(t) = \mathrm{AoI}(t). \tag{4}$$

where (a) indicates that AoI is semantics-agnostic, *i.e.*, it does not relate to X_t and $\hat{X}_t$. The process of reducing GoT to AoI is visually depicted in Fig. 2(a). From Fig. 2(a), we can notice that the tensor values given a specific Φ_t are constant, signifying that AoI is semantics-agnostic.

- **Degeneration to AoII**: The degeneration of the GoT to the AoII serves as a demonstration of the GoT's adaptability to existing semantic-aware metrics. AoII is defined as $\mathrm{AoII}(t) \triangleq f(\mathrm{AoS}(t)) \cdot g(X_t, \hat{X}_t)$, where $\mathrm{AoS}(t) \triangleq t - \max\left\{\tau : t \leq \tau, X_t = \hat{X}_t\right\}$. AoII is semantics-aware and is hence regarded as an enabler of semantic communications [55]. By setting $\Phi_t \triangleq f(\mathrm{AoS}(t))$, the GoT is succinctly expressed as:

$$\begin{aligned} \mathrm{GoT}(t) &= \mathcal{L}\left(X_t, \Phi_t, \hat{X}_t\right) \overset{(a)}{=} \Phi_t \cdot g\left(X_t, \hat{X}_t\right) \\ &= f(\mathrm{AoS}(t)) \cdot g(X_t, \hat{X}_t) = \mathrm{AoII}(t), \end{aligned} \tag{5}$$

where $g(X_t, \hat{X}_t)$ represents the error function, commonly defined by $\mathbb{1}_{\{X_t \neq \hat{X}_t\}}$. The equation labeled (a) highlights the intrinsic multiplicative nature of AoII, integrating the temporal dimension (AoS) with the error function to yield AoII. The visual representation of the GoT characterizing AoII is depicted in Fig. 2(b). From Fig. 2(b), we note the presence of a foundational layer within the tensor representation. This foundational layer serves as the cornerstone for subsequent layers, which are generated by the multiplication of Φ_t across this foundational layer.

- **Degeneration to MSE**: The GoT can also be reduced to the MSE, a fundamental metric in the reconstruction-oriented communications. MSE is defined as $\mathrm{MSE}(t) \triangleq \left(X_t - \hat{X}_t\right)^2$, which is intuitively context-agnostic. This reduction is evident in scenarios where the error function $g(X_t, \hat{X}_t)$ aligns with the MSE formulation, i.e., $g(X_t, \hat{X}_t) = (X_t - \hat{X}_t)^2$. Under these conditions, the GoT simplifies to the MSE as follows:

$$\begin{aligned} \mathrm{GoT}(t) &= \mathcal{L}\left(X_t, \Phi_t, \hat{X}_t\right) \overset{(a)}{=} g\left(X_t, \hat{X}_t\right) \\ &= \left(X_t - \hat{X}_t\right)^2 = \mathrm{MSE}(t). \end{aligned} \tag{6}$$

The equation labeled (a) establishes due to the context-agnostic nature of MSE. The visualization of the GoT reducing to MSE is shown in Fig. 2(c).

- **Degeneration to UoI**: The degeneration of the GoT to the AoII serves as a demonstration of the GoT's adaptability to existing context and semantics-aware metrics. UoI is defined by $\mathrm{UoI}(t) \triangleq \Phi_t \cdot \left(X_t - \hat{X}_t\right)^2$, where the context-aware weighting coefficient Φ_t is further introduced [43]. When

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2024.3416864

$g(X_t, \hat{X}_t) = (X_t - \hat{X}_t)^2$, the GoT could be transformed into the UoI by

$$\begin{aligned} \mathrm{GoT}(t) &= \mathcal{L}\left(X_t, \Phi_t, \hat{X}_t\right) \overset{(a)}{=} \Phi_t \cdot g\left(X_t, \hat{X}_t\right) \\ &= \Phi_t \cdot \left(X_t - \hat{X}_t\right)^2 = \mathrm{UoI}(t), \end{aligned} \tag{7}$$

where (a) indicates the inherent multiplicative characteristic of UoI. The visualization of the GoT reduction to UoI is shown in Fig. 2(d). From Fig. 2(d), we can also note the presence of a foundational layer within the tensor. This foundational layer serves as the cornerstone for other layers, which are generated by the multiplication of Φ_t across this foundational layer.

- **Degeneration to CoAE**: GoT can also be reduced to CoAE CoAE is defined by $C_{X_t,\hat{X}_t}$, which indicates the instantaneous system cost if the source status is X_t and the estimated one $\hat{X}_t$ mismatch [50]. Let $g(X_t, \hat{X}_t) = C_{X_t,\hat{X}_t}$, the GoT collapses to CoAE:

$$\mathrm{GoT}(t) = \mathcal{L}\left(X_t, \Phi_t, \hat{X}_t\right) \overset{(a)}{=} g\left(X_t, \hat{X}_t\right) = C_{X_t,\hat{X}_t}, \tag{8}$$

where (a) establishes due to the context-agnostic nature of CoAE. This reduction is visually represented in Fig. 2(e), highlighting the context-independent nature of CoAE within the GoT framework; that is, the value of the tensor remains constant regardless of Φ_t.

C. Goal Characterization Through GoT

In this subsection, we further demonstrate The GoT framework extends beyond flexibility of characterizing established metrics; it also offers its utility in characterizing generalized communication goals that are ultimately achieved by decision making. Specifically, as shown in Fig. 2 (f), within a broader GoT framework, the cost associated with semantics mismatches under a specific context can be assigned with any non-negative value, which is dependent on the decision-making policy designed for achieving a certain goal.

To characterize a goal through GoT, here we propose a method that formulates GoT by meticulously considering both the scenario at hand and the intended goal. Specific examples are also give in this subsection.

1) Steps to Formulate a GoT to Characterize the Goal

- **Step 1:** Clarify the scenario and the communication goal.
- **Step 2:** Define the sets of semantic status, $\mathcal{S}$, and contextual status, $\mathcal{V}$. These sets can be modeled as collections of discrete components. Here $X_t \in \mathcal{S}$ and $\Phi_t \in \mathcal{V}$.
- **Step 3:** The GoT could be decoupled by three factors: [2]

i) The status inherent cost $C_1(X_t, \Phi_t)$. It quantifies the cost associated with different status pairs (X_t, Φ_t) in the absence of external influences;
ii) The actuation gain $C_2(\pi_A(\hat{X}_t))$, where π_A is a deterministic decision policy contingent upon $\hat{X}_t$. This cost quantifies the positive *utility* resulting from the actuation $\pi_A(\hat{X}(t))$;
iii) The actuation resource consumption $C_3(\pi_A(\hat{X}_t))$, which reflects the resources consumed by implementing $\pi_A(\hat{X}(t))$.

- **Step 4:** Constructing the GoT. The GoT that characterizes the goal is calculated by

$$\mathrm{GoT}^{\pi_A}(t) = \Big[C_1(X_t, \Phi_t) - aC_2(\pi_A(\hat{X}_t))\Big]^+ + bC_3(\pi_A(\hat{X}_t)), \tag{9}$$

where a and b serve as scaling factors or unit conversion constants. Their role is to adjust the magnitudes of the terms involved, ensuring *dimensional consistency* across the equation. For example, when $C_1(X_t, \Phi_t)$ represents energy consumption with its own units and $C_2(\pi_A(\hat{X}_t))$ denotes financial cost in a different unit system, the constant a harmonizes these diverse units. The ramp function $[\cdot]^+$ ensures that any actuation $\pi_A(\hat{X}_t)$ reduces the cost to a maximum of 0. To further illustrate the above steps, we provide an example in the following. For more intricate applications, the GoT formulation can be similarly derived as outlined in this subsection.

2) A Specific Example

As a case study, we consider a forest fire remote monitoring and rescue system. In such a system, the goal is to execute timely and appropriate rescue operation so that the cost resulted from fire and rescue resource consumption are minimized in long term. We assume that the sensors monitor the forest situation and perceived the semantic status $X_t \in \mathcal{S}$, indicating the fire occurrence at time slot t. Here, the semantics set is defined as $\mathcal{S} = \{\text{Big Fire}, \text{Middle Fire}, \text{No Fire}\}$. The weather (or context) around the monitored area at time slot t is defined as $\Phi_t \in \mathcal{V}$, where $\mathcal{V} = \{\text{Rainy}, \text{Sunny}\}$.

The decision maker should make decisions based on the observed status $\hat{X}_t$. The decision policy is defined as $\pi_A(\hat{X}_t)$, where $\hat{X}_t \in \mathcal{S}$ is the observed status at the receiver at time slot t, and the $\pi_A(\cdot) : \mathcal{S} \rightarrow \mathcal{A}$ generates decisions $a_A(t) \in \mathcal{A}$ based on $\hat{X}_t$. We consider the decision making space as $\mathcal{A}_A = \{a_0, \cdots, a_{10}\}$. The variables a_i, where $0 \le i \le 10$, denote distinct levels of rescue operations. Each operation level varies in the amount of resources it consumes and the degree of rescue gain it yields. As an example, we define the decision making policy as $\pi_A(\text{Big Fire}) = a_7$, $\pi_A(\text{Fire}) = a_3$, and $\pi_A(\text{No Fire}) = a_0$.

