Agency for Science, Technology and Research
SINGAPORE

Goal-oriented Semantic Communications: An Integrated Sensing-Communications-Control Co-Design

Presentor : Aimin Li
Supervisor: Prof. Sumei Sun

Outline

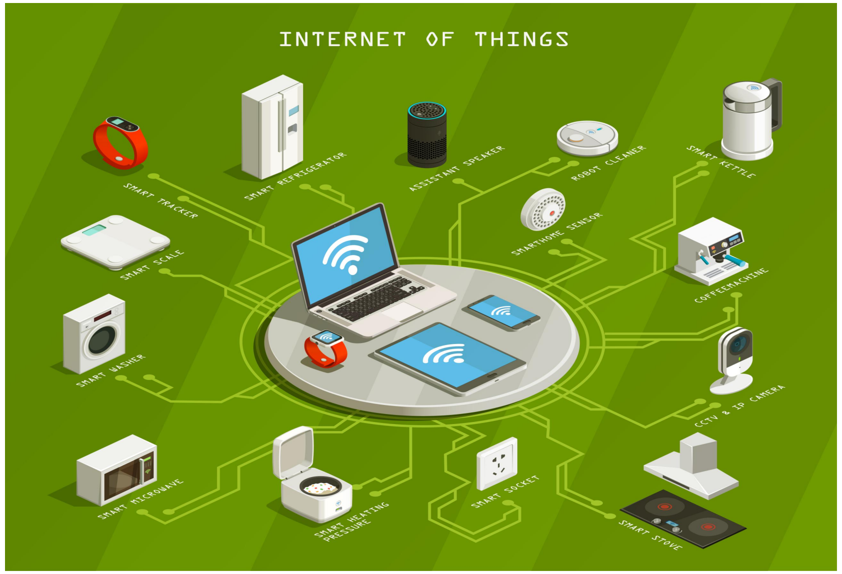
- Introduction and Motivation
- Current Approaches for Mesearing Information Importance
  - Age of Information
  - Beyond Age of Information
  - Goal-oriented Tensor
- Goal-oriented Semantic Communications——System Model
- Solutions and Results
- Furture Plan

Introduction and Motivation

Emerging Wireless Internet of Everything

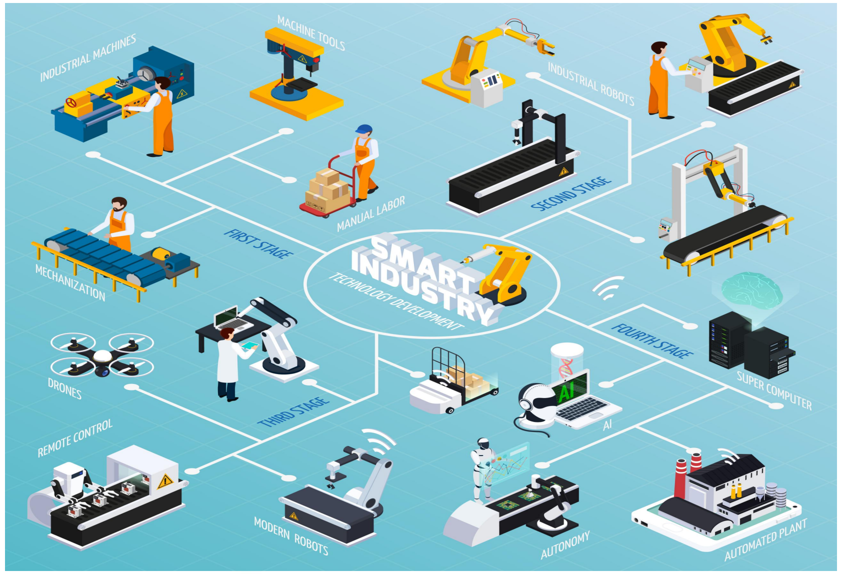
- Internet of Everything
  - Real-time multi-modal data exchange, such as video, text, AR, VR, video streams...
  - Real time information reconstruction, network and device computation, traffic flows with synced requirements, autonomous interactions

Internet of Things

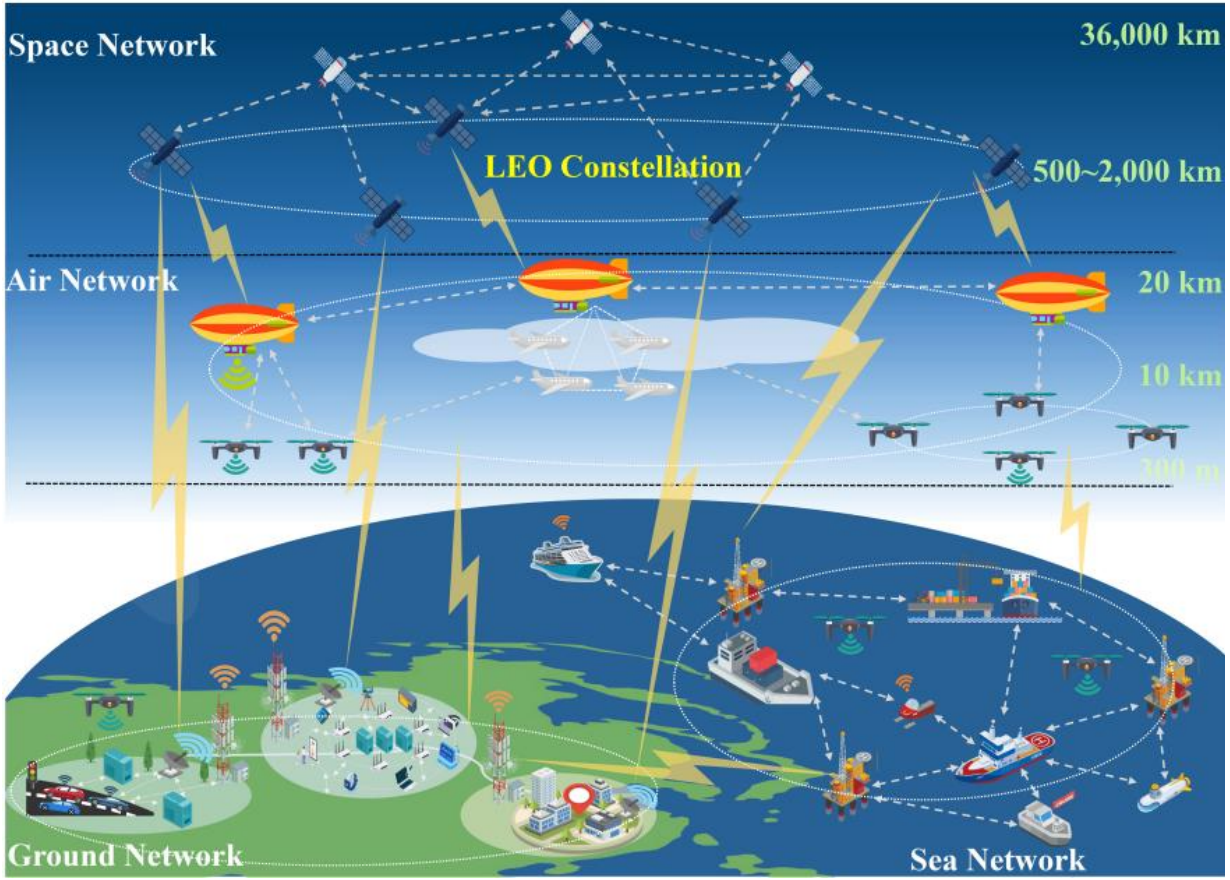


Internet of Vehicle

Smart Industry 4.0



Challange—Data Exploding vs. Communication Resources



Sketch for the 6G IoV *Space-Air-Ground-Sea Integrated Networks* （SAGSIN）

Data Explosion

- **Multi-modal** raw data from the seamless-coverage SAGSIN-based Internet of Everything are generated **every seconds**.

