Semantic Source Coding for Two Users with Heterogeneous Goals

Pouya Agheli*, Nikolaos Pappas†, and Marios Kountouris*
*Communication Systems Department, EURECOM, Sophia Antipolis, France
†Department of Computer and Information Science, Linköping University, Sweden
Email: pouya.agheli@eurecom.fr, nikolaos.pappas@liu.se, marios.kountouris@eurecom.fr

***Abstract*—We study a multiuser system in which an information source provides status updates to two monitors with heterogeneous goals. Semantic filtering is first performed to select the most useful realizations for each monitor. Packets are then encoded and sent so that each monitor can timely fulfill its goal. In this regard, some realizations are important for both monitors, while every other realization is informative for only one monitor. We determine the optimal real codeword lengths assigned to the selected packet arrivals in the sense of maximizing a weighted sum of semantics-aware utility functions for the two monitors. Our analytical and numerical results provide the optimal design parameters for different arrival rates and highlight the improvement in timely status update delivery using semantic filtering and source coding.**

I. Introduction

Goal-oriented semantic communication has recently been considered as a promising and timely research avenue towards realizing the long-standing vision of Shannon and Weaver [1] through incorporating the significance and the importance of information into the existing theoretic edifice. Despite various past endeavors [2]–[5], which remained at a conceptual level, leading to hardly any or no practically relevant applications, the quest for such theory has recently gained new impetus [6], [7], fueled by the emergence of networked intelligent systems and autonomous networks. This communication paradigm has the potential to render various network processes more efficient and effective, providing a parsimonious usage of communication and computation resources. Key to this is the definition of a concise yet insightful metric of the importance or the usefulness of information. Age of information (AoI) [8], [9], which measures information freshness in networks, and value of information (VoI) [10], [11], which quantifies the information utility or gain in decision making, can be viewed as simple, quantitative surrogates for information semantics.

In this paper, we investigate a *multiuser* system in which two monitors (users) receive status updates from a transmitter observing an information source. The updates may correspond to observations or measurements of a random phenomenon and are generated from a known discrete distribution with finite support. Semantic filtering is first performed as a means to select only the most important or valuable realizations according to each monitor's goal. Status updates are then encoded and sent to the monitor(s) over packet erasure channels (PECs). To harness packet transmission failures, an automatic repeat request (ARQ) protocol is applied. Our objective is to design a timely source coding scheme for two users with heterogeneous goals. Specifically, we consider that only a fraction of the "least" ("most") frequent source realizations is important for the first (second) monitor. This setting models for instance the case in which one user is interested in regular/standard information for monitoring purposes or typical actuation (normal mode), whereas the other monitor tracks the outliers that could potentially represent some kind of threat to the system or a possibly dangerous situation (alarm mode). The notion of semantics (importance) is captured here through a metric of timeliness, which is a nonlinear function of AoI, for the received updates at both monitors.

This work falls within the realm of source coding problem for status update systems in which the goal is to minimize the average age of information, such as in [12]–[15]. In [16], a semantics-aware encoding scheme for a single user over an error-free point-to-point status update link is proposed. Our paper extends prior work into multiuser systems with heterogeneous goals, which could also be competing or diverging for certain realizations. Specifically, we derive the optimal real codeword lengths that maximize a weighted sum of the semantics-aware utility functions for two heterogeneous monitors. Our analytical and numerical results characterize the promising performance gains by properly designing semantic filtering and source coding for timely status update delivery.

II. System Model

We consider a multiuser network in which a transmitter observes a physical phenomenon/event and sends status updates (samples) to two monitors (named 1 and 2, c.f. Fig. 1). Samples from the information source/process are generated in the form of packets carrying different realizations from a finite set $\mathcal{X} = \{x_i \mid i \in \mathcal{I}_n\}$, $\mathcal{I}_n = \{1, 2, ..., n\}$, each having a probability of occurrence $\tilde{p}_i = P_X(x_i)$ where $P_X(\cdot)$ is a known probability mass function (pmf). The observation sequence is independent and identically distributed (i.i.d.) and

The work of P. Agheli and M. Kountouris has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101003431). The work of N. Pappas is supported by the Swedish Research Council (VR), ELLIIT, and CENIIT.

