GOAL-ORIENTED COMMUNICATION FOR REAL-TIME TRACKING IN AUTONOMOUS SYSTEMS

Nikolaos Pappas[1], *Marios Kountouris*[2]

[1]Department of Science and Technology, Linköping University, Norrköping 60174, Sweden
[2]Communication Systems Department, EURECOM, Sophia-Antipolis 06904, France
Email: nikolaos.pappas@liu.se, marios.kountouris@eurecom.fr

ABSTRACT

Real-time remote tracking using under-sampled and delayed measurements is considered here. We study an autonomous system where a transmitter monitors the evolution of a discrete Markov source and sends status updates to a destination over an unreliable wireless channel. The destination is tasked with real-time source reconstruction for remote actuation. We introduce new goal-oriented sampling and communication policies, which leverage the significance and effectiveness of messages, as a means to generate and transmit only the most "informative" samples for real-time actuation. Our results illustrate that semantics-empowered policies significantly reduce both the real-time reconstruction and the cost of actuation errors, as well as the amount of ineffective updates.

Index Terms— Goal-oriented communication, semantics of information, cost of actuation error, real-time tracking.

1. INTRODUCTION

Networked autonomous systems are ubiquitous in various domains, with applications spanning from swarm robotics and healthcare to autonomous transportation and environmental monitoring. These systems require reliable real-time communication, autonomous interactions, and timely computations. In this context, information is valuable when it is fresh, accurate, and effective. For instance, real-time knowledge of the trajectory and the velocity of a mobile robot is essential in autonomous navigation. Timely and accurate updates are of cardinal importance in mission-critical decision-making. In this setting, real-time tracking and reconstruction of an information source/process from a set of under-sampled and delayed measurements is an important yet challenging problem. Information freshness is assessed by the Age of Information (AoI) [1, 2], i.e., the time elapsed since the newest successfully received update was generated. However, AoI does not take into account the source/process evolution and the application context. Several metrics [3–7] have been employed to address the shortcomings of AoI. A line of work that considers AoI and its variants as a criterion for remote estimation can be found in [8–14]. A new communication paradigm has recently been proposed, which accounts for the *semantics* (i.e., significance, goal-oriented usefulness, and contextual value) of information and leverages the synergy between data processing, information transmission, and signal reconstruction [15–17].

In this work, we consider the problem of real-time tracking and reconstruction of an information source. A transmitter (observer) tracks the state of a Markov source and sends status updates (samples) to a receiver over an unreliable wireless channel. Real-time reconstruction is performed at the destination for the purpose of remote actuation. This fundamental setting could model various real-time applications in autonomous networked systems. We introduce new goal-oriented, semantics-empowered sampling and communication policies, which account for the temporal evolution of the source/process and the semantic and application-dependent value of data being generated and transmitted. Our approach is shown to significantly reduce both reconstruction error and cost of actuation error, as well as the number of uninformative/ineffective samples.

2. SYSTEM MODEL

We consider a time-slotted communication system, in which a monitoring device (transmitter) observes a process X_t and informs a remote actuator (receiver) about its state by sending updates (samples) over a communication channel. This is depicted in Fig. 1. We assume that transmissions take place over a wireless erasure channel. The channel realization h_t is equal to 1 if the packet is successfully decoded at the receiver and 0 otherwise. The success probability is defined as $p_s = \mathbb{P}(h_t = 1)$. Successful/failed transmissions are declared to the transmitter using acknowledgement (ACK)/negative ACK packets, assumed to be delivered instantaneously and error-free. When the transmission fails, the update is discarded (no retransmission). The information source is modeled by a two-state discrete-time Markov chain (DTMC) $\{X_t, t \in \mathbb{Z}_0^+\}$, as-

This work is supported in part by the Swedish Research Council (VR), ELLIIT, and CENIIT. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 101003431).

sumed to be ergodic. At each timeslot t, the state X_t of the source can be either 0 or 1. The self-transition probabilities are denoted by $1-p$ and $1-q$ for states 0 and 1, respectively. Therefore, $\mathbb{P}\left(X_{t+1}=i|X_t=i\right)=\mathbb{1}\left(i=0\right)(1-p)+\mathbb{1}\left(i=1\right)(1-q)$, where $\mathbb{1}(\cdot)$ is the indicator function.