Limited Communications Resources

- Maximizing throughput, minimizing delay and achieving bit-by-bit information reconstruction are **energy-intensive**.

Challenge: Communicatioins Resource Constraints in the Face of Data Explosion

How to Solve this Challange?——Go Beyond Bit-Orient Communications

Challenge: Communicatioins Resource Constraints in the Face of Data Explosion

Trace Back To...

Shannon and Weaver in 1949

Recent Contributions to
The Mathematical Theory of Communication

Warren Weaver

September, 1949

Claude Shannon

Warren Weaver

Abstract

This paper is written in three main sections. In the first and third, W. W. is responsible both for the ideas and the form. The middle section, namely "2) Communication Problems at Level A" is an interpretation of mathematical papers by Dr. Claude E. Shannon of the Bell Telephone Laboratories. Dr. Shannon's work roots back, as von Neumann has pointed out, to Boltzmann's observation, in some of his work on statistical physics (1894), that entropy is related to "missing information," inasmuch as it is related to the number of alternatives which remain possible to a physical system after all the macroscopically observable information concerning it has been recorded. L. Szilard

Three Levels of Communications

Challenge: Communicatioins Resource Constraints in the Face of Data Explosion

Trace Back To...

Shannon and Weaver in 1949

Recent Contributions to
The Mathematical Theory of Communication

Warren Weaver

September, 1949

Claude Shannon

Warren Weaver

Abstract

This paper is written in three main sections. In the first and third, W. W. is responsible both for the ideas and the form. The middle section, namely "2) Communication Problems at Level A" is an interpretation of mathematical papers by Dr. Claude E. Shannon of the Bell Telephone Laboratories. Dr. Shannon's work roots back, as von Neumann has pointed out, to Boltzmann's observation, in some of his work on statistical physics (1894), that entropy is related to "missing information," inasmuch as it is related to the number of alternatives which remain possible to a physical system after all the macroscopically observable information concerning it has been recorded. L. Szilard

Three Levels of Communications

How to Solve this Challange?——Go Beyond Bit-Orient Communications

Challenge: Communicatioins Resource Constraints in the Face of Data Explosion

Trace Back To...

Shannon and Weaver in 1949

Recent Contributions to
The Mathematical Theory of Communication

Warren Weaver

September, 1949

Claude Shannon

Warren Weaver

Abstract

This paper is written in three main sections. In the first and third, W. W. is responsible both for the ideas and the form. The middle section, namely "2) Communication Problems at Level A" is an interpretation of mathematical papers by Dr. Claude E. Shannon of the Bell Telephone Laboratories. Dr. Shannon's work roots back, as von Neumann has pointed out, to Boltzmann's observation, in some of his work on statistical physics (1894), that entropy is related to "missing information," inasmuch as it is related to the number of alternatives which remain possible to a physical system after all the macroscopically observable information concerning it has been recorded. L. Szilard

Three Levels of Communications

How to Solve this Challange?——Go Beyond Bit-Orient Communications

Challenge: Communicatioins Resource Constraints in the Face of Data Explosion

Trace Back To...

Shannon and Weaver in 1949

Recent Contributions to
The Mathematical Theory of Communication

Warren Weaver

September, 1949

Claude Shannon

Warren Weaver

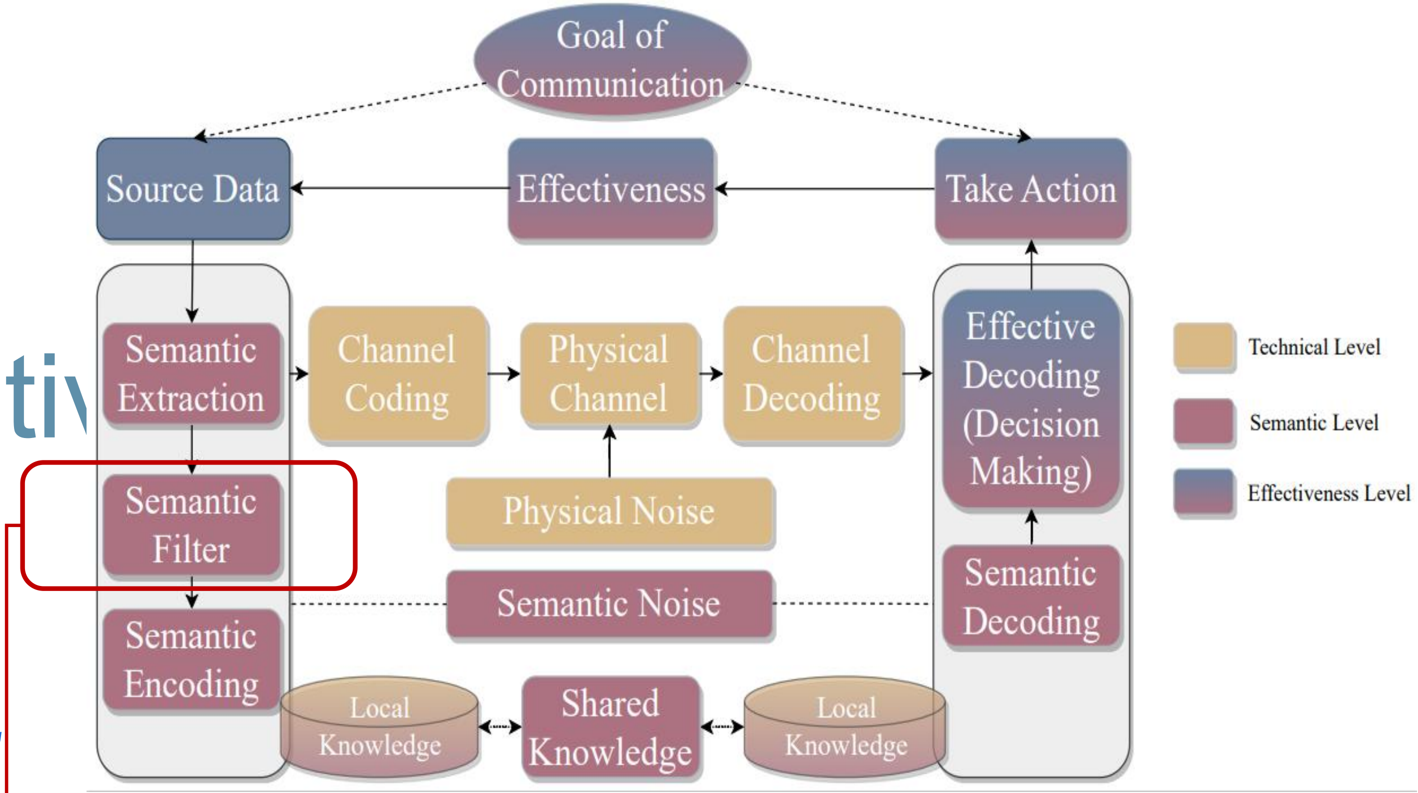
Abstract

This paper is written in three main sections. In the first and third, W. W. is responsible both for the ideas and the form. The middle section, namely "2) Communication Problems at Level A" is an interpretation of mathematical papers by Dr. Claude E. Shannon of the Bell Telephone Laboratories. Dr. Shannon's work roots back, as von Neumann has pointed out, to Boltzmann's observation, in some of his work on statistical physics (1894), that entropy is related to "missing information," inasmuch as it is related to the number of alternatives which remain possible to a physical system after all the macroscopically observable information concerning it has been recorded. L. Szilard