packets arrivals are Poisson distributed with rate λ. Only the most *important* status updates are transmitted, hence a semantic filter admits the k most important arrivals and discards the rest, i.e., $n-k$, done via a flow controller. The difference between the two monitors (receivers) lies in their significance/value assessment policies, i.e., one observation can be essential for one monitor while not relevant for the other. However, some observations are important or valuable to *both* monitors. The importance (semantics) of each arrival is assumed to be related to its probability of occurrence. Specifically, we consider that monitor 1 is only interested in the set of the k_1 least frequent arrivals ($\mathcal{I}_{k_1}$), whereas the set of the k_2 most frequent ones ($\mathcal{I}_{k_2}$) is highly important for monitor 2. Thus, the index set of the admitted packets, i.e., $\mathcal{I}_k \subseteq \mathcal{I}_n$, is defined as $\mathcal{I}_k = \mathcal{I}_{k_1} \cup \mathcal{I}_{k_2}$. In this regard, some arrivals are important for both monitors with some non-zero probability. A source encoder assigns codewords of length ℓ_i to the admitted status packets x_i, $\forall i \in \mathcal{I}_k$, using prefix-free coding, based on the truncated distribution calculated as follows $p_i = \tilde{p}_i/q_k, \forall i \in \mathcal{I}_k$ (and zero otherwise), where $q_k = \sum_{i\in\mathcal{I}_k} \tilde{p}_i$.

Assuming no buffer is employed at the transmitter, an admitted status arrival is blocked when the channel is busy. Besides, it is assumed that each packet is delivered to both monitors at the same time. Transmission occurs over a noisy network, which is modeled by two identical packet erasure channels with erasure probability δ.[1] A simple ARQ protocol is used for fixing potential transmission errors. In the event of failure at either monitor, the transmitter, upon receipt of negative feedback, retransmits the same packet to both monitors. All propagation and decoding delays are considered negligible.

After successfully transferring each packet, the transmitter waits for a new admitted arrival. We define $t_{r,j-1}$ the time instant that the j-th packet is received at both monitors concurrently but is valuable for monitor 1 ($r=1$) and/or monitor 2 ($r=2$). The update interval between the j-th successive and valuable arrival and the next one at the same monitor is then modeled as a random variable (r.v.) $Y_{r,j} = t_{r,j} - t_{r,j-1}$. This interval consists of the service time $S_{r,j}$ and the waiting time $W_{r,j}$, such that $Y_{r,j} = W_{r,j} + S_{r,j}$. $W_{r,j}$ is the time between the transmitted j-th packet and its previously delivered packet at the same monitor; thus, the waiting time is written as $W_{r,j} = \sum_{m'=1}^{M_{r,j}} Z_{r,m'}$, where $M_{r,j}$ is an r.v. of the number of admitted arrivals that are generated before finding the channel idle for the j-th packet. Also, $Z_{r,m'}$ is the time between two arrivals and is exponentially distributed with rate λ. Thus, the admitted arrivals for each monitor are generated under a Poisson process with rate λq_{k_r}. Let us define $q_{k_r} = \sum_{i\in\mathcal{I}_{k_r}} \tilde{p}_i \leq q_k$. The transmission time is proportional to the codeword length and the number of ARQ repetitions. Thereby, the service time of realization x_i, $\forall i \in \mathcal{I}_{k_r}$, which is counted important for monitor 1 and/or 2 during the j-th arrival becomes $S_{r,j} = \psi_j \ell_i$ time units, where ψ_j and ℓ_i are independent. Here, ψ_j shows the total number of transmissions for the j-th packet and is geometrically distributed with success probability $1-\epsilon_0$, mean $\pi_1 = \frac{1}{1-\epsilon_0}$, and second moment $\pi_2 = \frac{1+\epsilon_0}{(1-\epsilon_0)^2}$.

[1] Our analysis can be extended to the case of different erasure probabilities.

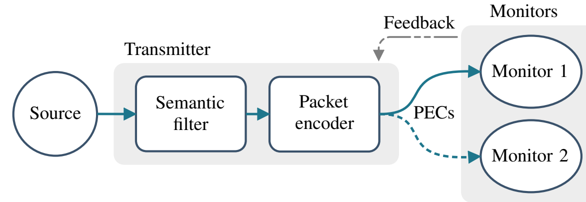

Fig. 1. Semantics-aware transmission over a multiuser network.

III. Problem Formulation

In this section, we formulate the problem of optimal semantic source coding for randomly arriving status updates.