The transmitter is capable of generating update X_t by sampling the source at will. The action of sampling at timeslot t is denoted by α_t^{s}, where $\alpha_t^{\mathrm{s}}=1$ if the source is sampled and $\alpha_t^{\mathrm{s}}=0$ otherwise. The action of transmitting a sample is denoted by $\alpha_t^{\mathrm{tx}}=1$, otherwise the transmitter remains silent ($\alpha_t^{\mathrm{tx}}=0$). The source is reconstructed at the destination based on successfully received status updates. The state of the reconstructed source at timeslot t is denoted as $\hat{X}_t$. A set of operation policies for sampling and transmission are presented in Section 4.

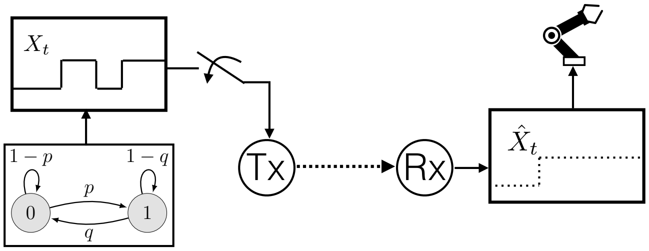


Fig. 1. Real-time tracking of a two-state Markov source.

3. KEY PERFORMANCE METRICS

In this section, we introduce a new, comprehensive system metric, namely the Semantics of Information (SoI), which captures the significance and usefulness of information with respect to the goal of data exchange and the application requirements. We then present two performance metrics used to evaluate the accuracy of real-time reconstruction and its effect on the actuation in our problem under study.

3.1. Semantics-empowered metrics

Let $\mathcal{I}\in\mathbb{R}^m$ denote the vector of m information attributes, which can be decomposed into *innate* (objective) and *contextual* (subjective). Innate are the attributes inherent to information regardless of its use, such as *freshness* or AoI, i.e., $\Delta_t=t-U_t$ where U_t is the generation time of the newest sample that has been delivered at the destination by time instant t, *precision*, and *correctness*. Contextual are attributes that depend on the particular context or application for which information is being used. The most relevant ones are *timeliness*, a function $g:\mathbb{R}_{\geq 0}\rightarrow\mathbb{R}$ of AoI, i.e., $g(\Delta_t)$, and *completeness*. Another relevant attribute is *accuracy (distortion)* $\delta:\mathcal{X}\times\mathcal{X}\rightarrow\mathbb{R}_{\geq 0}$ where $\mathcal{X}$ is the state space of X_t. For example, $\delta(X_t,\hat{X}_t)=(X_t-\hat{X}_t)^2$. We can also include the notion of *perception* via some divergence or distance function $D(\cdot\|\cdot):\mathcal{D}\times\mathcal{D}\rightarrow\mathbb{R}$ between probability distributions defined in the same probability space $\mathcal{D}$ (e.g., Wasserstein, Levenshtein, Hellinger, etc., depending on the application).

Formally, SoI is a composite function $\mathcal{S}_t=\nu(\psi(\mathcal{I})$, where $\psi:\mathbb{R}^m\rightarrow\mathbb{R}^z, m\geq z$ is a nonlinear function and $\nu:\mathbb{R}^z\rightarrow\mathbb{R}$ is a context-dependent, cost-aware function that maps qualitative information attributes to their application-dependent value. Below, we provide two such metrics, which are relevant to remote real-time tracking and actuation.