Three Levels of Communications

- ***Goal-oriented***
  - Communication is ***about achieving specific goals.***
  - For example, the reliable, low-latency communications eventually serve for ***goal-oriented decision making***.
- ***Semantics-empowered***
  - The semantics of information is further explored and extrated to enable ***data compressing.***
  - The semantics also means how the semantics of information helps atain a specific goal.
- ***A holistic redesign of the entire process of information generation, processing, transmission, and reconstruction.***

Motiv

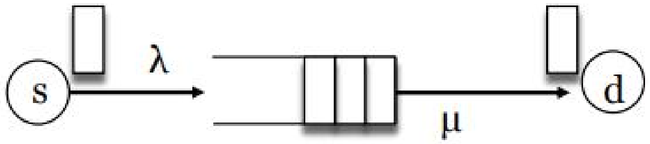


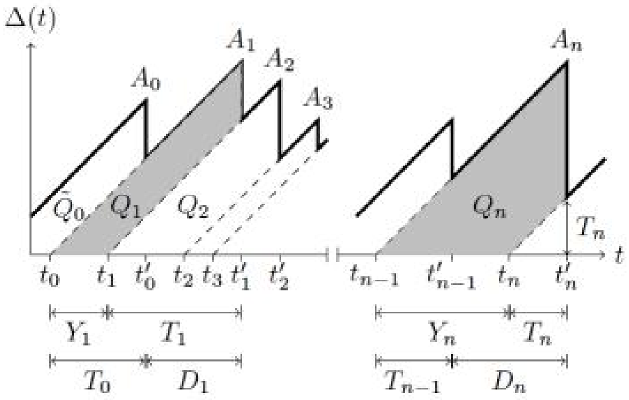
Based on the importance of Information

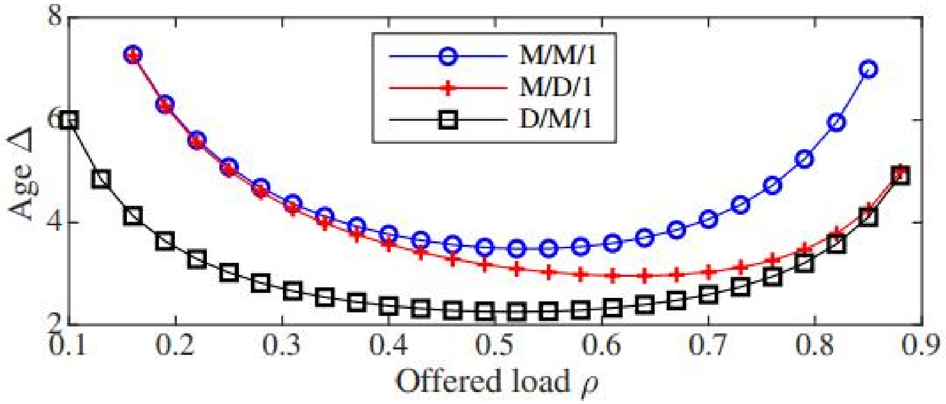
Current Approaches for Measuring Information Importance: AoI and Beyond

Quantify the Importance of Information in Achieving a Goal?

- **Real-time / freshness-critical systems:** ***Information usually has the highest value when it is fresh!*** (e.g., autonomous driving: info about location/speed/sensors)
- **Age of Information (AoI):** A new metric that could quantify the freshness of Information. (The fresher the information, the higher value the semantics)
- **Definition:*****An update packet with timestamp u has age t-u at a time t.***
- An update is fresh if its age is zero.
- AoI and its variants...





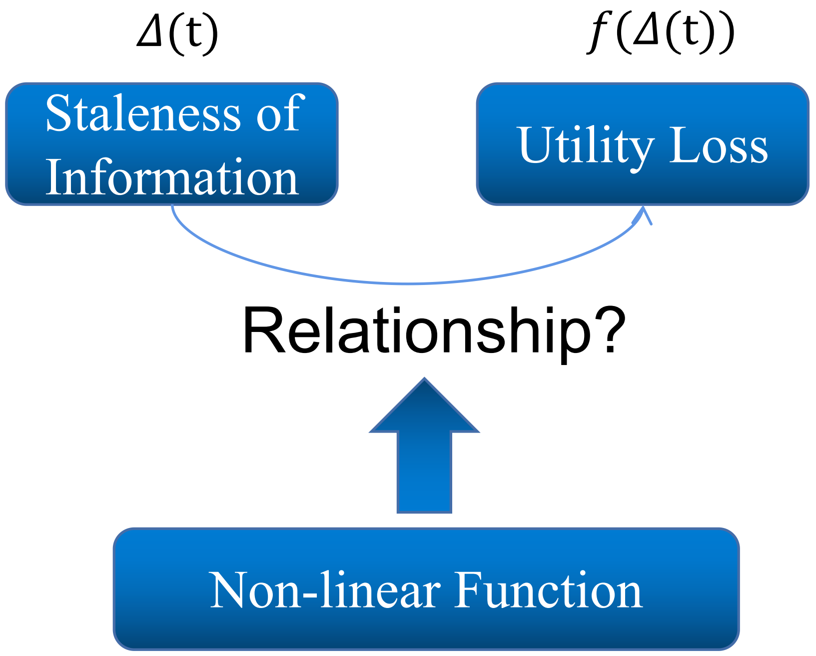


S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in 2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad-Hoc Communications and Networks, 2011.

Quantify the Importance of Information in Achieving a Goal?

Non-linear Age (Value of Information)

- AoI grows over time linearly, it describes the staleness property of the packet.
- However, ***the causality of how a stale packet will result in utility loss is not known.***
  - the performance degradation caused by information aging may not be a linear function of time.
- One way to capture the nonlinear behavior of information aging is to ***define freshness and staleness as nonlinear functions of AoI.***



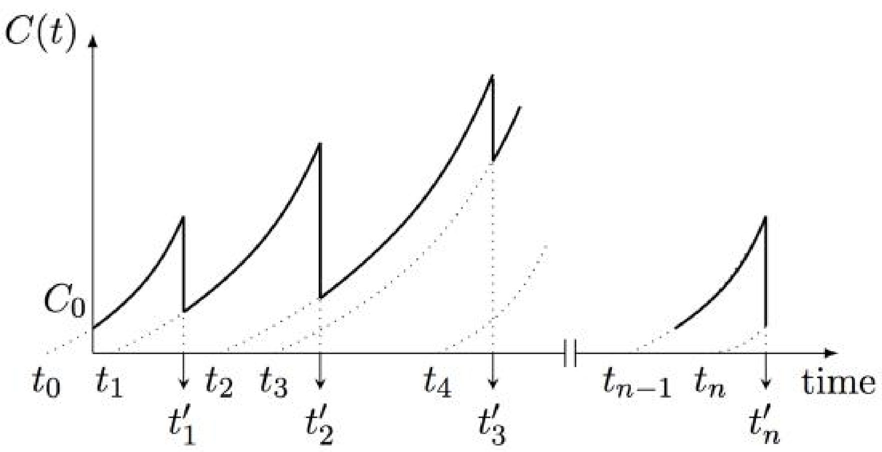
[1] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case", IEEE ISIT 2017.