The importance of received packets at the monitor(s) is measured using a special case of the semantics of information (SoI) [7], [17]. Specifically, we consider timeliness as a contextual attribute under the form of a non-increasing utility function $f : \mathbb{R}_0^+ \to \mathbb{R}$ of information freshness, i.e., $\mathcal{S}(t) = f(\Delta(t))$. $\Delta(t) = t - u(t)$ is the instantaneous AoI at the receiver, defined as the difference between the current time instant and the timestamp $u(t)$ of the most recently arrived update. Thereafter, $\Delta_r(t)$ with indices $r=1$ and $r=2$ is the AoI of monitors 1 and 2, respectively. The SoI at each monitor decreases according to $f(\cdot)$ until a valuable arrival for that monitor is observed. Then, the SoI rises to the value of the new update at that time. Hence, monitors could potentially attain different SoIs over similar monitoring time spans. The average SoI for an observation interval $(0,T)$, assuming a stationary ergodic process, is given by $\bar{\mathcal{S}}_r = \lim_{T\to\infty} \dfrac{1}{T} \displaystyle\int_0^T f(\Delta_r(t))\mathrm{d}t$.

A. Semantics-Aware Source Coding

Our objective is to optimally assign codeword lengths ℓ_i, $\forall i \in \mathcal{I}_k$, for semantics-aware encoding as a means to maximize the weighted sum of the monitors' average SoI, i.e., $\bar{\mathcal{S}}_1$ and $\bar{\mathcal{S}}_2$. For convenience of analytical derivation and to ensure positiveness, maximizing the average SoI can be turned into minimizing the average penalty of lateness defined as

$$L_r(\Delta_r) = \lim_{T\to\infty} \frac{1}{T} \int_0^T g(\Delta_r(t))\mathrm{d}t \tag{1}$$

where $g : \mathbb{R}_0^+ \to \mathbb{R}$ is a non-decreasing function [9]. For feasibility, codeword lengths should be positive integers, i.e., $\ell_i \in \mathbb{Z}^+$, and for constructing uniquely decodable codes, the Kraft-McMillan inequality has to be satisfied. Thus, the optimization problem is formulated as

$$\begin{aligned} \mathcal{P}_1 : \min_{\{\ell_i\}\in\mathbb{Z}^+} \quad & w_1 L_1(\Delta_1) + w_2 L_2(\Delta_2) \\ \text{s.t.} \quad & \sum_{i\in\mathcal{I}_k} 2^{-\ell_i} \leq 1 \end{aligned} \tag{2}$$

where $w_1, w_2 \geq 0$ are weight parameters. To solve $\mathcal{P}_1$, we relax the integer constraint for ℓ_i to allow for non-negative real-

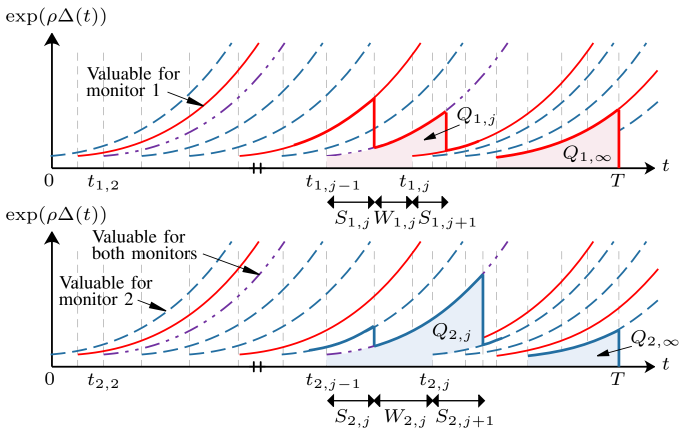

Fig. 2. Sample evolution for the EDT case over time for $\rho = 0.2$.

valued codeword lengths.[2] To explicitly find $L_r(\Delta_r)$, hence solve (2), we need to define the penalty function. In this paper, we propose three different forms of $g(.)$ as follows.

$$g(\Delta_r(t)) = \begin{cases} \exp(\rho\Delta_r(t)) & \text{EDT case} \\ \ln(\rho\Delta_r(t)) & \text{LDT case} \\ \rho(\Delta_r(t))^{\kappa} & \text{PDT case} \end{cases} \tag{3}$$

where $\rho \geq 0$ and $\kappa \in \mathbb{Z}^+$ are constant coefficients. The above cases correspond to *exponentially*, *logarithmically*, and *polynomially* decreasing timeliness (PDT), respectively.