The GoT aims to characterize the instantaneous cost resulted by different types of semantics mismatch and decision making policies in real-time scenarios. Specifically, we consider a “Big Fire” status incurs a variable cost depending on weather conditions: 50 units per time slot on a “Sunny” day and 20 units during “Rainy” conditions. Meanwhile, “Fire” status incurs 20 units per time slot on a “Sunny” day and 10 units during “Rainy” conditions. In contrast, a “No Fire” status does not generate any cost under either weather scenario. Furthermore, we can consider $C_2(\pi_A(\hat{X}_t))$ and $C_3(\pi_A(\hat{X}_t))$ are both linear to the actuation with $C_2(\pi_A(\hat{X}_t)) = C_g \cdot \mathtt{Index}(\pi_A(\hat{X}_t))$ and $C_3(\pi_A(\hat{X}_t)) = C_I \cdot \mathtt{Index}(\pi_A(\hat{X}_t))$, where $\mathtt{Index}(a_i) = i$. This signifies that the action a_i will consume $i \cdot C_I$ unit resources per time slot and will brings about at most C_g gain per time slot. These cost parameters are established based on historical data and past experiences. From the above discussion, we have the costs as:

[2] The definitions of costs are diverse, covering aspects such as financial loss, energy consumption, depending on the specific goal they are designed to address. Different types of costs are scaled through the scaling factors a and b in (9)

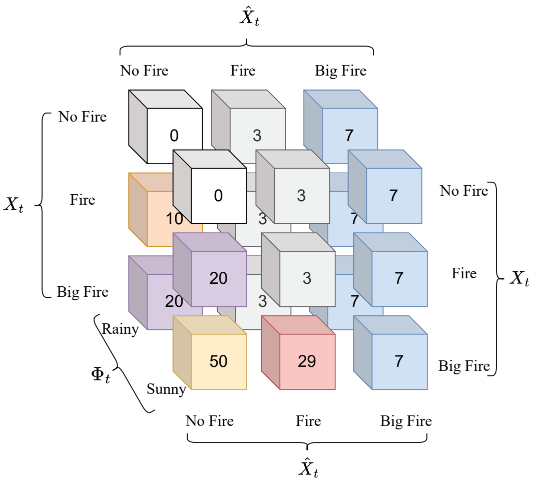


Fig. 3. GoT to characterize the goal in forest fire remote monitoring and rescue system: an example. For instance, the red square with value 29 represents the cost per time slot when $(X_t, \Phi_t, \hat{X}_t) = $ (Big Fire, Sunny, Fire).

i) Status inherent cost: (Unit cost per time slot)

$$\begin{aligned} C_1(\text{Big Fire}, \text{Sunny}) = 50, C_1(\text{Big Fire}, \text{Rainy}) = 20, \\ C_1(\text{Fire}, \text{Sunny}) = 20, C_1(\text{Fire}, \text{Rainy}) = 10, \\ C_1(\text{No Fire}, \text{Sunny}) = 0, C_1(\text{No Fire}, \text{Rainy}) = 0. \end{aligned} \tag{10}$$

i) Actuation gain: (Unit cost per time slot)

$$\begin{aligned} C_2(\pi_A(\text{Big Fire})) &= C_2(a_7) = 7C_g, \\ C_2(\pi_A(\text{Fire})) &= C_2(a_3) = 3C_g, \\ C_2(\pi_A(\text{No Fire})) &= C_2(a_0) = 0. \end{aligned} \tag{11}$$

iii) Actuation resource consumption: (Unit cost per time slot)

$$\begin{aligned} C_3(\pi_A(\text{Big Fire})) &= C_2(a_7) = 7C_I, \\ C_3(\pi_A(\text{Fire})) &= C_2(a_3) = 3C_I, \\ C_3(\pi_A(\text{No Fire})) &= C_2(a_0) = 0. \end{aligned} \tag{12}$$

Then, by leveraging (9), we obtain the GoT as visualized in Fig. 3. Here we assume that the units in (10)-(12) are the same, and thus the scaling factors satisfy $a = b = 1$. In addition, we set $C_g = 8$ and $C_I = 1$ as an example to visualize the tensor.

III. System Model

This section aims to explore strategies for leveraging an established GoT to achieve effective goal-oriented semantic communications. We aim at designing efficient algorithms that recognize the significance of the semantics in goal achievement. As shown in Fig. 4, we consider a time-slotted perception-actuation loop , where the semantics of the source during the time slot t, denoted by $X_t \in \mathcal{S} = \left\{s_1, \cdots, s_{|\mathcal{S}|}\right\}$, and the context around the source during the time slot t, denoted by $\Phi_t \in \mathcal{V} = \left\{v_1, \cdots, v_{|\mathcal{V}|}\right\}$ are fed into a semantic sampler, tasked with determining the significance of the present status X_t and subsequently deciding if it warrants transmission via an unreliable channel. Here, X_t and Φ_t belongs to the semantic set $\mathcal{S}$ and the context set $\mathcal{V}$, respectively.

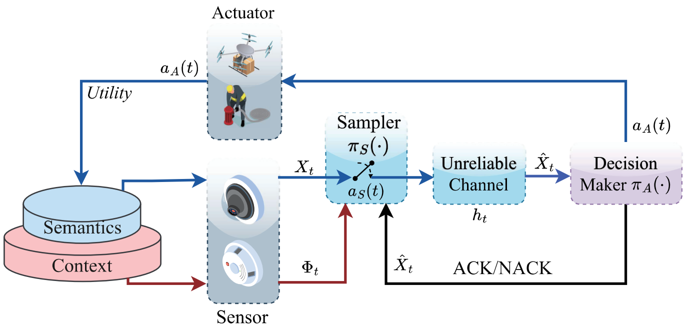


Fig. 4. Illustration of our considered system where transmitted semantic status arrives at a receiver for decision-making to achieve a certain goal.

The semantics and context are extracted and assumed to perfectly describe the status of the observed process. 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 the value 1 representing the execution of sampling and transmission at the beginning of the time slot, and the value 0 indicating the idleness of the sampler. π_S here represents the sampling policy. We consider a perfect and delay-free feedback channel [48]–[51], with ACK representing a successful transmission and NACK representing the otherwise. The decision-maker at the receiver will make decisions $a_A(t) \in \mathcal{A}_A = \left\{a_1, \cdots, a_{|\mathcal{A}_A|}\right\}$ based on the most recently received semantic status at time slot t, denoted by $\hat{X}_t$[3]. The decision making process will ultimately affect the goal-achieving *utility* of the system.

A. Semantics and Context Dynamics

Thus far, a plethora of studies have delved into the analysis of various discrete Markov sources, encompassing Birth-Death Markov Processes elucidated in [51], binary Markov sources elucidated in [56], and etc. In real-world situations, the context and the actuation also affect the source's evolution. For instance, a rescue actuation will affect the evolution of the fire situation; a control command will affect the position of a vehicle. In this paper, we consider a context-dependent controlled Discrete Markov source:

$$\Pr\left(X_{t+1} = s_u \left| X_t = s_i, a_A(t) = a_m, \Phi_t = v_k \right.\right) = p_{i,u}^{(k,m)}, \tag{13}$$

where the dynamics of the source is dependent on both the decision-making $a_A(t)$ and context Φ_t. Furthermore, we take into account the variations in context Φ_t, characterized by:

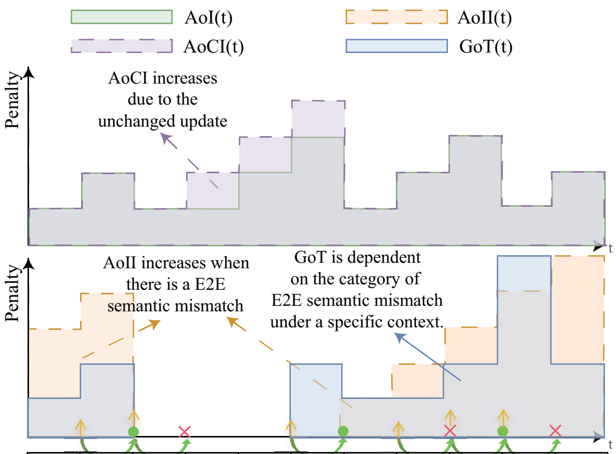
$$\Pr\left(\Phi_{t+1} = v_r \left| \Phi_t = v_k \right.\right) = p_{k,r}. \tag{14}$$

Note that the semantic status X_t and context status Φ_t could be tailored according to the specific application scenario. In general, these two processes are independent with each other.

B. Unreliable Channel

We assume that each transmission will incur a transmission delay of 1 time slot and posit that the channel realizations are independent and identically distributed (i.i.d.) across time

[3] We consider a general abstract decision-making set $\mathcal{A}_A$ that exhibits adaptability to diverse applications. Notably, this decision-making set can be customized and tailored to suit specific requirements. Examples have been discussed in Section II.C



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	1	0	\
X_t	NF	NF	NF	NF	NF	F	NF	NF	NF	F	F
$\hat{X}_t$	F	F	NF	NF	NF	NF	F	F	F	NF	NF
Φ_t	R	S	S	S	R	R	R	R	S	S	R
GoT(t)	1	2	0	0	0	2	1	1	2	5	2

Fig. 5. An illustration of AoI, AoCI, AoII, and GoT in a time-slotted status update system. Here, the value of GoT is obtained from Fig. 3. “NF” represents “No Fire”, “F” indicates “Fire”, “R” represents “Rainy”, and “S” represents “Sunny”. The yellow arrow represents the sampling action. The green circle represents the successful delivery of an update. The red cross indicates a failed update.

slots and are independent of X_t and Φ_t, following a Bernoulli distribution. Particularly, the channel realization h_t assumes a value of 1 in the event of successful transmission, and 0 otherwise. Accordingly, we define the probability of successful transmission as $\Pr\left(h_t=1\right)=p_S$ and the failure probability as $\Pr\left(h_t=0\right)=1-p_S$. To characterize the dynamics of the observation $\hat{X}_t$, we consider two cases as described below:

- $a_S(t)=0$. In this case, the sampler and transmitter remain idle at time slot t, manifesting that there is no new knowledge delivered to the receiver at time slot $t{+}1$. Thus, the observation at the receiver at time slot $t+1$ remains remains the same as that at time slot t, and we have $\hat{X}_{t+1}=\hat{X}_t$. As such, we have:

$$\Pr\left(\hat{X}_{t+1}=x\,\middle|\,\hat{X}_t=s_j,a_S(t)=0\right)=\mathbb{1}_{\{x=s_j\}}. \tag{15}$$

- $a_S(t)=1$. 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 observed status at the receiver $\hat{X}_{t+1}$ remains $X(t)$, and the transition probability is

$$\Pr\left(\hat{X}_{t+1}=x\,\middle|\,\hat{X}_t=s_j,X_t=s_i,a_S(t)=1,h_t=1\right)=\mathbb{1}_{\{x=s_i\}}. \tag{16}$$

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

$$\Pr\left(\hat{X}_{t+1}=x\,\middle|\,\hat{X}_t=s_j,X_t=s_i,a_S(t)=1,h_t=0\right)=\mathbb{1}_{\{x=s_j\}}. \tag{17}$$

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

$$\begin{aligned}&\Pr\left(\hat{X}_{t+1}=x\,\middle|\,\hat{X}_t=s_j,X_t=s_i,a_S(t)=1\right)\\&=\sum_{h_t}p\left(h_t\right)\Pr\left(\hat{X}_{t+1}=x\,\middle|\,\hat{X}_t=s_j,X_t=s_i,a_S(t)=1,h_t\right)\\&=p_S\cdot\mathbb{1}_{\{x=s_i\}}+\left(1-p_S\right)\cdot\mathbb{1}_{\{x=s_j\}}.\end{aligned} \tag{18}$$

Combing (15) with (18) yields the dynamics of the observed status $\hat{X}_t$.

C. Goal-oriented decision-making and Actuating

We note that the previous works primarily focused on minimizing the open-loop freshness-related or error-related penalty for a transmitter-receiver system. Nevertheless, irrespective of the *fresh* delivery or accurate end-to-end timely reconstruction, the ultimate goal of such optimization efforts is to ensure precise and effective decision-making. To this end, we broaden the open-loop transmitter-receiver information flow to include a perception-actuation closed-loop *utility* flow by incorporating the decision-making and actuation processes. As a result, decision-making and actuation enable the conversion of status updates into ultimate effectiveness. Here the decision-making at time slot t follows that $a_A(t)=\pi_A(\hat{X}_t)$, with π_A representing the deterministic decision-making policy.