[2] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The cost of delay in status updates and their value: Non-linear ageing", IEEE Trans. Comm., 2020.

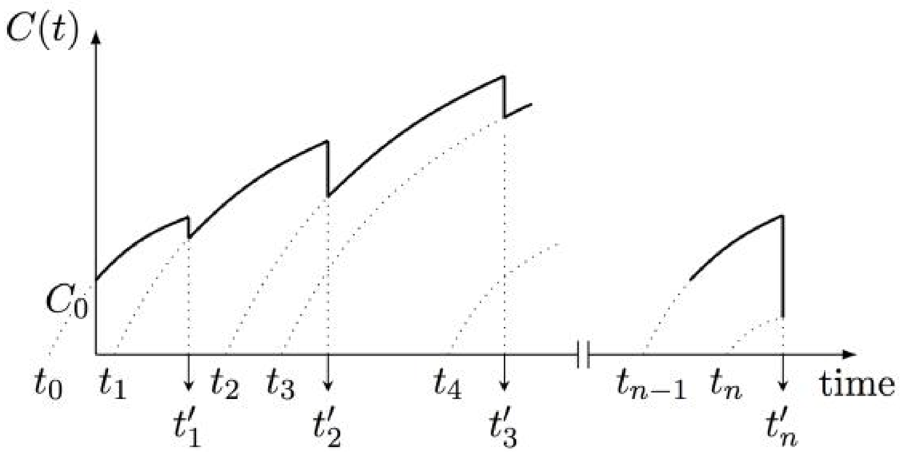
Quantify the Importance of Information in Achieving a Goal?

Non-linear Age (Value of Information)

$f_s(t) = e^{\alpha t} - 1 \longleftrightarrow$ low autocorrelation



$f_s(t) = \log(\alpha t + 1) \longleftrightarrow$ high autocorrelation



[1] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, “Age and value of information: Non-linear age case”, IEEE ISIT 2017.
[2] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, “The cost of delay in status updates and their value: Non-linear ageing”, IEEE Trans. Comm., 2020.

Quantify the Importance of Information in Achieving a Goal?

Mean Square Error (MSE)

- Mean Square Error is used to measure the reconstruction performance.
  - ***The goal of minimizing MSE is to achieve accurate reconstruction.***
  - It is assumed that accurate reconstruction will lead to the subsequent accurate decisiong making.
- MSE could be derived as a non-linear AoI if the sampling is content-agnostic.

$MSE(t) = f(\Delta(t))$, ***if sampling is content-agnostic***

[1] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the Wiener Process for Remote Estimation over a Channel with Random Delay", IEEE Transactions on Information Theory 2017.
[2] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," in IEEE WiOPT, 2019, pp. 1–8.
[3] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor, "Sample, quantize, and encode: Timely estimation over noisy channels," IEEE Transactions on Communications, vol. 69, no. 10, pp. 6485–6499, 2021.

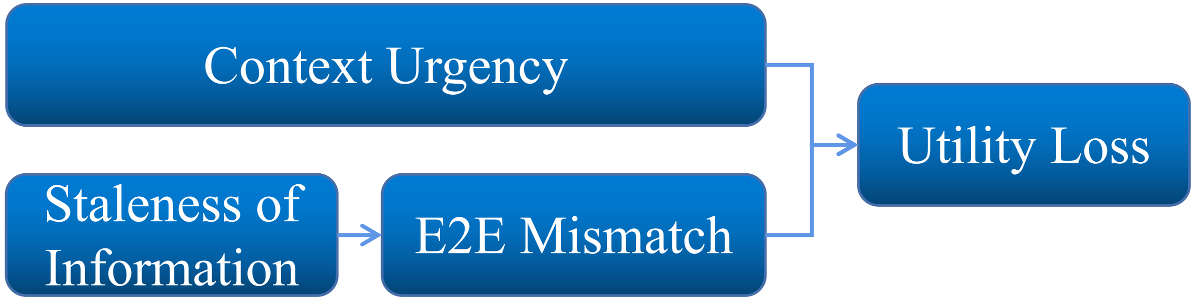
Quantify the Importance of Information in Achieving a Goal?

Urgency of Information (UoI)

- The mulplication of the context-aware coeficient $\varphi(\mathrm{t})$ and the MSE:

$$\mathrm{UoI}(t) = \varphi(t) \cdot (X_t - \widehat{X}_t)^2$$

  - A car dirving in a complex congested road requires a higher coeficient $\varphi(\mathrm{t})$.
  - In contrast, A car dirving in a straight road requires a lower coeficient $\varphi(\mathrm{t})$.



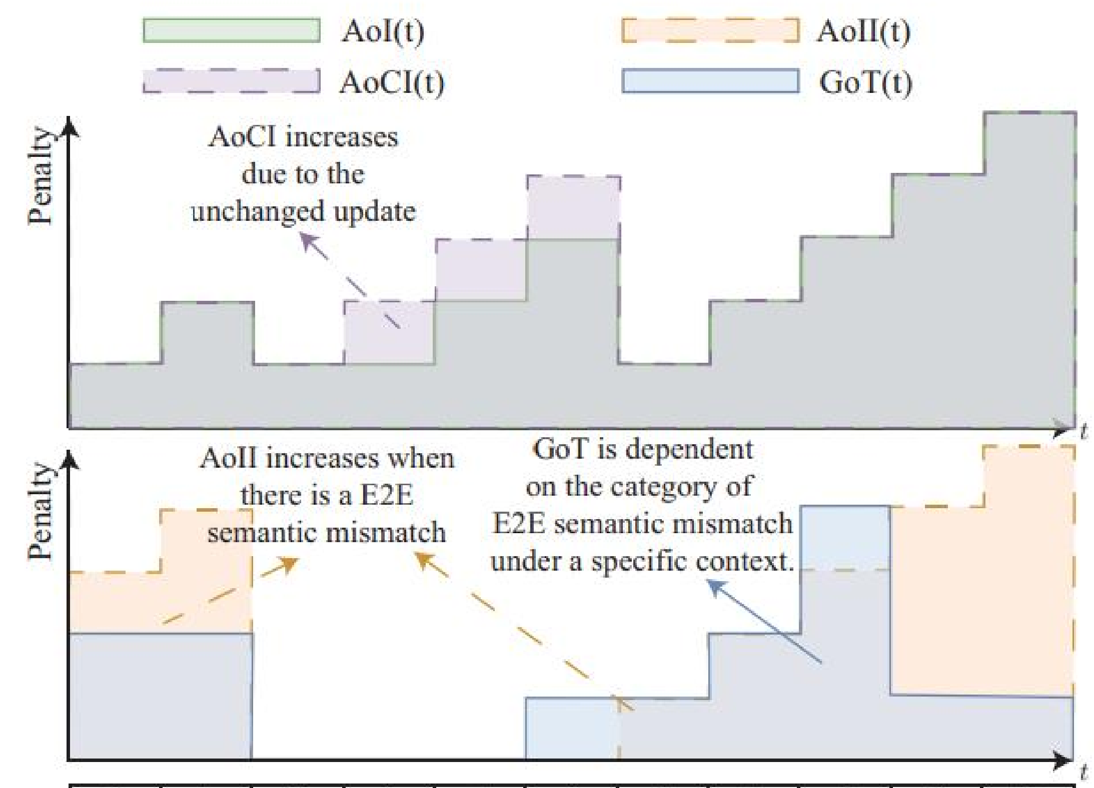
X. Zheng, S. Zhou, and Z. Niu, “Urgency of information for context-aware timely status updates in remote control systems,” IEEE Transactions on Wireless Communications, vol. 19, no. 11, pp. 7237–7250, 2020.