The average penalty of lateness is computed for all cases using (1) and (3). We divide the positive area below the curve of $g(\Delta_r(t))$ over the interval $(0, T)$ into polygons of $Q_{r,j}$, $j = 1, 2, ..., \mathcal{N}_r(T)$, and $Q_{r,\infty}$, as in Fig. 2 for the EDT case. Herein, $\mathcal{N}_r(T) \leq \mathcal{N}(T)$ is the number of admitted packets for monitor 1 ($r = 1$) or monitor 2 ($r = 2$) by time T, where $\mathcal{N}(T)$ is the number of all admitted packets. Thus, we have

$$L_r(\Delta_r) = \lim_{T\to\infty} \frac{1}{T}\left\{ \sum_{j=1}^{\mathcal{N}_r(T)} Q_{r,j} + Q_{r,\infty} \right\} = \eta_r \mathbb{E}[Q_r] \tag{4}$$

where $\eta_r = \lim\limits_{T\to\infty} \frac{\mathcal{N}_r(T)-1}{T}$ is the steady-state time average arrival rate. From (2)–(4), after merging η_1 with w_1 and η_2 with w_2 as all being positive constants, we have

$$\begin{aligned} \mathcal{P}_2 \colon & \min_{\{\ell_i\}\in\mathbb{R}^+} \quad w_1\mathbb{E}[Q_1] + w_2\mathbb{E}[Q_2] \coloneqq \mathcal{J}_{\text{SoI}} \\ & \text{s.t.} \sum_{i\in\mathcal{I}_k} 2^{-\ell_i} \leq 1. \end{aligned} \tag{5}$$

IV. Semantic Multiuser Codeword Design

In what follows, we find the optimal semantics-aware real codeword lengths by solving $\mathcal{P}_2$ for the particular cases in (3).

[2] After computing real-valued ℓ_i, $\forall i \in \mathcal{I}_k$, its corresponding integer value can be obtained by using the rounded-off operation, as $\lceil \ell_i \rceil$.

A. EDT Case

Using second-order Taylor expansion for the exponential function, the area $Q_{r,j}$ for $j \geq 2$ yields

$$\begin{aligned} Q_{r,j} &= \int_{t_{r,j-1}}^{t_{r,j}+S_{r,j+1}} e^{\rho(t-t_{r,j-1})}\mathrm{d}t - \int_{t_{r,j}}^{t_{r,j}+S_{r,j+1}} e^{\rho(t-t_{r,j})}\mathrm{d}t \\ &\approx \frac{\rho}{2}Y_{r,j}^2 + \rho S_{r,j+1}Y_{r,j} + Y_{r,j}. \end{aligned} \tag{6}$$

Therefore, $\mathbb{E}[Q_r]$ is derived as

$$\begin{aligned} \mathbb{E}[Q_r] &\approx \frac{\rho}{2}\mathbb{E}[Y_r^2] + \rho\mathbb{E}[S_r]\mathbb{E}[Y_r] + \mathbb{E}[Y_r] \\ &= \frac{\rho}{2}\pi_2\mathbb{E}[L^2]_r + \rho\pi_1^2(\mathbb{E}[L]_r)^2 \\ &\quad + (1+2\rho\gamma_r)\pi_1\mathbb{E}[L]_r + \rho\gamma_r^2 + \gamma_r, \end{aligned} \tag{7}$$

using $\mathbb{E}[Y_r] = \pi_1\mathbb{E}[L]_r + \gamma_r$ and $\mathbb{E}[Y_r^2] = \pi_2\mathbb{E}[L^2]_r + 2\gamma_r\pi_1\mathbb{E}[L]_r + 2\gamma_r^2$ with $\gamma_r = (\lambda q_{k_r})^{-1}$ [15]. We also have $\mathbb{E}[S_r] = \pi_1\mathbb{E}[L]_r$ and $\mathbb{E}[S_r^2] = \pi_2\mathbb{E}[L^2]_r$ where $\mathbb{E}[L]_r = \sum_{i\in\mathcal{I}_{k_r}} p_i\ell_i$ and $\mathbb{E}[L^2]_r = \sum_{i\in\mathcal{I}_{k_r}} p_i\ell_i^2$ indicate the first and second moments of the codeword lengths, respectively.

Applying (7), the objective function for the optimization problem $\mathcal{P}_2$ becomes

$$\begin{aligned} \mathcal{J}_{\text{SoI}} = & \frac{\rho}{2}\pi_2\Big(w_1\mathbb{E}[L^2]_1 + w_2\mathbb{E}[L^2]_2\Big) \\ & + \rho\pi_1^2\Big(w_1(\mathbb{E}[L]_1)^2 + w_2(\mathbb{E}[L]_2)^2\Big) \\ & + w_1(1+2\rho\gamma_1)\pi_1\mathbb{E}[L]_1 + w_1\rho\gamma_1^2 + w_1\gamma_1 \\ & + w_2(1+2\rho\gamma_2)\pi_1\mathbb{E}[L]_2 + w_2\rho\gamma_2^2 + w_2\gamma_2. \end{aligned} \tag{8}$$