3.2. Real-time reconstruction error

The real-time reconstruction error measures the discrepancy between the original X_t and the reconstructed source $\hat{X}_t$ at timeslot t, i.e., $E_t=\mathbb{1}\left(X_t\neq\hat{X}_t\right)=\left|X_t-\hat{X}_t\right|$. In other words, $\delta(\cdot,\cdot)$ is the $0-1$ loss function or the Hamming distortion measure. For a two-state DTMC, E_t takes values 0 or 1. Our analysis can easily be generalized to an N-state Markov source ($N<\infty$), in which case the reconstruction error can take any value from 0 to N. The system can be either in an erroneous state ($E_t=1$) or in a synced state ($E_t=0$). The time-averaged reconstruction error is given by

$$\bar{E}=\lim_{T\rightarrow\infty}\frac{\sum_{t=1}^{T}E_t}{T}=\lim_{T\rightarrow\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left(X_t\neq\hat{X}_t\right). \quad (1)$$

The evolution of the state of the system, E_t, can be described by a Markov Chain as depicted in Fig. 2. The synced state is denoted by 0 ($E_t=0$ at timeslot t), whereas 1 denotes the erroneous state. The one-step transitions probabilities are defined as

$$p_{ji}=\mathbb{P}\left(E_{t+1}=j|E_t=i\right), \forall i,j\in\{0,1\}. \quad (2)$$

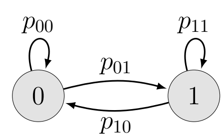


Fig. 2. DTMC describing the state E_t of the system.

We now give general expressions for the transition probabilities. To obtain p_{00} we need to calculate

$$p_{00}=\sum_{i=0}^{1}\mathbb{P}\left(E_{t+1}=0|E_t=0,X_t=i\right)\mathbb{P}\left(X_t=i\right), \quad (3)$$

with $\mathbb{P}\left(X_t=i\right)=\mathbb{1}\left(i=0\right)\frac{q}{p+q}+\mathbb{1}\left(i=1\right)\frac{p}{p+q}$ and

$$\begin{aligned}&\mathbb{P}\left(E_{t+1}=0|E_t=0,X_t=0\right)=\\&=1-p+p\mathbb{P}\left(\alpha_{t+1}^{\mathrm{s}}=1,\alpha_{t+1}^{\mathrm{tx}}=1,h_{t+1}=1\right). \quad (4)\end{aligned}$$

In a similar way, we obtain $\mathbb{P}\left(E_{t+1}=0|E_t=0,X_t=1\right)$. We also have

$$\begin{aligned}&\mathbb{P}\left(E_{t+1}=1|E_t=1,X_t=0\right)=(1-p)\\&\times\left[\mathbb{P}\left(\alpha_{t+1}^{\mathrm{s}}=1,\alpha_{t+1}^{\mathrm{tx}}=1,h_{t+1}=0\right)+\mathbb{P}\left(\alpha_{t+1}^{\mathrm{s}}=0\right)\right]. \quad (5)\end{aligned}$$

The exact values of transition probabilities depend on the sampling and transmissions actions, dictated by the policies introduced in Section 4. We can also compute the stationary distribution of this two-state DTMC, where π_0 is the probability the system is synced (or the percentage of time the system is synced), and π_1 the probability the system is in an erroneous state.

3.3. Cost of actuation error

The second performance metric is the cost of actuation error, which captures the significance of the error at the point of actuation. Note that some errors may have larger impact than others. At timeslot t, $C_{i,j}$ denotes the cost of being in state i at the original source and in $j \neq i$ at the reconstructed, i.e., $E_t = 1$. Whenever $i = j$, there is no error, and consequently no cost. We consider the general and practically relevant case of non-commutative errors, i.e., $C_{0,1} \neq C_{1,0}$. This means that different erroneous actions may have different cost (penalty) due to different repercussions for the system performance. We assume that C_{ij} does not change over time.[1]

In order to calculate the average cost of actuation error, we can use a two-dimensional Markov chain describing the joint status of the system regarding the current state at the original source and whether the reconstructed source is synced or not. That way, we can obtain expressions for the average real-time reconstruction error and the average cost of actuation error. The latter can be written as

$$\bar{C}_A = \pi_{(0,1)} C_{0,1} + \pi_{(1,0)} C_{1,0} \tag{6}$$

where $\pi_{(i,i+1 \pmod 2)}$ is obtained from the stationary distribution of the two-dimensional DTMC.