Quantify the Importance of Information in Achieving a Goal?

Other metrics beyond AoI...

- Age of Incorrect Information(AoII)
- Age of Changed Information(AoCI)
- Age of Loop (AoL)
- Age of Actuation(AoA)

[1] A. Maatouk, S. Kriouile, M. Assaad and A. Ephremides, "The Age of Incorrect Information: A New Performance Metric for Status Updates", IEEE/ACM Trans. on Networking 2020.
[2] 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 Transactions on Communications, 2022.
[3] 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 Transactions on Wireless Communications, 2023.
[4] A. Nikkhah, A. Ephremides, and N. Pappas, “Age of Actuation in a Wireless Power Transfer System,” in INFOCOM Age of Information WKSHPS, 2023.



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

Quantify the Importance of Information in Achieving a Goal?

Motivation:

- The current metrics minimize the utility loss ***indirectly***!
  - AoI attempts to avoid staleness of information to minimize the utility loss.
  - MSE attempts to avoid E2E mismatch to minimize the utility loss.
  - Some other metrics also follow the way...
- There are many variants, while there is ***not yet a unified framework***!

Our Method:

- New metric: Goal-oriented Tensor
  - ✓ It could evaluate the goal-achieving utility ***directly***.
  - ✓ It could ***derade to existing metrics***, e.g., AoI, UoI, AoII, etc.
  - ✓ The dimension is ***extensible***, enabling various goal-oriented applications.

[1] **Aimin Li**, Shaohua Wu, Siqi Meng, Rongxing Lu, Sumei Sun, and Qinyu Zhang, "Towards Goal-Oriented Semantic Communications: New Metric, ", Submitted to IEEE Wireless Communications Magazine.

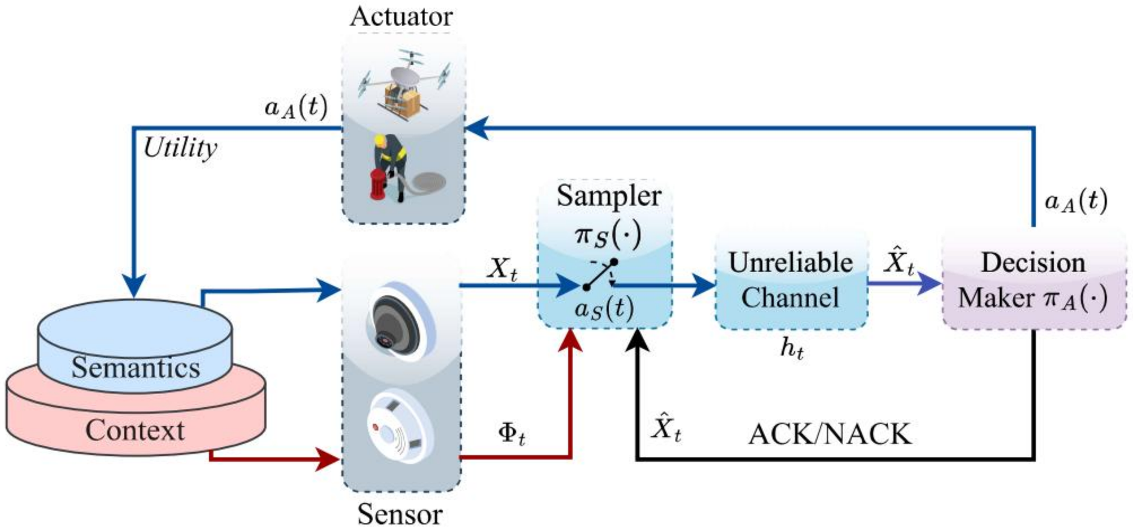
Outline

- Improve the Manuscript
  - Review the System Model
  - Optimality of the RVI-Brute-Force-Search Algorithm
  - Some Investigations of the Nash Equibrium
  - Revision and Improvement on the AIA Algorithm (JESP as a term)
  - Other Minor Improvements
- Limitations and Future Plan
  - Long constant delay or random delay
  - AoI-based prediction enhances the utility

System Model

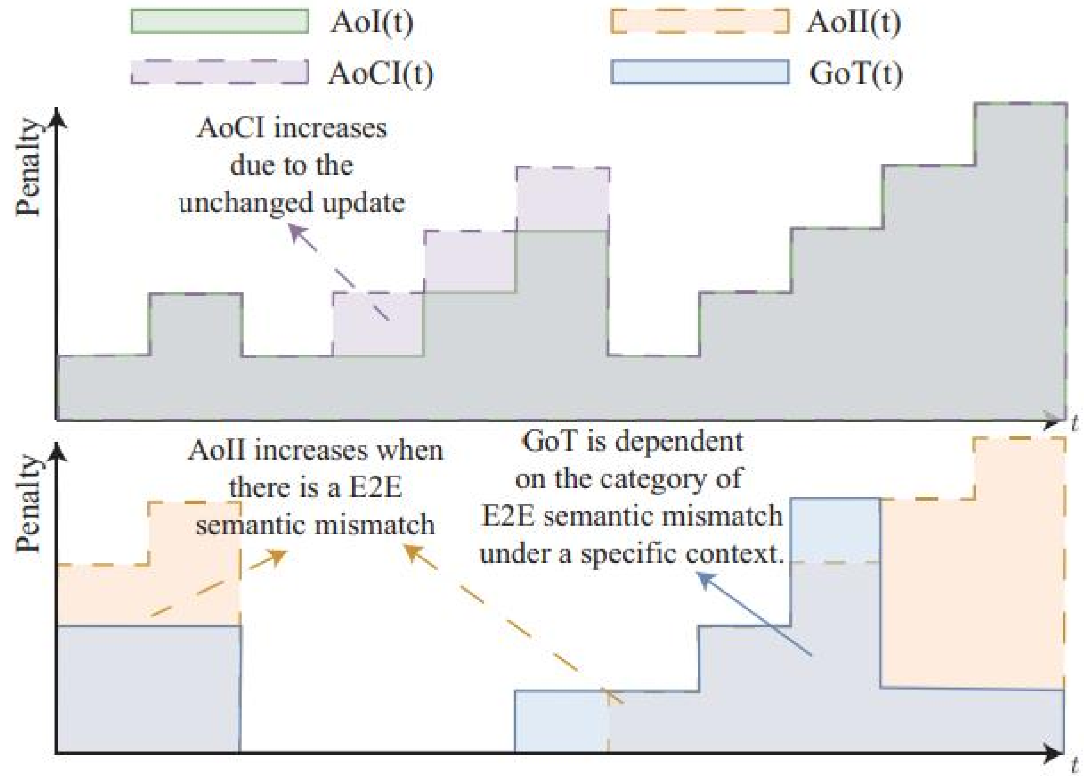
System Model----Sampling-Decision-Making Co-Design

- Sensors extract the semantics X_t and context φ_t from a source.
- The ***sampler*** decides whether the current semantic status should be sampled and transmitted.
- The receiver is informed about the semantic status if the transmission succeed.
- Due to delay and unreliability, $X_t \neq \widehat{X}_t$ is trivial, thus resulting error decision.
- The ***decision maker*** will generate a decision $a_A(t)$ based on the current $\widehat{X}_t$.
- The decision $a_A(t)$ will lead to three aspects of ***utility***:
  - Actuation cost
  - Actuation gain
  - Semantics Evolution