IV. Problem Formulation and Solution

Traditionally, the development of sampling strategies has been designed separately from the decision-making process. An archetypal illustration of this two-stage methodology involves first determining the optimal sampling policy based on AoI, MSE, and their derivatives, such as AoII, and subsequently accomplishing goal-oriented decision-making. This two-stage separate design arises from the inherent limitation of existing metrics that they fail to capture the closed-loop decision *utility*. Nevertheless, the metric GoT empowers us to undertake a co-design of sampling and decision-making.

We adopt the *team decision theory*, wherein 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 $\text{GoT}^{\pi_A}(t)+C_S\cdot a_S(t)$, and the problem is characterized as:

$$\mathcal{P}1:\quad\min_{\boldsymbol{\pi}_C\in\Upsilon}\ \limsup_{T\to\infty}\frac{1}{T}\mathbb{E}^{\boldsymbol{\pi}_C}\left(\sum_{t=0}^{T-1}\text{GoT}^{\pi_A}(t)+C_S\cdot a_S(t)\right), \tag{19}$$

where $\boldsymbol{\pi}_C=(\pi_S,\pi_A)$ denotes the joint sampling and decision-making 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. Note that $\text{GoT}^{\pi_A}(t)$ is characterized by (9).

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2024.3416864

A. Dec-POMDP Formulation

The problem in (19) aims to find the optimal decentralized policy $\boldsymbol{\pi}_C$ so that the long-term average cost of the system is minimized. To solve the problem $\mathcal{P}1$, we ought to formulate a Dec-POMDP problem, which is initially introduced in [53] to solve the cooperative sequential decision issues for distributed multi-agents. Unlike MDP and POMDP frameworks, which concentrate on the decision-making processes of a single centralized agent, Dec-POMDP involves a group of agents working collaboratively towards a shared goal, depending entirely on their individual, localized, and partially observable knowledge. In our context, the semantics-aware sampler and the goal-oriented decision-maker work in a naturally decentralized manner to achieve a shared goal, and we thus formulate it as a Dec-POMDP problem. A typical Dec-POMDP is 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 this instance, we have $n = 2$, signifying the presence of two agents within this context: one agent $\mathcal{A}gent_S$ embodies the semantics-context-aware sampler and transmitter, while the other represents the $\hat{X}_t$-dependent decision-maker, denoted by $\mathcal{A}gent_A$.
- $\mathcal{I}$ is the finite set of the global system status, 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 squeal.
- $\mathcal{T}$ is the transition function defined by

$$\mathcal{T}(\mathbf{w}, \mathbf{a}, \mathbf{w}') \triangleq \Pr(\mathbf{W}_{t+1} = \mathbf{w}' | \mathbf{W}_t = \mathbf{w}, \mathbf{a}_t = \mathbf{a}), \tag{20}$$

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. Then, by taking into account the *conditional independence* among X_{t+1}, Φ_{t+1}, and $\hat{X}_{t+1}$, given $(X_t, \Phi_t, \hat{X}_t)$ and $\mathbf{a}(t)$, the transition functions can be calculated in lemma 1.

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

$$\begin{aligned} &p\left((s_u, x, v_r)\,|(s_i, s_j, v_k), (1, a_m)\right) = \\ &p_{i,u}^{(k,m)} \cdot p_{k,r} \cdot \left(p_S \cdot \mathbb{1}_{\{x=s_i\}} + (1-p_S) \cdot \mathbb{1}_{\{x=s_j\}}\right), \end{aligned} \tag{21}$$

$$p\left((s_u, x, v_r)\,|(s_i, s_j, v_k), (0, a_m)\right) = p_{i,u}^{(k,m)} \cdot p_{k,r} \cdot \mathbb{1}_{\{x=s_j\}}, \tag{22}$$

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. The transition function can be derived by incorporating the dynamics in equations (13), (14), (15), and (18). A more comprehensive proof is presented in Appendix A. □

- $\mathcal{A} = \mathcal{A}_S \times \mathcal{A}_A$, with $\mathcal{A}_S \triangleq \{0, 1\}$ representing the set of binary sampling actions executed by the sampler, and $\mathcal{A}_A \triangleq \{a_0, \cdots, a_{M-1}\}$ representing the set of decision actions undertaken by the actuator.
- $\Omega = \Omega_S \times \Omega_A$ constitutes a finite set of joint observations. Generally, the observation made by a single agent regarding the system status is partially observable. 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 actuator (or 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:

$$\begin{aligned} \mathcal{O}(\mathbf{w}, \mathbf{o}) &\triangleq \Pr(\mathbf{o}_t = \mathbf{o} | \mathbf{W}_t = \mathbf{w}), \\ \mathcal{O}_S(\mathbf{w}, o_S) &\triangleq \Pr(o_S^{(t)} = o_S | \mathbf{W}_t = \mathbf{w}), \\ \mathcal{O}_A(\mathbf{w}, o_A) &\triangleq \Pr(o_A^{(t)} = o_A | \mathbf{W}_t = \mathbf{w}). \end{aligned} \tag{23}$$

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:*

$$\begin{aligned} p_S\left((s_u, s_r, v_q)|(s_i, s_j, v_k)\right) &= \mathbb{1}_{\{(s_u, s_r, v_q) = (s_i, s_j, v_k)\}} \\ p_A\left(s_z|(s_i, s_j, v_k)\right) &= \mathbb{1}_{\{s_z = s_j\}}. \end{aligned} \tag{24}$$

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 as 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 $\min\limits_{\boldsymbol{\pi}_C \in \Upsilon} \limsup\limits_{T \to \infty} \frac{1}{T} \mathbb{E}^{\boldsymbol{\pi}_C} \left(-\sum\limits_{t=0}^{T-1} r(t) \right)$. Subsequently, to establish congruence with the problem in (19), the reward function is defined as:

$$r(t) = \mathcal{R}^{\pi_A}(\mathbf{w}, a_S) = -\mathrm{GoT}^{\pi_A}(t) - C_S \cdot a_S(t). \tag{25}$$

B. Solutions to the Infinite-Horizon Dec-POMDP

In general, solving a Dec-POMDP is known to be NEXP-complete for the finite-horizon setup [53], signifying that it necessitates formulating a conjecture about the solution non-deterministically, while each validation of a conjecture demands exponential time. For an infinite-horizon Dec-POMDP problem, finding an optimal policy for a Dec-POMDP problem is known to be undecidable. Nevertheless, within our considered model, both the sampling and decision-making processes are deterministic, given as $a_S(t) = \pi_S(\mathbf{w})$ and $a_A(t) = \pi_A(o_A)$. In such a circumstance, it is feasible to determine a joint optimal deterministic policy via Brute-Search-across the decision-making policy space.

1) Optimal Solution

The idea is based on the finding that, given a deterministic decision-making policy π_A, the sampling problem can be formulated as a standard fully observed MDP problem denoted by $\mathscr{M}_{\text{MDP}}^{\pi_A} \triangleq \langle \mathcal{I}, \mathcal{T}^{\pi_A}, \mathcal{A}_S, \mathcal{R}\rangle$.

Definition 1. *Given a deterministic decision-making policy π_A, the optimal sampling problem could be formulated by a typical fully observed MDP problem $\mathscr{M}_{\text{MDP}}^{\pi_A} \triangleq \langle \mathcal{I}, \mathcal{A}_S, \mathcal{T}_{\text{MDP}}^{\pi_A}, \mathcal{R}\rangle$, where the elements are given as follows:*

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2024.3416864

- *$\mathcal{I}$: 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-making policy π_A, which is*

$$\begin{aligned}\mathcal{T}^{\pi_A}(\mathbf{w}, a_S, \mathbf{w}') &= p^{\pi_A}\left(\mathbf{w}'|\mathbf{w}, a_S\right) \\ &= \sum_{o_A \in \mathcal{O}_A} p\left(\mathbf{w}'|\mathbf{w}, (a_S, \pi_A(o_A))\right) p_A(o_A|\mathbf{w}),\end{aligned} \tag{26}$$

  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 $\mathscr{M}^{\pi_A}_{\text{MDP}}$, which is characterized by a tuple $\langle \mathcal{I}, \mathcal{T}^{\pi_A}, \mathcal{A}_S, \mathcal{R}\rangle$. In order to deduce the optimal sampling policy under a deterministic decision-making policy π_A, it is imperative to resolve the Bellman equations [57]:

$$\theta^*_{\pi_A} + V_{\pi_A}(\mathbf{w}) = \max_{a_S \in \mathcal{A}_S}\left\{\mathcal{R}^{\pi_A}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}'|\mathbf{w}, a_S) V_{\pi_A}(\mathbf{w}')\right\}, \tag{27}$$

where $V^{\pi_A}(\mathbf{w})$ is the value function and $\theta^*_{\pi_A}$ is the optimal long-term average reward given the decision-making 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 π_A

```
Input: The MDP tuple $\langle \mathcal{I}, \mathcal{A}_S, \mathcal{T}^{\pi_A}, \mathcal{R}\rangle$, $\epsilon$, $\pi_A$;
Initialization: $\forall \mathbf{w} \in \mathcal{I}$, $\tilde{V}^0_{\pi_A}(\mathbf{w}) = 0$, $\tilde{V}^{-1}_{\pi_A}(\mathbf{w}) = \infty$, $k = 0$ ;
Choose $\mathbf{w}^{ref}$ arbitrarily;
while $||\tilde{V}^k_{\pi_A}(\mathbf{w}) - \tilde{V}^{k-1}_{\pi_A}(\mathbf{w})|| \geq \epsilon$ do
    $k = k + 1$;
    $g_k = \max_{a_S}\left\{\mathcal{R}(\mathbf{w}^{ref}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}'|\mathbf{w}^{ref}, a_S)\tilde{V}^{k-1}_{\pi_A}(\mathbf{w}')\right\}$;
    for $\mathbf{w} \in \mathcal{I} - \mathbf{w}^{ref}$ do
        $\tilde{V}^k_{\pi_A}(\mathbf{w}) = -g_k + \max_{a_S}\left\{\mathcal{R}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I} - \mathbf{w}^{ref}} p(\mathbf{w}'|\mathbf{w}, a_S)\tilde{V}^{k-1}_{\pi_A}(\mathbf{w}')\right\}$;
$\theta^*(\pi_A, \pi^*_S) = -\tilde{V}^k_{\pi_A}(\mathbf{w}) + \max_{a_S \in \mathcal{A}_S}\left\{\mathcal{R}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}'|\mathbf{w}, a_S)\tilde{V}^k_{\pi_A}(\mathbf{w}')\right\}$;
for $\mathbf{w} \in \mathcal{I}$ do
    $\pi^*_S(\pi_A, \mathbf{w}) = \arg\max_{a_S}\left\{\mathcal{R}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}'|\mathbf{w}, a_S)\tilde{V}^k_{\pi_A}(\mathbf{w}')\right\}$;
Output: $\pi^*_S(\pi_A)$, $\theta^*(\pi_A, \pi^*_S)$
```

With Definition 1 and Algorithm 1 in hand, we could then perform a Brute-Force-Search across the decision-making policy space Υ_A, thereby acquiring the joint sampling-decision-making policy. The algorithm is called RVI-Brute-Force-Search Algorithm, which is elaborated in Algorithm 2. In the following theorem, we discuss the optimality of the RVI-Brute-Force-Search Algorithm.