Remark. The above Markov chain formulation provides a very general view of the system, which can be used to derive optimal online policies using Markov Decision Processes or Deep Reinforcement Learning.

4. GOAL-ORIENTED SAMPLING AND COMMUNICATION POLICIES

In this section, we propose two semantics-empowered policies of information acquisition (sampling) and transmission for real-time reconstruction of a Markov source with the purpose of actuation. We start by presenting two conventional policies, which are used for comparison purposes. Due to space limitations, expressions only for the last goal-oriented policy are provided; analyzing the other policies involves a modification of the general expressions given in Section 3.

4.1. Uniform

In this baseline policy, sampling is performed periodically, independently of the temporal evolution of the source. Despite being simple and easy to implement, process-agnostic policies could result in missing several state transitions during the time interval between two collected samples. In the case of erasure, the most recently acquired sample is transmitted.

4.2. Age-aware

In this policy, the receiver triggers the acquisition and transmission of a new sample, once the AoI reaches a predefined threshold A_{th}. This can be extended to different AoI thresholds depending on the state [18].

Whenever a transmission fails, the receiver tries to anticipate the update based on the statistics of the source process. In that case, the receiver, given its current state, tries to predict the next state based on the state transition probabilities, assumed to be known (or learned after a period of time). This policy remains source-agnostic regarding the value of information but takes into account the timeliness.

4.3. Change-aware

In this policy, sample generation is triggered at the transmitter whenever a change at the state of the source (with respect to the previous sample) is observed. Consider that for a certain period of time $X_{t+kt} = i, k = 0, 1, \ldots, K$, at the end of which the state changes, i.e., $X_{t(K+1)+1} = j, j \neq i$, and hence the transmitter generates and transmits a new status update sample.

4.4. Semantics-aware

This policy extends the previous one into that the amount of change is not solely measured at the source, but is also tracked by the difference in state between receiver and transmitter. Sample generation is triggered whenever there is discrepancy between X_t and $\hat{X}_t$. Assume that at a given timeslot t, $X_t = \hat{X}_t$. Then, a change in X_t (source's state) occurs in $t+1$ with a given probability, resulting in transmission of a newly acquired sample. If transmission fails, $\hat{X}_{t+1} = \hat{X}_t$. Suppose now that in the next slot, $X_{t+2} = X_t$. Thus, there is no discrepancy between the original and the reconstructed source ($\hat{X}_{t+2} = X_{t+2}$), thus, there is no need for sending an update. Particularizing the general expressions in Section 3, the transition matrix P_E for the DTMC that models the system state is given by

$$\mathrm{P}_E = \begin{bmatrix} 1 - \frac{2pq(1-p_s)}{p+q} & \frac{2pq(1-p_s)}{p+q} \\ p_s + \frac{2pq(1-p_s)}{p+q} & 1 - p_s - \frac{2pq(1-p_s)}{p+q} \end{bmatrix}.$$

Then, we obtain the the probability that the system is in an erroneous (not synced) state

$$\bar{E} = \frac{2pq(1-p_s)}{p_s(p+q) + 4pq(1-p_s)}. \tag{7}$$

Note that the previous expression is also the percentage of time that the system is not synced or the time-average reconstruction error.