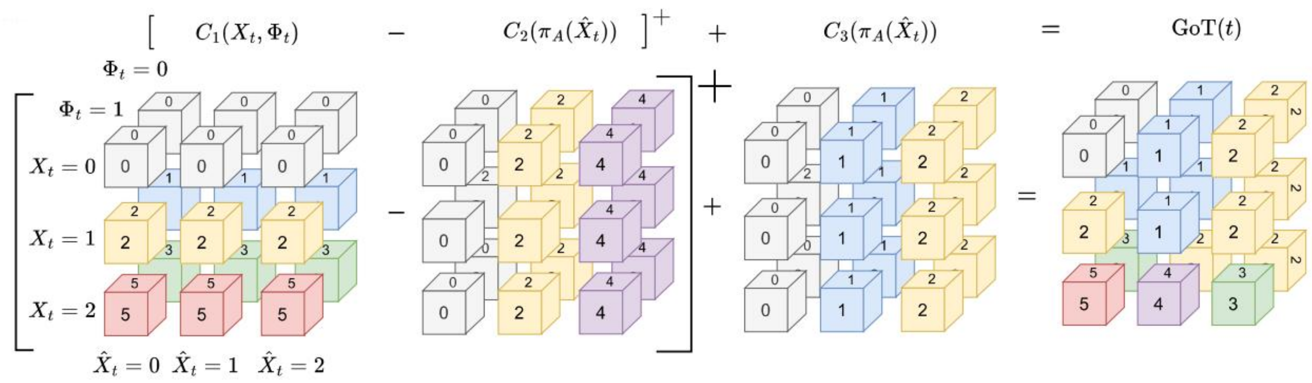


Goal-oriented Tensor (GoT)

- The cost is actually resulted by the error decision making due to status mismatch!
- Three types of cost are considered:
  - Status inherent cost $C_1(X_t, \varphi_t)$
  - Actuation gain $C_2(a_A(t))$
  - Actuation cost $C_3(a_A(t))$

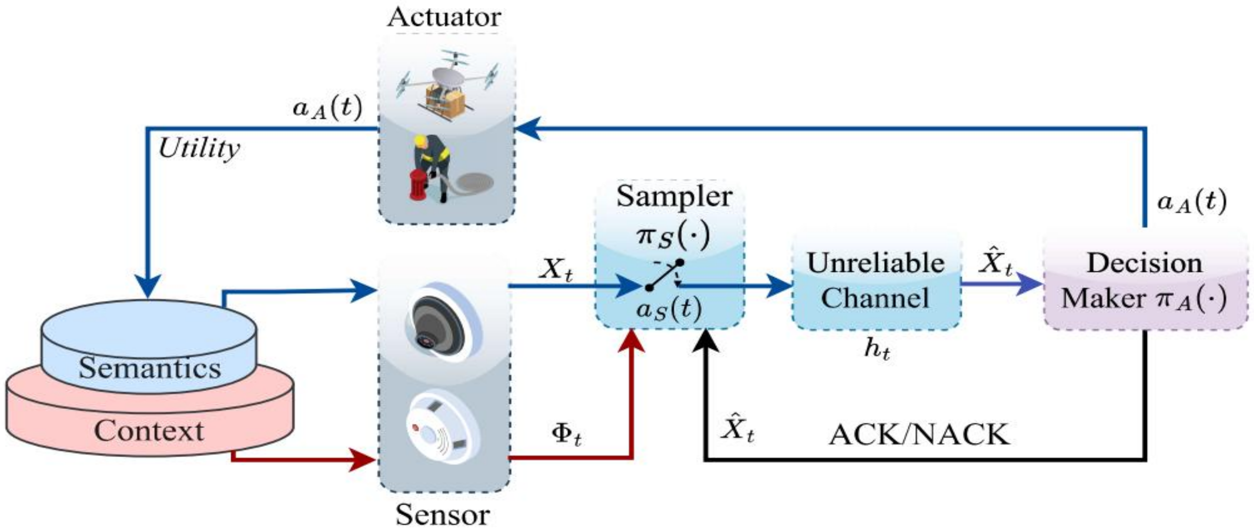


t	0	1	2	3	4	5	6	7	8	9	10
$a_S(t)$	0	1	1	0	0	1	0	1	1	1	0
h_t	\	1	1	\	\	1	\	0	0	0	\
X_t	0	0	0	0	0	1	1	2	2	0	0
$\hat{X}_t$	2	2	0	0	0	0	1	1	1	1	1
Φ_t	0	1	1	1	0	0	0	0	1	1	0



Problem Formulation

- **Goal:** to minimize the long term average cost.
- **Method:** find the optimal deterministic joint sampling-decision-making policy!



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

Joint Policy ($\pi_C \in \Upsilon$, $\mathbb{E}^{\pi_C}$)

Actuation Policy (π_A)

Optimality of the RVI-Brute-Force-Search Algorithm

Optimality of the RVI-Brute-Force-Search Algorithm

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

Joint Policy **Actuation Policy**

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

Proof. *If the the transition function T^{π_A} follows a unichan, we obtain from [57, Theorem 8.4.5] that for any π_A and π_S, we could obtain the optimal deterministic policy π_S^* such that $\theta^*(\pi_A, \pi_S^*) \le \theta^*(\pi_A, \pi_S)$. Also, The Brute-Force-Search assures that for any π_A, we have $\theta^*(\pi_A^*, \pi_S^*) \le \theta^*(\pi_A, \pi_S^*)$. This leads to the conclusion that for any π_S and π_A, we have*

$$\theta^*(\pi_A^*, \pi_S^*) \le \theta^*(\pi_A, \pi_S^*) \le \theta^*(\pi_A, \pi_S)$$

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

Optimality of the RVI-Brute-Force-Search Algorithm

- Brute-Force-Search is computationally expensive, its complexity is $O({\mathcal{A}_A}^{\mathcal{O}_{\mathcal{A}}})$.
- We propose a sub-optimal algorithm based on the search of the Nash equibrium between the sampler and the decision maker.

Nash Equibrium

Nash Equibrium

- Nash equibrium is introduced in team decision theory. (Prisoner's dilemma)
- In the Nash equilibrium, each agent's policy is optimal when considering the decisions of other players.
- In our system setup, denote π_S^N, π_A^N as the Nash equibrium policies. It satisfies that $\forall \pi_A, \theta^*(\pi_A^N, \pi_S^N) \le \theta^*(\pi_A, \pi_S^N)$ and $\forall \pi_S, \theta^*(\pi_A^N, \pi_S^N) \le \theta^*(\pi_A^N, \pi_S)$.
- In this case, we could not assure that $\forall \pi_A, \pi_A, \theta^*(\pi_A^N, \pi_S^N) \le \theta^*(\pi_A, \pi_S)$. It is sub-optimal!
- The search of the Nash equibrium is called Joint Equilibrium-Based Search for Policies (JESP) in the literature, not Alternative Iterative Algorithm (AIA) as called in the original manuscript!

R. Nair, M. Tambe, M. Yokoo, D. V. Pynadath, and S. Marsella, "Taming Decentralized POMDPs: Towards efficient policy computation for multia-gent settings," in Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 2003.