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 unichan.*

Proof. If the the transition function $\mathcal{T}^{\pi_A}$ follows a unichian, we obtain from [58, 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

$$\theta^*(\pi^*_A, \pi^*_S) \leq \theta^*(\pi_A, \pi^*_S) \leq \theta^*(\pi_A, \pi_S). \tag{28}$$

Nonetheless, the Brute-Force-Search across the decision-making policy space remains computationally expensive, as the size of the decision-making policy space Υ_A amounts to $|\Upsilon_A| = \mathcal{A}_A^{\mathcal{O}_A}$. This implies that executing the RVI algorithm $\mathcal{A}_A^{\mathcal{O}_A}$ times is necessary to attain the optimal policy. Consequently, although proven to be optimal, such an algorithm is ill-suited for scenarios where $\mathcal{O}_A$ and $\mathcal{A}_A$ are considerably large. To ameliorate this challenge, we propose a sub-optimal, yet computation-efficient alternative in the subsequent section.

2) A Sub-optimal Solution

The method here is to instead find a locally optimal algorithm to circumvent the high complexity of the Brute-Force-Search-based approach. We apply the Joint Equilibrium-Based Search for Policies (JESP) for *Nash equilibrium* solutions [59]. Within this framework, the sampling policy is optimally responsive to the decision-making policy and vice versa, *i.e.*, $\forall \pi_S, \pi_A, \theta(\pi^*_S, \pi^*_A) \leq \theta(\pi_S, \pi^*_A), \theta(\pi^*_S, \pi^*_A) \leq \theta(\pi^*_S, \pi_A)$.

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

```
Input: The Dec-POMDP tuple
    $\mathscr{M}_{DEC-POMDP} \triangleq \langle n, \mathcal{I}, \mathcal{A}, \mathcal{T}, \Omega, \mathcal{O}, \mathcal{R}\rangle$;
for $\pi_A \in \Upsilon_A$ do
    Formulate the MDP problem
      $\mathscr{M}^{\pi_A}_{\text{MDP}} \triangleq \langle \mathcal{I}, \mathcal{A}_S, \mathcal{T}^{\pi_A}_{\text{MDP}}, \mathcal{R}\rangle$ as given in Definition 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: $\pi^*_S$, $\pi^*_A$
```

To search for the *Nash equilibrium*, we first search for the optimal sampling policy prescribed a decision-making policy. This problem can be formulated as a standard fully observed MDP problem denoted by $\mathscr{M}^{\pi_A}_{\text{MDP}} \triangleq \langle \mathcal{I}, \mathcal{A}_S, \mathcal{T}^{\pi_A}_{\text{MDP}}, \mathcal{R}\rangle$ (see Definition 1). Next, we alternatively fix the sampling policy π_S and solve for the optimal decision-making policy π_A. This problem can be modeled as a memoryless partially observable Markov decision process (POMDP), denoted by $\mathscr{M}^{\pi_S}_{\text{POMDP}} \triangleq \langle \mathcal{I}, \mathcal{A}_A, \Omega_A, \mathcal{O}_A, \mathcal{T}^{\pi_S}_{\text{POMDP}}, \mathcal{R}\rangle$ (see Definition 2). Then, by alternatively iterating between $\mathcal{A}gent_S$ and $\mathcal{A}gent_A$, we could obtain the *Nash equilibrium* between the two agents.

Definition 2. *Given a deterministic sampling policy π_S, the optimal sampling problem could be formulated as a memoryless POMDP problem $\mathscr{M}^{\pi_S}_{\text{POMDP}} \triangleq \langle \mathcal{I}, \mathcal{A}_A, \Omega_A, \mathcal{O}_A, \mathcal{T}^{\pi_S}_{\text{POMDP}}, \mathcal{R}\rangle$, where the elements are given as follows:*

- *$\mathcal{I}$, Ω_A, $\mathcal{A}_A$, and $\mathcal{O}_A$: the same as the pre-defined Dec-POMDP tuple.*

- $\mathcal{T}_{\text{POMDP}}^{\pi_S}$: *the transition function given a deterministic sampling policy* π_S*, which is*

$$\begin{aligned}\mathcal{T}_{\text{POMDP}}^{\pi_S}(\mathbf{w}, a_A, \mathbf{w}') &= p^{\pi_S}\left(\mathbf{w}'|\mathbf{w}, a_A\right) \\ &= p\left(\mathbf{w}'|\mathbf{w}, (\pi_S(\mathbf{w}), a_A)\right),\end{aligned} \tag{29}$$

where $p\left(\mathbf{w}'|\mathbf{w}, (\pi_S(\mathbf{w}), a_A)\right)$ *could be obtained by Lemma 1.*

- $\mathcal{R}$: *the reward function is denoted as* $\mathcal{R}^{\pi_S}(\mathbf{w}, a_A)$*, which could be obtained by (9).*

We then proceed to solve the memoryless POMDP problem discussed in Definition 2 to obtain the deterministic decision-making policy. Denote $p_{\pi_A}^{\pi_S}(\mathbf{w}'|\mathbf{w})$ as the transition probability $\Pr\left\{\mathbf{W}_{t+1} = \mathbf{w}'|\mathbf{W}_t = \mathbf{w}\right\}$ under the sampling policy π_A and π_S, we then have that

$$p_{\pi_A}^{\pi_S}(\mathbf{w}'|\mathbf{w}) = \\ \sum_{o_A \in \mathcal{O}_A} p(o_A|\mathbf{w}) \sum_{a_A \in \mathcal{A}_A} p^{\pi_S}(\mathbf{w}'|\mathbf{w}, a_A)\pi_A(a_A|o_A), \tag{30}$$

where $p(o_A|\mathbf{w})$ could be obtained by Lemma 1 and $p^{\pi_S}(\mathbf{w}'|\mathbf{w}, \pi_A(o_A))$ is obtained by (29). By assuming the ergodicity of the $p_{\pi_A}^{\pi_S}(\mathbf{w}'|\mathbf{w})$ and rewrite it as a matrix $\mathbf{P}_{\pi_A}^{\pi_S}$, we could then solve out the stationary distribution of the system status under the policies π_S and π_A, denoted as $\boldsymbol{\mu}_{\pi_A}^{\pi_S}$, by solving the balance equations:

$$\boldsymbol{\mu}_{\pi_A}^{\pi_S}\mathbf{P}_{\pi_A}^{\pi_S} = \boldsymbol{\mu}_{\pi_A}^{\pi_S},\ \boldsymbol{\mu}_{\pi_A}^{\pi_S}\mathbf{e} = 1, \tag{31}$$

where $\mathbf{e}$ is the all one vector $[1, \cdots, 1]_{|\mathcal{I}|\times 1}$, $\boldsymbol{\mu}_{\pi_A}^{\pi_S}$ could be solved out by Cramer's rules. Denote $\mu_{\pi_A}^{\pi_S}(\mathbf{w})$ as the stationary distribution of $\mathbf{w}$. Also, we denote $r_{\pi_A}^{\pi_S}(\mathbf{w})$ as the expectation reward of global system status $\mathbf{w}$ under policies π_A and π_S. It could be calculated as:

$$r_{\pi_A}^{\pi_S}(\mathbf{w}) = \sum_{o_A \in \mathcal{O}_A} p(o_A|\mathbf{w})\mathcal{R}^{\pi_S}(\mathbf{w}, \pi_A(o_A)). \tag{32}$$

The performance measure is the long-term average reward:

$$\eta_{\pi_A}^{\pi_S} = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^{(\pi_S,\pi_A)}\left(\sum_{t=0}^{T-1} r(t)\right) = \sum_{\mathbf{w}\in\mathcal{I}} \mu_{\pi_A}^{\pi_S}(\mathbf{w}){\cdot}r_{\pi_A}^{\pi_S}(\mathbf{w}) \tag{33}$$

With $\eta_{\pi_A}^{\pi_S}$ in hand, we then introduce the relative reward $g_{\pi_A}^{\pi_S}(\mathbf{w})$, defined by

$$g_{\pi_A}^{\pi_S}(\mathbf{w}) \triangleq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^{(\pi_S,\pi_A)}\left[\sum_{t=0}^{T-1}\left(r(t) - \eta_{\pi_A}^{\pi_S}\right)|\mathbf{W}_0 = \mathbf{w}\right], \tag{34}$$

which satisfies the Poisson equations [60]:

$$\eta_{\pi_A}^{\pi_S} + g_{\pi_A}^{\pi_S}(\mathbf{w}) = r_{\pi_A}^{\pi_S}(\mathbf{w}) + \sum_{\mathbf{w}'\in\mathcal{I}} p_{\pi_A}^{\pi_S}(\mathbf{w}'|\mathbf{w})g_{\pi_A}^{\pi_S}(\mathbf{w}'). \tag{35}$$

Denote $\mathbf{g}_{\pi_A}^{\pi_S}$ as the vector consisting of $g_{\pi_A}^{\pi_S}(\mathbf{w})$, $\mathbf{r}_{\pi_A}^{\pi_S}$ as the vector consisting of $r_{\pi_A}^{\pi_S}(\mathbf{w}), \mathbf{w} \in \mathcal{I}$. $\mathbf{g}_{\pi_A}^{\pi_S}$ could be solved by utilizing [58]:

$$\mathbf{g}_{\pi_A}^{\pi_S} = \left[(I - \mathbf{P}_{\pi_A}^{\pi_S} + \mathbf{e}\boldsymbol{\mu}_{\pi_A}^{\pi_S})^{-1} - \mathbf{e}\boldsymbol{\mu}_{\pi_A}^{\pi_S}\right]\mathbf{r}_{\pi_A}^{\pi_S} \tag{36}$$

With the relative reward $g_{\pi_A}^{\pi_S}$ in hand, we then introduce $Q_{\pi_A}^{\pi_S}(\mathbf{w}, a_A)$ and $Q_{\pi_A}^{\pi_S}(o_A, a_A)$ as follows:

Lemma 3. $Q_{\pi_A}^{\pi_S}(\mathbf{w}, a_A)$ *and* $Q_{\pi_A}^{\pi_S}(o_A, a_A)$ *are defined and calculated as:*

$$\begin{aligned}&Q_{\pi_A}^{\pi_S}(\mathbf{w}, a_A) \triangleq \\ &\limsup_{T\to\infty} \tfrac{1}{T}\mathbb{E}^{(\pi_S,\pi_A)}\left[\sum_{t=0}^{T-1}\left(r(t) - \eta_{\pi_A}^{\pi_S}\right)|\mathbf{W}_0 = \mathbf{w}, a_A(0) = a_A\right] \\ &= \mathcal{R}^{\pi_S}(\mathbf{w}, a_A) - \eta_{\pi_A}^{\pi_S} + \textstyle\sum_{\mathbf{w}'\in\mathcal{I}} p^{\pi_S}(\mathbf{w}'|\mathbf{w}, a_A)g_{\pi_A}^{\pi_S}(\mathbf{w}'),\end{aligned} \tag{37}$$

$$\begin{aligned}&Q_{\pi_A}^{\pi_S}(o_A, a_A) \triangleq \\ &\limsup_{T\to\infty} \tfrac{1}{T}\mathbb{E}^{(\pi_S,\pi_A)}\left[\sum_{t=0}^{T-1}\left(r(t) - \eta_{\pi_A}^{\pi_S}\right)\Big|o_A^{(0)} = o_A, a_A(0) = a_A\right] \\ &= \sum_{\mathbf{w}\in\mathcal{I}} p_{\pi_A}^{\pi_S}(\mathbf{w}|o_A)Q_{\pi_A}^{\pi_S}(\mathbf{w}, a_A),\end{aligned} \tag{38}$$

where $p_{\pi_A}^{\pi_S}(\mathbf{w}'|\mathbf{w})$ *can be obtained by (30) and* $p_{\pi_A}^{\pi_S}(\mathbf{w}|o_A)$ *can be obtained by the Bayesian formula:*

$$p_{\pi_A}^{\pi_S}(\mathbf{w}|o_A) = \frac{\mu_{\pi_A}^{\pi_S}(\mathbf{w})p(o_A|\mathbf{w})}{\sum_{\mathbf{w}\in\mathcal{I}} \mu_{\pi_A}^{\pi_S}(\mathbf{w})p(o_A|\mathbf{w})}. \tag{39}$$

Proof. Please refer to Appendix B. □

With $Q_{\pi_A}^{\pi_S}(o_A, a_A)$ in hand, it is then easy to conduct the Policy Iteration (PI) Algorithm with Step Sizes [61] to iteratively improve the deterministic memoryless decision-making policy π_A. The detailed steps are shown in Algorithm 3.

Algorithm 3: The PI Algorithm with Step Size to Solve the POMDP Given π_S

```
Input: The POMDP tuple
    $\mathscr{M}_{\text{POMDP}}^{\pi_S} \triangleq \langle \mathcal{I}, \mathcal{A}_A, \Omega_A, \mathcal{O}_A, \mathcal{T}_{\text{POMDP}}^{\pi_S}, \mathcal{R}\rangle$, $\epsilon$, $\pi_S$;
Initialization: randomly choose decision-making policy $\pi_A^{(1)}$,
  $\eta_{\pi_A{}^{(0)}}^{\pi_S} = 0$, $\eta_{\pi_A{}^{(-1)}}^{\pi_S} = \infty$, $k = 0$;
while $|\eta_{\pi_A{}^{(k)}}^{\pi_S} - \eta_{\pi_A{}^{(k-1)}}^{\pi_S}| \geq \epsilon$ do
    $k = k + 1$;
    Calculate the transition probability $p_{\pi_A{}^{(k)}}^{\pi_S}(\mathbf{w}'|\mathbf{w})$ by (30);
    Solve the stationary distribution $\boldsymbol{\mu}_{\pi_A{}^{(k)}}^{\pi_S}$ by the stationary equations (31);
    Calculate the expectation reward $r_{\pi_A{}^{(k)}}^{\pi_S}(\mathbf{w})$ by (32) and
      the long-term average reward $\eta_{\pi_A{}^{(k)}}^{\pi_S}$ by (33);
    Calculate the relative reward $g_{\pi_A{}^{(k)}}^{\pi_S}(\mathbf{w})$ by (36);
    for $o_A \in \mathcal{O}_A$ do
        for $a_A \in \mathcal{A}_A$ do
            Calculate $Q_{\pi_A{}^{(k)}}^{\pi_S}(o_A, a_A)$ by Lemma 3;
    for $o_A \in \mathcal{O}_A$ do
        $\pi_A(\cdot|o_A) = \arg\max_{\pi_A(\cdot|o_A)} Q_{\pi_A^{(k)}}^{\pi_S}(o_A, a_A)$;
        $\pi_A^{(k)}(\cdot|o_A) =$
          $\pi_A^{(k-1)}(\cdot|o_A) + \delta_k * (\pi_A^{(k-1)}(\cdot|o_A) - \pi_A(o_A))$;
for $o_A \in \mathcal{O}_A$ do
    $\pi_A^*(o_A) = \arg\max_{a_A} \pi_A^{(k)}(a_A|o_A)$;
$\theta^*(\pi_S, \pi_A^*) = \eta_{\pi_A{}^*}^{\pi_S}$;
Output: $\pi_A^*(\pi_S)$, $\theta^*(\pi_S, \pi_A^*)$
```

Thus far, we have solved two problems: i) by capitalizing on Definition 1 and Algorithm 1, we have ascertained an optimal sampling strategy π_S^* contingent upon the decision-making policy π_A; *ii)* by harnessing Definition 2 and Algorithm 3, we have determined an optimal actuation strategy π_S^* predicated on the decision-making policy π_A. Consequently, we could iteratively employ Algorithm 1 and Algorithm 3 in an alternating fashion, whereby Algorithm 1 yields the optimal sampling strategy $\pi_S^{*(k)}(\pi_A^{(k-1)})$, subsequently serving as an input for Algorithm 3 to derive the decision-making policy $\pi_A^{*(k)}(\pi_S^{*(k)})$. The procedure shall persist until the average reward $\theta^*(\pi_S^{*(k)}, \pi_A^{*(k)})$ reaches convergence, indicating that the solution achieves a *Nash equilibrium* between the sampler and the actuator. The intricacies of the procedure are delineated in Algorithm 4.

Remark 1. *Generally, the JESP algorithm should restart the algorithm by randomly choosing the initial decision-making policy $\pi_A^{*(1)}$ to ensure a good solution, as the initialization of decision-making policy $\pi_A^{*(1)}$ may often lead to poor local optima. We here investigate a heuristic initialization to find the solution quickly and reliably. Specifically, we assume that the decision-maker is fully observable and solve a MDP problem:*

Definition 3. $\mathscr{M}_{\text{MDP}} \triangleq \langle \mathcal{I}, \mathcal{A}_A, \mathcal{T}_{\text{MDP}}, \mathcal{R} \rangle$*, where the elements are given as follows:*

- $\mathcal{I}$*: the set of* $(X_t, \Phi_t) \in \mathcal{S} \times \mathcal{V}$.
- $\mathcal{A}_A = \{a_0, a_{M-1}\}$*: the decision-making set.*
- $\mathcal{T}$*: the transition function, given as*

$$\begin{aligned}&\mathcal{T}((X_t, \Phi_t), a_A, (X_{t+1}, \Phi_{t+1}))\\&= p(X_{t+1}|X_t, a_A, \Phi_t) \cdot p(\Phi_{t+1}|\Phi_t),\end{aligned} \tag{40}$$

  where $p(X_{t+1}|X_t, a_A, \Phi_t)$ *and* $p(\Phi_{t+1}|\Phi_t)$ *could be obtained by (13) and (14), respectively.*
- $\mathcal{R}$*: the same as the pre-defined Dec-POMDP tuple.*

Through solving the above MDP problem, we could explicitly obtain the Q function $Q(X_t, \Phi_t, a_A)$, define $Q(X_t, a_A)$ as $\mathbb{E}_{\Phi_t}[Q(X_t, \Phi_t, a_A)]$, the initial decision-making policy $\pi_A^{*(1)}$ is given as

$$\pi_A^{*(1)}(\hat{X}_t) = \arg\min_{a_A} Q(\hat{X}_t, a_A). \tag{41}$$

V. Simulation Results

Tradition metrics such as Age of Information have been developed under the assumption that a fresher packet or more accurate packet, capable of aiding in source reconstruction, holds a higher value for the receiver, thus promoting goal-oriented decision-making. Nevertheless, the manner in which a packet update impacts the system's *utility* via decision-making remains an unexplored domain. Through the simulations, we endeavor to elucidate the following observations of interest:

• *GoT-optimal vs. State-of-the-art*. In contrast with the state-of-the-art sampling policies, the proposed goal-oriented *sampler & decision-maker* co-design is capable of concurrently maximizing goal attainment and conserving communication resources, accomplishing a closed-loop *utility* optimization via sparse sampling. (See Fig. 6 and 7)

• *Separate Design vs. Co-Design*. Compared to the two-stage sampling-decision-making separate framework, the co-design of sampling and decision-making not only achieves superior goal achievement but also alleviates resource expenditure engendered by communication and actuation implementation. (See Fig. 6 and 7)

Algorithm 4: The Improved JESP Algorithm

```
Input: The Dec-POMDP tuple
       $\mathscr{M}_{DEC-POMDP} \triangleq \langle n, \mathcal{I}, \mathcal{A}, \mathcal{T}, \Omega, \mathcal{O}, \mathcal{R} \rangle$, $\epsilon$;
Initialization: $\theta_0^* = 0$, $\theta_{-1}^* = \infty$, $k = 0$;
Initialize $\pi_A^{*(1)}$ by calculating (41);
while $||\theta_k^* - \theta_{k-1}^*|| \geq \epsilon$ do
    $k = k + 1$;
    Formulate the MDP problem
      $\mathscr{M}_{\text{MDP}}^{\pi_A^{*(k)}} \triangleq \langle \mathcal{I}, \mathcal{A}_S, \mathcal{T}_{\text{MDP}}^{\pi_A^{*(k)}}, \mathcal{R} \rangle$ as given in Definition 1;
    Run Algorithm 1 to obtain $\pi_S^{*(k)}$;
    Formulate the POMDP problem
      $\mathscr{M}_{\text{POMDP}}^{\pi_S^{*(k)}} \triangleq \langle \mathcal{I}, \mathcal{A}_A, \Omega_A, \mathcal{O}_A, \mathcal{T}_{\text{POMDP}}^{\pi_S^{*(k)}}, \mathcal{R} \rangle$ as given
      in Definition 2;
    Run Algorithm 3 to obtain $\theta^*(\pi_S^{*(k)}, \pi_A^{*(k)})$ and $\pi_A^{*(k)}$;
    $\theta_k^* = \theta^*(\pi_S^{*(k)}, \pi_A^{*(k)})$;
The joint sub-optimal policy is: $\pi_S^* = \pi_S^{*(k)}, \pi_A^* = \pi_A^{*(k)}$;
The sub-optimal average reward is: $\theta^* = \theta_k^*$;
Output: $\pi_S^*$, $\pi_A^*$, $\theta^*$
```