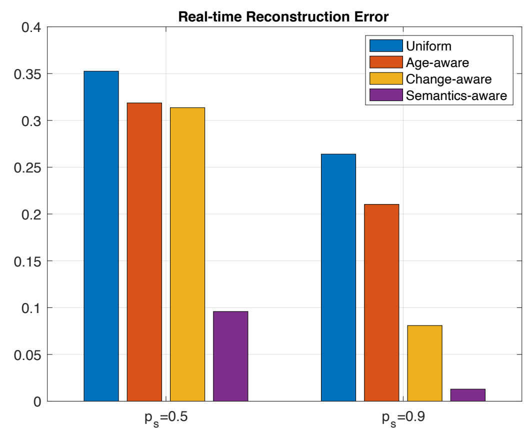
Remark. Evidently, sampling and transmission at every timeslot could provide the best performance for achieving the goal (real-time reconstruction). However, this comes at the

[1] Penalty functions based on non-linear aging [3], where the cost of being in an erroneous state is increasing over time, can be employed.

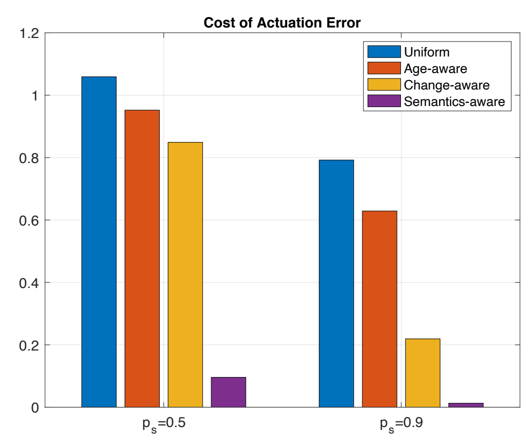
expense of very large number of samples, which are not necessarily useful and which require excessive resources (e.g., energy, network) for their acquisition, transmission, and processing. The proposed semantics-empowered policies reduce or even eliminate the generation of uninformative sample updates, thus improving network resource usage and being scalable.

5. NUMERICAL RESULTS

We evaluate the performance of above policies in terms of average real-time reconstruction error and average cost of actuation error. We consider two scenarios regarding the source variability, the first being when the source is slowly changing ($p = 0.1$, $q = 0.15$), depicted in Fig. 3, and the second being when the source is rapidly changing ($p = 0.2$, $q = 0.7$), depicted in Fig. 4. In addition, regarding the probability of successful transmission we consider two distinct cases, $p_s = 0.5$ and $p_s = 0.9$, to investigate the impact of transmission errors on reconstruction and actuation. In uniform sampling, a sample is acquired every 5 timeslots, and for the age-aware policy we set $A_{\mathrm{th}} = 5$. Additionally, the actuation errors are $C_{0,1} = 5$ and $C_{1,0} = 1$.