Joint Equilibrium-Based Search for Policies (JESP)

JESP

- JESP Process
  - Step 1: Initialize $\pi_A^{*(1)}$
  - Step 2: Given $\pi_A^{*(k)}$, solve the $\pi_S^{*(k)}$.
  - Step 3: Given $\pi_S^{*(k)}$, solve the $\pi_A^{*(k+1)}$.
  - Step 4: Repeat Step 2 and Step 3 until $||\theta_k^* - \theta_{k-1}^*|| \leq \varepsilon$.
- Generally, the JESP algorithm should **restart the algorithm** by randomly choosing the initial decision-making policy $\pi_A^{*(1)}$ to ensure a good solution.
- The random initialization of decision-making policy may often lead to **poor local optima.**

Algorithm 4: The JESP Algorithm

```
Input: The Dec-POMDP tuple
    𝓜_DEC−POMDP ≜ ⟨n, 𝓘, 𝓐, 𝓣, Ω, 𝓞, 𝓡⟩, ϵ;
Initialization: θ*_0 = 0, θ*_{−1} = ∞, k = 0;
Initialize π_A^{*(1)} by calculating (46);
while ||θ*_k − θ*_{k−1}|| ≥ ϵ do
    k = k + 1;
    Formulate the MDP problem
      𝓜_MDP^{π_A^{*(k)}} ≜ ⟨𝓘, 𝓐_S, 𝓣_MDP^{π_A^{*(k)}}, 𝓡⟩ as given in
      Definition 1;
    Run Algorithm 1 to obtain π_S^{*(k)};
    Formulate the POMDP problem
      𝓜_POMDP^{π_S^{*(k)}} ≜ ⟨𝓘, 𝓐_A, Ω_A, 𝓞_A, 𝓣_POMDP^{π_S^{*(k)}}, 𝓡⟩ as
      given in Definition 2;
    Run Algorithm 3 to obtain θ*(π_S^{*(k)}, π_A^{*(k)}) and
      π_A^{*(k)};
    θ*_k = θ*(π_S^{*(k)}, π_A^{*(k)});
The joint sub-optimal policy is:
  π*_S = π_S^{*(k)}, π*_S = π_A^{*(k)};
The sub-optimal average reward is: θ* = θ*_k;
Output: π*_S, π*_A, θ*
```

Improved JESP

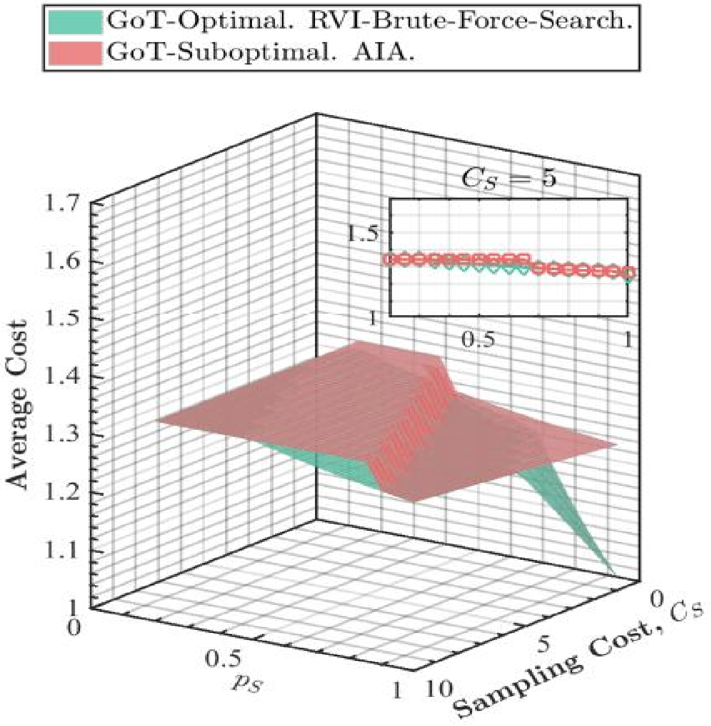
- We propose a method to heuristicly initialize $\pi_A^{*(1)}$, which achieves excellent solutions:
  - We assume that the decision maker is ***fully observable*** and formulate a MDP problem to initialize $\pi_A^{*(1)}$
  - **Definition 3**. Fully observable MDP to Intialize $\pi_A^{*(1)}$: $\mathcal{M} = \langle \mathcal{I}, \mathcal{A}_A, \mathcal{T}, \mathcal{R} \rangle$
    - $\mathcal{I}$: the set of (X_t, φ_t).
    - $\mathcal{A}_A$: the decision making space.
    - $\mathcal{T}$: the transition function.
    - $\mathcal{R}$: the reward function.
  - Applying RIV algorithm to solve out $\mathcal{M} = \langle \mathcal{I}, \mathcal{A}_A, \mathcal{T}, \mathcal{R} \rangle$, and we have the intial $\pi_A^{*(1)}$.

Simulation Results

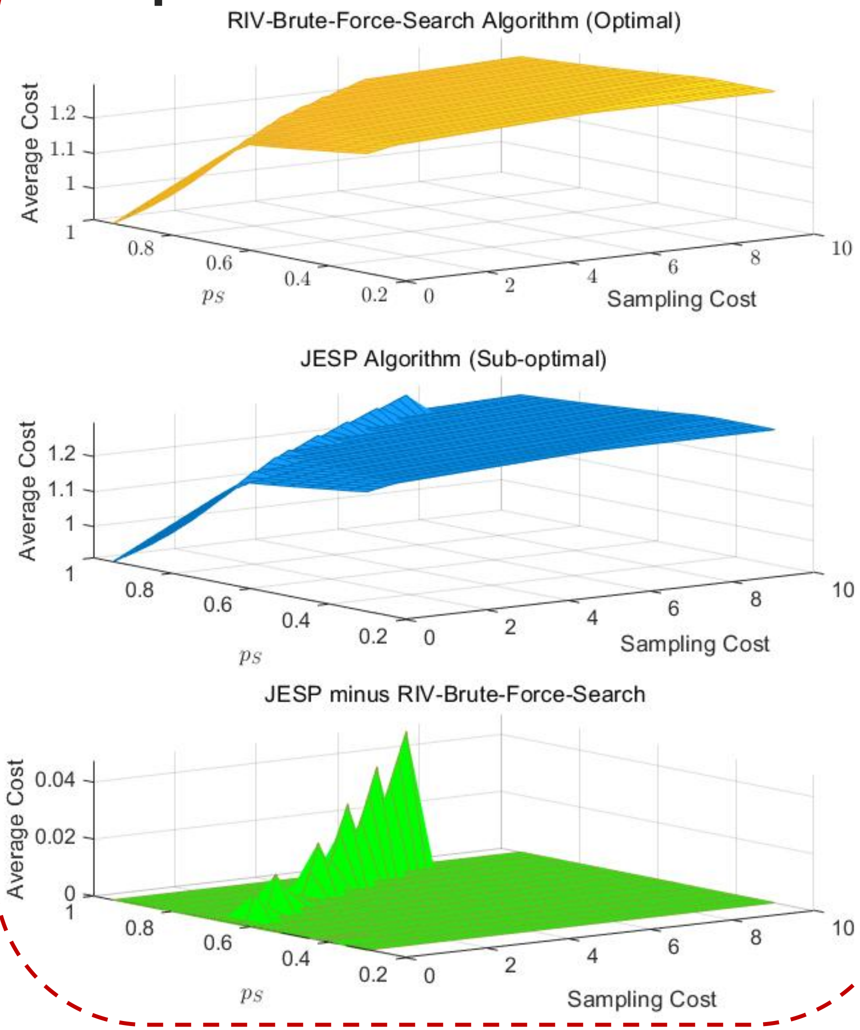
- Compared to the original random intialization, the improved heuristic intialization is **closer to the *optimal solution*!**

- Under different sampling costs and successful transmission probability p_S, the improved sub-optimal JESP could achieve a good solution!