• *Optimal Brute-Force-Search vs. Sub-optimal JESP*. Under different successful transmission probability p_S and sampling cost C_S, the sub-optimal yet computation-efficient JESP algorithm will converge to near-optimal solutions. (See Fig. 8)

• *Trade-off: Transmission vs. Actuation*. There is a trade-off between transmission and actuation in terms of resource expenditure: under reliable channel conditions, it is apt to increase communication overhead to ensure effective decision-making; conversely, under poor channel conditions, it is advisable to curtail communication expenses and augment actuation resources to attain maximal system *utility*. (See Fig. 9)

A. Comparing Benchmarks

Fig. 6 illustrates the simulation results, which characterizes the *utility* by the average cost composed by status inherent cost $C_1(X_t, \Phi_t)$, actuation gain cost $C_2(\pi_A(\hat{X}_t))$, actuation inherent cost $C_3(\pi_A(\hat{X}_t))$, and sampling cost C_S. The simulation setup utilizes the parameters detailed in Section II-C2, which is a remote fire monitoring and rescue system.[4] The following comparing benchmarks are considered:

• **Uniform.** Sampling is triggered periodically in this policy. In this case, $a_S(t) = \mathbb{1}_{\{t=K*\Delta\}}$, where $K = 0, 1, 2, \cdots$ and $\Delta \in \mathbb{N}^+$. For each Δ, we can obtain the sampling rate as $1/\Delta$ and explicitly obtain the long-term average cost through Markov chain simulations under pre-defined greedy-based decision-making policy $\pi_A(s_0) = a_0, \pi_A(s_1) = a_3, \pi_A(s_2) = a_7$, which is obtained through (42). The setup of Δ represents a trade-off between *utility* and sampling rate, as depicted in Fig. 6: If C_S is minimal, sampling will contribute positively

[4] We omit specific transition probabilities $p_{i,u}^{(k,m)}$. In the fire monitoring and rescue system, action a_i with higher i increases the probability that the source will stabilize and transition to "No Fire." The "Rainy" context also enhances this likelihood. Typically, these probabilities are informed by empirical data and historical observations and may be arbitrarily chosen as long as the source remains a *unichain*.

to the *utility*; If the sampling action is expansive, sampling may not yield adequate *utility*; If a single sampling consumes moderate resource, the *utility* will exhibit a U-shaped pattern in terms of the sampling rate.

- **Age-aware.** Sampling is executed when the AoI attains a predetermined threshold, a principle that has been established as a threshold-based result for AoI-optimal sampling [30]. In this case, $a_S(t) = \mathbb{1}_{\{\mathrm{AoI(t)}>\delta\}}$, where the AoI-optimal threshold δ can be ascertained using the Bisection method delineated in Algorithm 1 of [30]. In this context, rather than determining a fixed threshold that minimized AoI, we dynamically shift the threshold to explore the balance between sampling and *utility*. As evidenced in Fig. 6, the *utility* derived from this sampling policy consistently surpasses that of the uniform sampling policy.
- **Change-aware.** Sampling is triggered whenever the source status changes. In such a case, $a_S(t) = \mathbb{1}_{\{X_t \neq X_{t-1}\}}$. The performance of this policy is dependent on the dynamics of the system, *i.e.*, if the semantics of the sources transfers frequently, then the sampling rate will be higher. The *utility* of this policy may be arbitrarily detrimental owing to its semantics-unaware nature. In our considered model, the Change-aware policy will turn out 89.5% to have the unsynchronized status $X_t = s_0$ while $\hat{X}_t = s_3$. In this case, the actuator will implement the actuation $a_A = a_7$ according to the observation $\hat{X}_t = s_3$, which will in turn make the status X_t converges to s_0.
- **Optimal MSE.** This is a type of E2E goal-oriented sampling policy if the goal is determined as achieving real-time reconstruction. Nevertheless, this policy disregards the semantics conveyed by the packet and the ensuing actuation updating precipitated by semantics updates. The problem could be formulated as a standard MDP formulation and solved out through RVI Algorithm. The sampling rate and average cost are obtained given the MSE-optimal sampling policy and the pre-defined decision making policy $\pi_A(s_0) = a_0, \pi_A(s_1) = a_3, \pi_A(s_2) = a_7$, which is obtained through (42).
- **Optimal AoII (also optimal AoCI).** From [45], it has been proven that the AoII-optimal sampling policy turns out to be $a_S(t) = \mathbb{1}_{\{X_t \neq \hat{X}_t\}}$. From [42], 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. The sampling rate and average cost are obtained given this sampling policy and the greedy-based decision-making policy $\pi_A(s_0) = a_0, \pi_A(s_1) = a_3, \pi_A(s_2) = a_7$, which is obtained through (42).

B. Separate Design vs. *Co-Design*

Conventionally, the sampling and actuation policies are designed in a two-stage manner: they first emphasize open-loop performance metrics such as average mean squared error (MSE) or average Age of Information (AoI), we then focus on the decision-making policy π_A design. Specifically, we consider that π_A is predetermined using a greedy methodology:

$$\pi_A(\hat{X}_t) = \underset{a_A \in \mathcal{S}_A}{\arg\min} \underset{\Phi_t}{\mathbb{E}} \left\{ \left[C_1(\hat{X}_t, \Phi_t) - C_2(\pi_A(\hat{X}_t)) \right]^+ + C_3(\pi_A(\hat{X}_t)) \right\}. \tag{42}$$

This greedy approach entails selecting the actuation that minimizes cost in the current step, given that th observation $\hat{X}_t$ is perfect. By calculating (42), we obtain $\pi_A(s_0) = a_0, \pi_A(s_1) = a_3, \pi_A(s_2) = a_7$.

However, we notice that sampling and actuation are closely intertwined, highlighting the potential for further co-design. In this paper, we have proposed the RVI-Brute-Force-Search and the Improved JESP algorithms for such optimal co-design. As shown in Fig. 6, the *sampler & decision-maker* co-design achieves the optimal utility through sparse sampling. Specifically, only semantically important information is sampled and transmitted, while non-essential data is excluded. This goal-oriented, semantic-aware, and sparse sampling design represents a significant advancement in sampling policy design. By incorporating a best-matching decision-making policy, the sparse sampling achieves superior performance compared to existing methods.

Fig. 7 presents a comparative analysis between the AoII (or AoCI)-optimal sampling policy, the MSE-optimal sampling policy, and our proposed GoT-optimal *sampler & decision-maker* co-design. It is verified that under different p_S and C_S, the proposed GoT-optimal *sampler & decision-maker* co-design achieves the optimal goal-oriented *utility*. Importantly, under the condition of an extremely unreliable channel $p_S = 0.2$ and high sampling cost $C_S = 10$, the proposed co-design facilitates a significant reduction in long-term average cost, exceeding 60%. This underscores the superiority of the GoT-optimal *sampler & decision-maker* co-design.

C. Optimal vs. Sub-Optimal

Fig. 8 presents a comparative visualization between optimal and sub-optimal solutions over a wide range of C_S and p_S values. The negligible zero-approaching value in Fig. 8 implies a trivial deviation between the optimal and sub-optimal solutions, suggesting the latter's potential for convergence towards near-optimal outcomes. The minimal variance testifies to the sub-optimal algorithm's consistent ability to approximate solutions with high proximity to the optimal. This critical observation underscores the practical advantages of employing sub-optimal improved JESP Algorithm, especially in scenarios with extensive $\mathcal{A}_A$ and $\mathcal{O}_A$.

D. Trade-off: Transmission vs. Actuation

Fig. 9 exemplifies the resource allocation trade-off between transmission and actuation when the long-term average cost is minimized. When the probability of p_S remains low (signifying an unreliable channel) or C_S is high (indicating expensive sampling), it becomes prudent to decrease sampling and transmission, while concurrently augmenting actuation resources for optimal system *utility*. In contrast, when the channel is reliable, sampling and transmission resource can be harmonized with actuation resources to achieve the goal better. This indicates that through the investigation of the optimal co-design of the *sampler & decision-maker* paradigm, a trade-off between transmission and actuation resources can be achieved.

VI. Conclusion

In this paper, we have investigated the GoT metric to directly describe the goal-oriented system decision-making

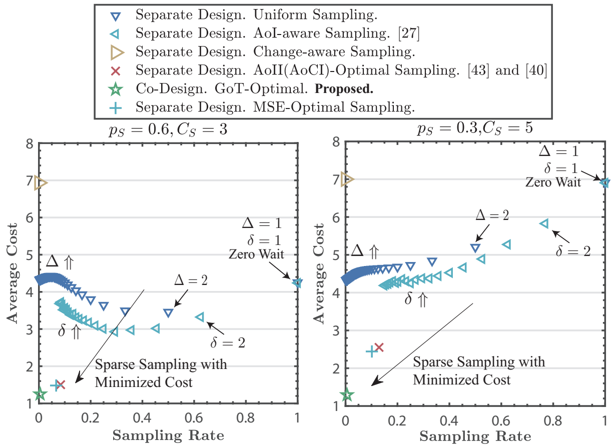


Fig. 6. Average Cost vs. Sampling Rate under different policies and parameters setup. Here we set $C_g = 8$ and $C_I = 1$.

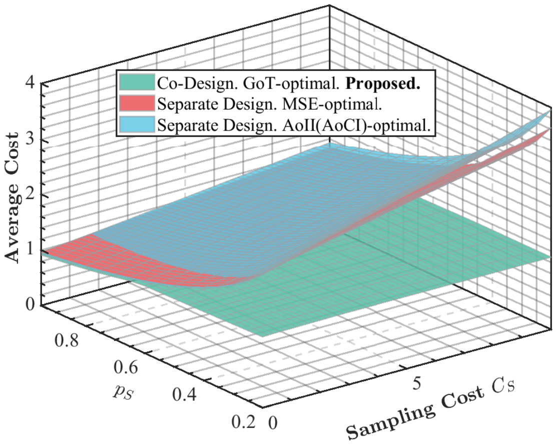


Fig. 7. Comparisons among GoT-optimal, AoII-optimal, and MSE-optimal policies. Here we set $C_g = 8$ and $C_I = 1$.