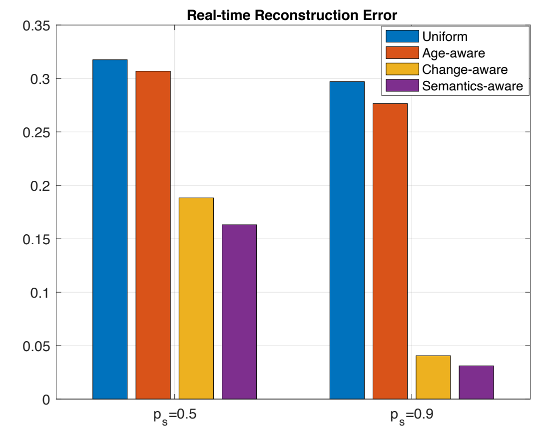
(a) Real-time reconstruction error



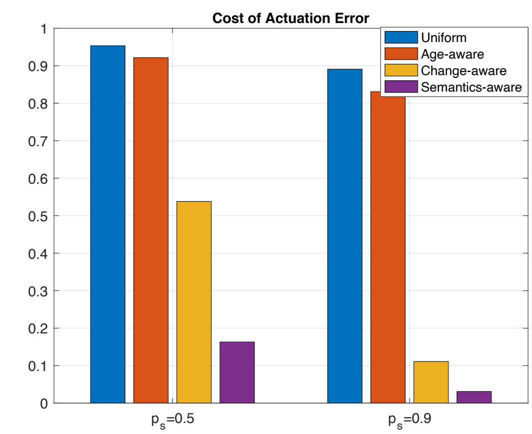
(b) Cost of actuation error

Fig. 3. Slowly varying source with $p = 0.1$ and $q = 0.15$.

For slow varying sources, the semantics-aware policy significantly outperforms the change-aware policy due to the fact that the system manages to rapidly eliminate the discrepancy between the original and the reconstructed state of the source, even in the case of low channel quality ($p_s = 0.5$). On the other hand, for rapidly varying sources, both semantics- and change-aware policies exhibit very good reconstruction error performance, while semantics-aware provides the lowest actuation error *without wasting any resources for transmitting uninformative samples*.



(a) Real-time reconstruction error



(b) Cost of actuation error

Fig. 4. Rapidly changing source with $p = 0.2$ and $q = 0.7$.

6. CONCLUSION

In this work, we showcased the potential of goal-oriented data generation and communication policies in a remote real-time tracking and actuation scenario. Accounting for the semantics of information semantics could lead to significant reduction in task-dependent error metrics and in ineffective traffic volume.

7. REFERENCES

[1] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Foundations and Trends in Networking*, vol. 12, no. 3, 2017.

[2] R. D. Yates, Y. Sun, D. Richard Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of Information: An introduction and survey," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, 2021.

[3] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The cost of delay in status updates and their value: Non-linear ageing," *IEEE Trans. on Communications*, vol. 68, no. 8, 2020.

[4] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *Journal of Communications and Networks*, vol. 21, no. 3, 2019.

[5] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, 2020.

[6] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, "Age-of-information vs. value-of-information scheduling for cellular networked control systems," in *10th ACM/IEEE ICCPS*, Apr. 2019.

[7] A. Molin, H. Esen, and K. H. Johansson, "Scheduling networked state estimators based on value of information," *Automatica*, vol. 110, 2019.

[8] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," in *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, 2019.

[9] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the Wiener process for remote estimation over a channel with random delay," *IEEE Transactions on Information Theory*, vol. 66, no. 2, 2020.

[10] C. Kam, S. Kompella, and A. Ephremides, "Age of incorrect information for remote estimation of a binary markov source," in *IEEE INFOCOM Workshops*, 2020.

[11] C. Tsai and C. Wang, "Unifying AoI minimization and remote estimation—optimal sensor/controller coordination with random two-way delay," in *IEEE Conference on Computer Communications (INFOCOM)*, 2020.

[12] K. Huang, W. Liu, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Real-time remote estimation with hybrid arq in wireless networked control," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, 2020.

[13] M. Wang, W. Chen, and A. Ephremides, "Real-time reconstruction of a counting process through first-come-first-serve queue systems," *IEEE Transactions on Information Theory*, vol. 66, no. 7, 2020.

[14] X. Zheng, S. Zhou, and Z. Niu, "Urgency of information for context-aware timely status updates in remote control systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, 2020.

[15] P. Popovski, O. Simeone, F. Boccardi, D. Gündüz, and O. Sahin, "Semantic-effectiveness filtering and control for post-5G wireless connectivity," *Journal of the Indian Institute of Science*, vol. 100, no. 2, 2020.

[16] E. Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Computer Networks*, p. 107930, 2021.

[17] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *to appear, IEEE Commun. Magazine*, 2021.

[18] G. Stamatakis, N. Pappas, and A. Traganitis, "Control of status updates for energy harvesting devices that monitor processes with alarms," in *IEEE Globecom Workshops*, 2019.