Original Intialization



Improved Intialization

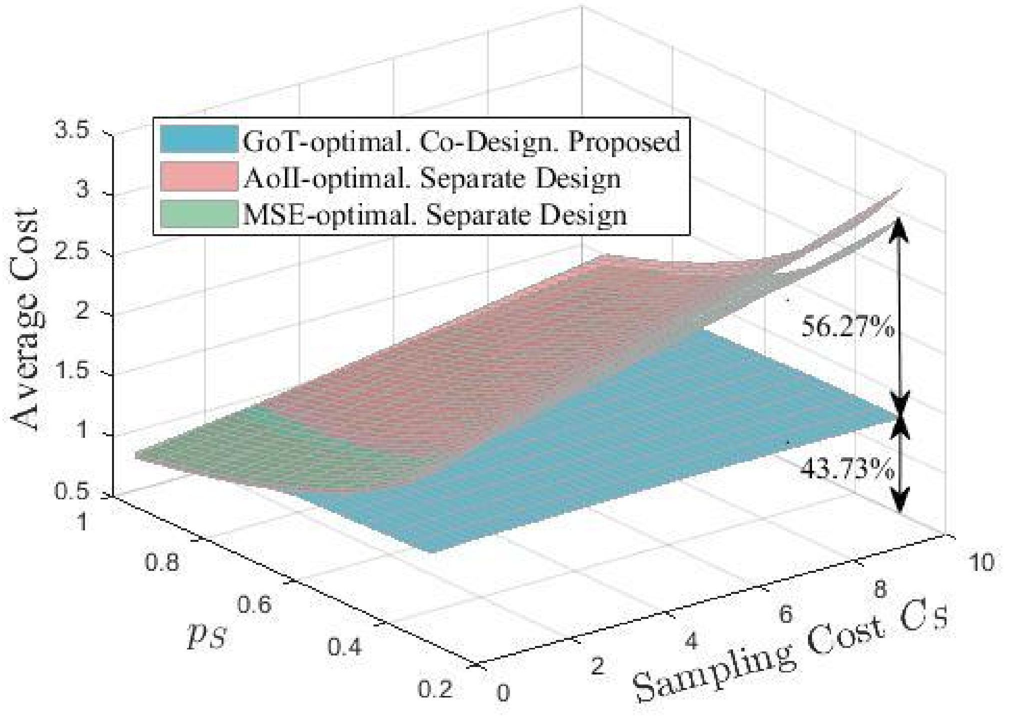


Simulation Results

- GoT-optimal ***sampling-decision-making co-design*** could achieve the goal of communicationg best!

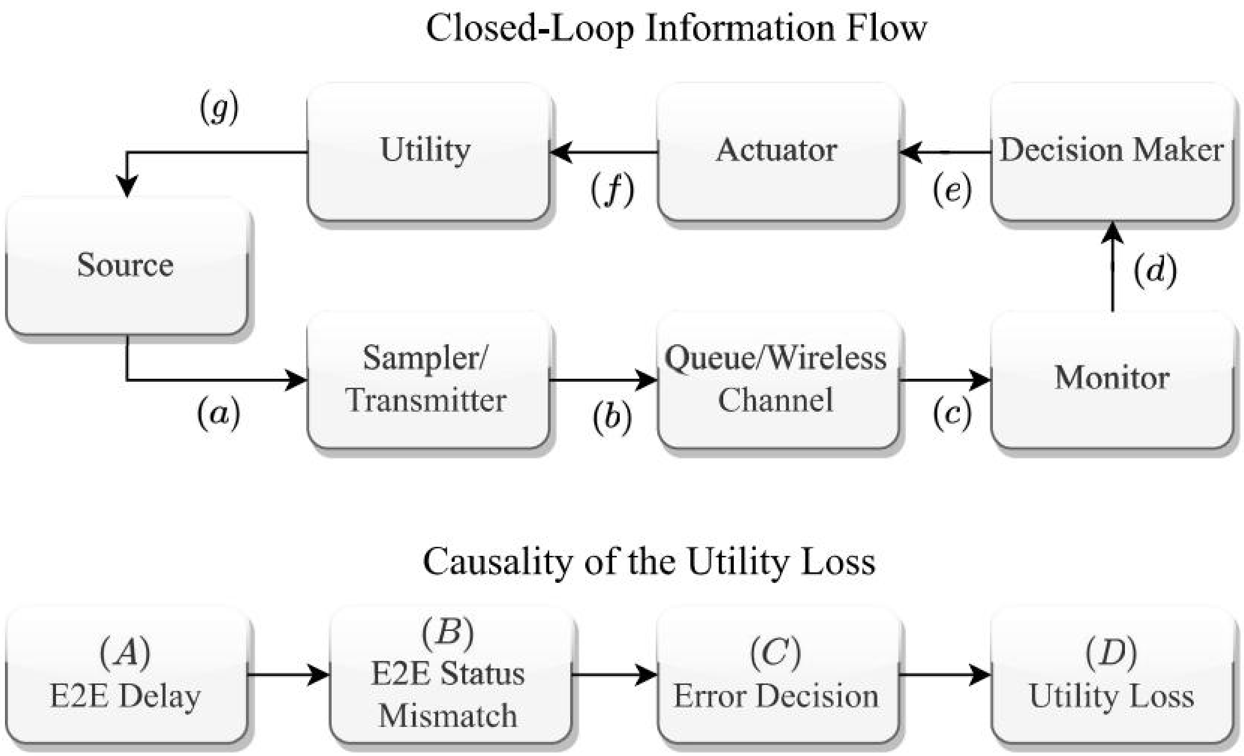
- MSE-optimal is not the end...
  - Achieving E2E reconstruction is not the ultimate goal of communications.

- When the sampling cost C_S is high and the channel is unreliable (p_S is low), it is more effective to ***assign more resources to the actuator rather than communicating.***



Other Minor Improvements on the Manuscript

Other Minor Improvements - Intorduction



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

- Information is transmitted from source to receiver to **achieve the goal by decision maker**.
- The causality of the utility loss...

Limitations and Future Plan

Limitations and Future Plan

- Consider random long delay
  - The current delay is set as a constant 1:
    - ***How to achieve the goal even when the delay is non-trivial?***
    - ✓ ***AoI-based prediction and decision-making may enhance the utility.***
  - ***Goal-oriented Communications (Sampling-dicision-making co-design) under nontrivial delay.***

Y Sun, Y Polyanskiy, E Uysal. "Sampling of the Wiener process for remote estimation over a channel with random delay," IEEE Transaction on Information Theory.

Y Chen, A Ephremides, "Minimizing Age of Incorrect Information over a Channel with Random Delay", arxiv, 2023

Pan, Jiayu, et al. "Optimal sampling for data freshness: Unreliable transmissions with random two-way delay." IEEE/ACM Transactions on Networking, 2022.

Tsai C H, Wang C C. Unifying AoI minimization and remote estimation—Optimal sensor/controller coordination with random two-way delay. IEEE/ACM Transactions on Networking, 2021.

Thanks for listening!!