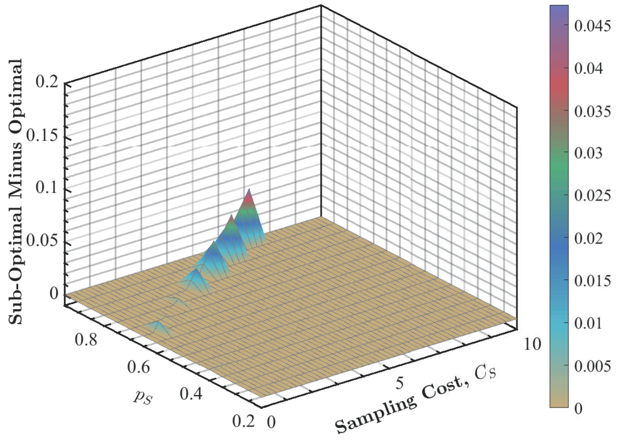


Fig. 8. Distance between optimal solutions through RVI-Brute-Force-Search algorithm and Sub-optimal solutions through JESP algorithm.

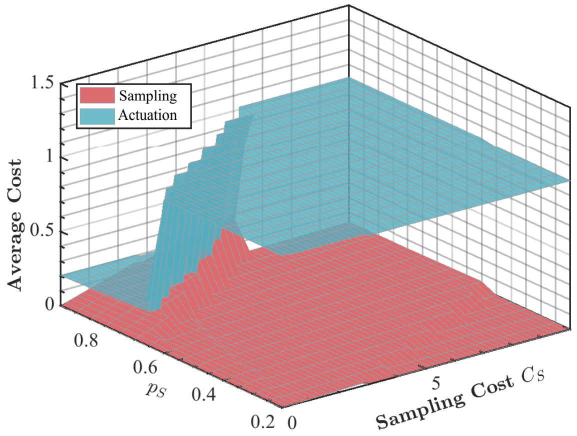


Fig. 9. Trade-off between long-term average transmission cost and long-term average actuation cost.

utility. Employing the proposed GoT, we have formulated an infinite horizon Dec-POMDP problem to accomplish the co-design of sampling and actuating. To address this problem, we have developed two algorithms: the computationally intensive RVI-Brute-Force-Search, which is proven to be optimal, and the more efficient, albeit suboptimal algorithm, named JESP Algorithm. Comparative analyses have substantiated that the proposed GoT-optimal *sampler & decision-maker* pair can achieve sparse sampling and meanwhile maximize the *utility*, signifying the initial realization of a sparse, goal-oriented, and semantics-aware sampler design.

REFERENCES

[1] A. Li, S. Wu, and S. Sun, "Goal-oriented Tensor: Beyond AoI Towards Semantics-Empowered Goal-Oriented Communications," in *Proc. IEEE WCNC, Dubai*, 2024.

[2] 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.

[3] 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*, vol. 63, no. 11, pp. 7492–7508, 2017.

[4] M. Costa, M. Codreanu, and A. Ephremides, "On the Age of Information in Status Update Systems with Packet Management," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1897–1910, 2016.

[5] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, 2018.

[6] C. Kam, S. Kompella, G. D. Nguyen, and A. Ephremides, "Effect of message transmission path diversity on status age," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1360–1374, 2015.

[7] A. M. Bedewy, Y. Sun, and N. B. Shroff, "The Age of Information in Multihop Networks," *IEEE/ACM Trans. Netw.*, vol. 27, no. 3, pp. 1248–1257, 2019.

[8] M. Moltafet, M. Leinonen, and M. Codreanu, "On the Age of Information in Multi-Source Queueing Models," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5003–5017, 2020.

[9] N. Akar, O. Dogan, and E. U. Atay, "Finding the Exact Distribution of (Peak) Age of Information for Queues of PH/PH/1/1 and M/PH/1/2 Type," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5661–5672, 2020.

[10] N. Akar and O. Dogan, "Discrete-Time Queueing Model of Age of

Information With Multiple Information Sources," *IEEE Internet Things J.*, vol. 8, no. 19, pp. 14 531–14 542, 2021.

[11] H. Sac, B. T. Bacinoglu, E. Uysal-Biyikoglu, and G. Durisi, "Age-Optimal Channel Coding Blocklength for an M/G/1 Queue with HARQ," *IEEE Proc. SPAWC*, pp. 1–5, 2018.

[12] M. Xie, Q. Wang, J. Gong, and X. Ma, "Age and Energy Analysis for LDPC Coded Status Update with and Without ARQ," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10 388–10 400, 2020.

[13] J. You, S. Wu, Y. Deng, J. Jiao, and Q. Zhang, "An Age Optimized Hybrid ARQ Scheme for Polar Codes via Gaussian Approximation," *IEEE Wirel. Commun. Lett.*, vol. 10, no. 10, pp. 2235–2239, 2021.

[14] S. Meng, S. Wu, A. Li, J. Jiao, N. Zhang, and Q. Zhang, "Analysis and Optimization of the HARQ-Based Spinal Coded Timely Status Update System," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6425–6440, 2022.

[15] H. Pan, T.-T. Chan, V. C. Leung, and J. Li, "Age of Information in Physical-layer Network Coding Enabled Two-way Relay Networks," *IEEE Trans. Mob. Comput.*, 2022.

[16] A. Maatouk, M. Assaad, and A. Ephremides, "Minimizing the Age of Information: NOMA or OMA?" *IEEE INFOCOM WKSHPS*, pp. 102–108, 2019.

[17] S. Wu, Z. Deng, A. Li, J. Jiao, N. Zhang, and Q. Zhang, "Minimizing Age-of-Information in HARQ-CC Aided NOMA Systems," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 2, pp. 1072–1086, 2023.

[18] E. T. Ceran, D. Gündüz, and A. György, "Average Age of Information with Hybrid ARQ under a resource constraint," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 3, pp. 1900–1913, 2019.

[19] D. Li, S. Wu, J. Jiao, N. Zhang, and Q. Zhang, "Age-Oriented Transmission Protocol Design in Space-Air-Ground Integrated Networks," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 7, pp. 5573–5585, 2022.

[20] F. Peng, Z. Jiang, S. Zhang, and S. Xu, "Age of Information Optimized MAC in V2X Sidelink via Piggyback-Based Collaboration," *IEEE Trans. Wirel. Commun.*, vol. 20, pp. 607–622, 2020.

[21] H. Pan, T.-T. Chan, J. Li, and V. C. M. Leung, "Age of Information With Collision-Resolution Random Access," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 11 295–11 300, 2022.

[22] A. Li, S. Wu, J. Jiao, N. Zhang, and Q. Zhang, "Age of Information with Hybrid-ARQ: A Unified Explicit Result," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 7899–7914, 2022.

[23] B. Zhou and W. Saad, "Joint Status Sampling and Updating for Minimizing Age of Information in the Internet of Things," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7468–7482, 2019.

[24] X. Xie, H. Wang, and X. Liu, "Scheduling for Minimizing the Age of Information in Multi-Sensor Multi-Server Industrial IoT Systems," *IEEE Trans. Industr. Inform.*, vol. 20, no. 1, pp. 573–582, 2024.

[25] M. A. Abd-Elmagid and H. S. Dhillon, "Closed-Form Characterization of the MGF of AoI in Energy Harvesting Status Update Systems," *IEEE Trans. Inf. Theory*, vol. 68, no. 6, pp. 3896–3919, 2022.

[26] M. Hatami, M. Leinonen, Z. Chen, N. Pappas, and M. Codreanu, "On-Demand AoI Minimization in Resource-Constrained Cache-Enabled IoT Networks With Energy Harvesting Sensors," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7446–7463, 2022.

[27] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *IEEE Proc. ISIT*, 2015, pp. 3008–3012.

[28] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 534–556, 2019.

[29] 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.

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

[31] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The cost of delay in status updates and their value: Non-linear ageing," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4905–4918, 2020.

[32] J. Cho and H. Garcia-Molina, "Effective page refresh policies for web crawlers," *ACM ACM Trans. Database Syst.*, vol. 28, no. 4, pp. 390–426, 2003.

[33] M. Bastopcu and S. Ulukus, "Information freshness in cache updating systems," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 3, pp. 1861–1874, 2020.

[34] K. T. Truong and R. W. Heath, "Effects of Channel Aging in Massive MIMO systems," *J. Commun. Networks*, vol. 15, no. 4, pp. 338–351, 2013.

[35] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," in *IEEE SPAWC Workshop*, 2018, pp. 1–5.

[36] Z. Wang, M.-A. Badiu, and J. P. Coon, "A Framework for Characterizing the Value of Information in Hidden Markov Models," *IEEE Trans. Inf. Theory*, vol. 68, no. 8, pp. 5203–5216, 2022.

[37] G. Chen, S. C. Liew, and Y. Shao, "Uncertainty-of-Information Scheduling: A Restless Multiarmed Bandit Framework," *IEEE Trans. Inf. Theory*, vol. 68, no. 9, pp. 6151–6173, 2022.

[38] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Towards an Effective Age of Information: Remote Estimation of a Markov Source," in *IEEE INFOCOM WKSHPS*, 2018, pp. 367–372.

[39] 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*, vol. 66, no. 2, pp. 1118–1135, 2019.

[40] T. Z. Ornee and Y. Sun, "Sampling for Remote Estimation Through Queues: Age of Information and Beyond," in *Proc. WiOPT*, 2019, pp. 1–8.

[41] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor, "Sample, quantize, and encode: Timely estimation over noisy channels," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6485–6499, 2021.

[42] 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.

[43] 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.

[44] J. Zhong, R. D. Yates, and E. Soljanin, "Two freshness metrics for local cache refresh," in *IEEE Proc. ISIT*, pp. 1924–1928.

[45] 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.

[46] J. Cao, X. Zhu, S. Sun, P. Popovski, S. Feng, and Y. Jiang, "Age of Loop for Wireless Networked Control System in the Finite Blocklength Regime: Average, Variance and Outage Probability," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 8, pp. 5306–5320, 2023.

[47] A. Nikkhah, A. Ephremides, and N. Pappas, "Age of Actuation in a Wireless Power Transfer System," in *IEEE INFOCOM WKSHPS*, 2023.

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

[49] N. Pappas and M. Kountouris, "Goal-oriented communication for real-time tracking in autonomous systems," in *IEEE Proc. ICAS*, 2021, pp. 1–5.

[50] M. Salimnejad, M. Kountouris, and N. Pappas, "Real-time Reconstruction of Markov Sources and Remote Actuation over Wireless Channels," *IEEE Trans. Commun.*, 2024.

[51] E. Fountoulakis, N. Pappas, and M. Kountouris, "Goal-oriented Policies for Cost of Actuation Error Minimization in Wireless Autonomous Systems," *IEEE Commun. Lett.*, vol. 27, no. 9, pp. 2323–2327, 2023.

[52] A. Li, S. Wu, S. Meng, S. Sun, R. Lu, and Q. Zhang, "Towards Goal-Oriented Semantic Communications: New Metrics, Open Challenges, and Future Research Directions," *IEEE Wirel. Commun. to appear*, 2024. [Online]. Available: https://arxiv.org/abs/2304.00848

[53] 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.

[54] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, B. Soret *et al.*, "Semantic communications in networked systems: A data significance perspective," *IEEE Netw.*, vol. 36, no. 4, pp. 233–240, 2022.

[55] A. Maatouk, M. Assaad, and A. Ephremides, "The Age of Incorrect Information: An Enabler of Semantics-Empowered Communication," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 4, pp. 2621–2635, 2023.

[56] C. Kam, S. Kompella, and A. Ephremides, "Age of Incorrect Information for Remote Estimation of a Binary Markov Source," in *IEEE INFOCOM WKSHPS*, 2020, pp. 1–6.

[57] D. Bertsekas, *Dynamic Programming and Optimal Control, 3rd Edition*. Athena scientific, 2005, vol. 1.

[58] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 2014.

[59] R. Nair, M. Tambe, M. Yokoo, D. V. Pynadath, and S. Marsella, "Taming Decentralized POMDPs: Towards Efficient Policy Computation for Multiagent Settings," in *Proc. IJCAI, 2003*.

[60] D. P. Bertsekas and J. N. Tsitsiklis, "Neuro-dynamic programming: an overview," in *IEEE Proc. CDC*, vol. 1, pp. 560–564.

[61] Y. Li, B. Yin, and H. Xi, "Finding optimal memoryless policies of pomdps under the expected average reward criterion," *Eur. J. Oper. Res.*, vol. 211, no. 3, pp. 556–567, 2011.

Appendix A
The Proof of Lemma 1

By taking into account the *conditional independence* among X_{t+1}, Φ_{t+1}, and X_{t+1}, given (X_t, Φ_t, X_t) and $\mathbf{a}(t)$, we we can express the following:

$$\begin{aligned}&\Pr\left\{\mathbf{W}_{t+1}=(s_u,x,v_r)\,|\mathbf{W}_t=(s_i,s_j,v_k),\mathbf{a}(t)=(1,a_m)\right\}\\&=\Pr\left\{X_{t+1}=s_u\,|\mathbf{W}_t=(s_i,s_j,v_k),\mathbf{a}(t)=(1,a_m)\right\}\times\\&\Pr\left\{\hat{X}_{t+1}=x\,|\mathbf{W}_t=(s_i,s_j,v_k),\mathbf{a}(t)=(1,a_m)\right\}\times\\&\Pr\left\{\Phi_{t+1}=v_r\,|\mathbf{W}_t=(s_i,s_j,v_k),\mathbf{a}(t)=(1,a_m)\right\},\end{aligned}\tag{43}$$

wherein the first, second, and third terms can be derived through *conditional independence*, resulting in simplified expressions of (13), (18), and (14), respectively:

$$\Pr\left\{X_{t+1}=s_u\,|\mathbf{W}_t=(s_i,s_j,v_k),\mathbf{a}(t)=(1,a_m)\right\}=p_{i,u}^{(k,m)},\tag{44}$$

$$\begin{aligned}&\Pr\left\{\hat{X}_{t+1}=x\,|\mathbf{W}_t=(s_i,s_j,v_k),\mathbf{a}(t)=(1,a_m)\right\}\\&=p_S\cdot\mathbb{1}_{\{x=s_i\}}+(1-p_S)\cdot\mathbb{1}_{\{x=s_j\}},\end{aligned}\tag{45}$$

$$\Pr\left\{\Phi_{t+1}=v_r\,|\mathbf{W}_t=(s_i,s_j,v_k),\mathbf{a}(t)=(1,a_m)\right\}=p_{k,r},\tag{46}$$

Substituting (44), (45), and (46) into (43) yields the (1) in Lemma 1. In the case where $a_S(t)=0$, we can obtain a similar expression by replacing $\mathbf{a}(t)=(1,m)$ with $\mathbf{a}(t)=(0,m)$. Substituting (13), (15), and (14) into this new expression results in the proof of (22) in Lemma 1.

Appendix B
Proof of Lemma 3

$Q^{\pi_S}_{\pi_A}(\mathbf{w},a_A)$ could be simplified as follows:

$$\begin{aligned}&\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}^{(\pi_S,\pi_A)}\left[\sum_{t=0}^{T-1}\left(r(t)-\eta^{\pi_S}_{\pi_A}\right)|\mathbf{W}_0=\mathbf{w},a_A(0)=a_A\right]\\&=\mathcal{R}^{\pi_S}(\mathbf{w},a_A)-\eta^{\pi_S}_{\pi_A}+\sum_{\mathbf{w}'\in\mathcal{I}}p^{\pi_S}(\mathbf{w}'|\mathbf{w},a_A)\cdot\\&\underbrace{\limsup_{T\to\infty}\frac{1}{T-1}\mathbb{E}^{(\pi_S,\pi_A)}\left[\sum_{t=1}^{T-1}\left(r(t)-\eta^{\pi_S}_{\pi_A}\right)|\mathbf{W}_1=\mathbf{w}'\right]}_{g^{\pi_S}_{\pi_A}(\mathbf{w}')}\\&=\mathcal{R}^{\pi_S}(\mathbf{w},a_A)-\eta^{\pi_S}_{\pi_A}+\sum_{\mathbf{w}'\in\mathcal{I}}p^{\pi_S}(\mathbf{w}'|\mathbf{w},a_A)g^{\pi_S}_{\pi_A}(\mathbf{w}'),\end{aligned}\tag{47}$$

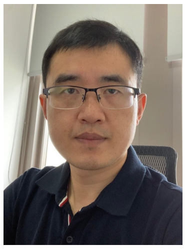
$Q^{\pi_S}_{\pi_A}(o_A,a_A)$ could be solved as:

$$\begin{aligned}&\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}^{(\pi_S,\pi_A)}\left[\sum_{t=0}^{T-1}\left(r(t)-\eta^{\pi_S}_{\pi_A}\right)\Big|o_A^{(0)}=o_A,a_A(0)=a_A\right]\\&=\sum_{\mathbf{w}\in\mathcal{I}}p^{\pi_S}_{\pi_A}(\mathbf{w}|o_A,a_A)\\&\cdot\underbrace{\limsup_{T\to\infty}\frac{1}{T-1}\mathbb{E}^{(\pi_S,\pi_A)}\left[\sum_{t=0}^{T-1}\left(r(t)-\eta^{\pi_S}_{\pi_A}\right)|\mathbf{W}_0=\mathbf{w},a_A(0)=a_A\right]}_{Q^{\pi_S}_{\pi_A}(\mathbf{w},a_A)}\\&=\sum_{\mathbf{w}\in\mathcal{I}}p^{\pi_S}_{\pi_A}(\mathbf{w}|o_A,a_A)Q^{\pi_S}_{\pi_A}(\mathbf{w},a_A).\end{aligned}\tag{48}$$

where $p^{\pi_S}_{\pi_A}(\mathbf{w}|o_A,a_A)$ is the posterior conditional probability. Note that the state $\mathbf{w}$ is independent of a_A when o_A is known, we have that

$$\begin{aligned}p^{\pi_S}_{\pi_A}(\mathbf{w}|o_A,a_A)=p^{\pi_S}_{\pi_A}(\mathbf{w}|o_A)&=\frac{p^{\pi_S}_{\pi_A}(\mathbf{w},o_A)}{p^{\pi_S}_{\pi_A}(o_A)}\\&=\frac{\boldsymbol{\mu}^{\pi_S}_{\pi_A}(\mathbf{w})p(o_A|\mathbf{w})}{\sum_{\mathbf{w}\in\mathcal{I}}\boldsymbol{\mu}^{\pi_S}_{\pi_A}(\mathbf{w})p(o_A|\mathbf{w})}.\end{aligned}\tag{49}$$

Aimin Li (Student Member, IEEE) received his B.E. from Harbin Institute of Technology Shenzhen (HITSZ) in 2020, where he was awarded the highest honor of Undergraduate Thesis. He is currently a Ph.D student at HITSZ and a visit student at Institute for Infocomm Research (I2R), Agency for Science, Technology, and Research (A*STAR). He has served as a Reviewer for IEEE TWC, IEEE TNNLS, IEEE TVT, IEEE CL, IEEE ISIT, *etc.* His current research interests include channel coding, age of information and goal-oriented semantic communications.



Shaohua Wu (Member, IEEE) received the Ph.D. degree in communication engineering from Harbin Institute of Technology, Harbin, China, in 2009. From 2009 to 2011, he held a postdoctoral position with the Department of Electronics and Information Engineering, Shenzhen Graduate School, Harbin Institute of Technology, where he has been with since 2012. From 2014 to 2015, he was a Visiting Researcher with BBCR, University of Waterloo, Canada. He is currently a Full Professor with the Harbin Institute of Technology (Shenzhen), China. He is also a Professor with Peng Cheng Laboratory, Shenzhen, China. His research interests include satellite and space communications, advanced channel coding techniques, space-air-ground-sea integrated networks, and B5G/6G wireless transmission technologies. He has authored or coauthored over 100 papers in these fields and holds over 40 Chinese patents.

Sumei Sun (Fellow, IEEE) is Executive Director of the Institute for Infocomm Research (I2R), Agency for Science, Technology, and Research (A*STAR), Singapore. She also holds an adjunct appointment with the National University of Singapore, and joint appointment with the Singapore Institute of Technology, both as a full professor. Her current research interests include next-generation wireless communications, joint communication-sensing-computing-control design, industrial internet of things, applied deep learning and artificial intelligence. She is a member of the IEEE Vehicular Technology Society Board of Governors (2022-2024), Fellow of the IEEE and the Academy of Engineering Singapore.

Jie Cao (Member, IEEE) received the Ph.D. degree from the Harbin Institute of Technology, Shenzhen, China, in 2022. He was a Research Assistant and subsequently a Research Scientist of the Institute for Infocomm Research, Agency for Science, Technology, and Research (A*STAR), Singapore from 2021-2023. He is currently an Assistant Professor of the Harbin Institute of Technology, Shenzhen, China. He has served as an Associate Editor for IEEE OJ-COMS, a Reviewer for the IEEE Transactions on Wireless Communications and a Co-chair for IEEE ICC, IEEE VTC Workshops. His research interests include URLLC, short packet communication, age of information and task-oriented communication. He has authored and co-authored more than 30 research papers published on top-tier journals and international conferences.