Proposition 1. *The codeword length ℓ_i, $\forall i \in \mathcal{I}_k$, that minimizes (8) in $\mathcal{P}_2$ for the EDT case is given by*

$$\ell_i = -\log_2\left(\frac{\alpha p_i}{\mu(\ln(2))^2} W_0\left(\frac{\mu(\ln(2))^2}{\alpha p_i} 2^{\frac{\beta}{\alpha}}\right)\right) \tag{9}$$

where $\mu \geq 0$ is the Lagrange multiplier, $\alpha = \rho\pi_2(\varpi_1 + \varpi_2)$,

$$\beta = \frac{2\mu\rho\ln(2)\pi_1^2(\varpi_1\chi_1 + \varpi_2\chi_2) + \alpha\theta\pi_1}{\alpha + 2\rho\pi_1^2(\varpi_1\chi_1 + \varpi_2\chi_2)}, \tag{10}$$

$\theta = \varpi_1(1+2\rho\gamma_1) + \varpi_2(1+2\rho\gamma_2)$, and $W_0(.)$ is the principal branch of Lambert W function. Moreover, $\varpi_1 \coloneqq w_1\varrho_1$ and $\varpi_2 \coloneqq w_2\varrho_2$, with $\varrho_1, \varrho_2 \in \{0, 1\}$ being indicator parameters, initialized as follows: $(\varrho_1 = 1, \varrho_2 = 0)$, $(\varrho_1 = \varrho_2 = 1)$, and $(\varrho_1 = 0, \varrho_2 = 1)$ if realization x_i belongs to set $\mathcal{A} = \mathcal{I}_{k_1} - \mathcal{B}$, $\mathcal{B} = \mathcal{I}_{k_1} \cap \mathcal{I}_{k_2}$, and $\mathcal{C} = \mathcal{I}_{k_2} - \mathcal{B}$, respectively.

Before proceeding with the proof, we remark that the values of μ, χ_1, and χ_2 are calculated using **Algorithm 1** through its inner and outer loops. First, we assume uniformly distributed p_i, $\forall i \in \mathcal{I}_k$, and assign $\chi_1^{(0)} = k_1/k$ and $\chi_2^{(0)} = k_2/k$. Then, we find ℓ_i and compute new χ_1 and χ_2 for a given μ. Based on them, the new values for ℓ_i are found. This process continues until the convergence criterion ε is satisfied. Then, the outer loop checks the Kraft-McMillan condition and resets μ if the condition is not satisfied. At the end, the algorithm converges to the final values of χ_1, χ_2, β, ℓ_i, and μ with the rate of $\mathcal{O}((N_nN_m)^{-1})$ in which N_n and N_m denote the maximum numbers of inner and outer iterations, respectively. With regard to (9) and (10), the dependency of ℓ_i on the changes of χ_1 and χ_2 reduces for higher order iterations and larger μ. Therefore, after a few iterations, the values of χ_1 and χ_2 remain fixed for the given μ, leading to final codeword lengths.

Proof: By (8), we define the Lagrange function for $\mathcal{P}_2$ as

$$\begin{aligned}\mathcal{L}(\ell_i;\mu) &= \frac{\rho}{2}\pi_2\Big[w_1\sum_{i\in\mathcal{I}_{k_1}} p_i\ell_i^2 + w_2\sum_{i\in\mathcal{I}_{k_2}} p_i\ell_i^2\Big] \\ &+ \rho\pi_1^2\Big[w_1\Big(\sum_{i\in\mathcal{I}_{k_1}} p_i\ell_i\Big)^2 + w_2\Big(\sum_{i\in\mathcal{I}_{k_2}} p_i\ell_i\Big)^2\Big] \\ &+ w_1(1+2\rho\gamma_1)\pi_1\sum_{i\in\mathcal{I}_{k_1}} p_i\ell_i + w_2(1+2\rho\gamma_2)\pi_1\sum_{i\in\mathcal{I}_{k_2}} p_i\ell_i \\ &+ w_1\rho\gamma_1^2 + w_2\rho\gamma_2^2 + w_1\gamma_1 + w_2\gamma_2 + \mu\Big(\sum_{i\in\mathcal{I}_k} 2^{-\ell_i} - 1\Big)\end{aligned} \tag{11}$$

where $\mu \geq 0$ is the Lagrange multiplier. Then, we write the Karush-Kuhn-Tucker (KKT) conditions for $i \in \mathcal{I}_k$, as follows:

$$\begin{aligned}\frac{\partial\mathcal{L}(\ell_i;\mu)}{\partial\ell_i} &= \rho\pi_2(\varpi_1+\varpi_2)p_i\ell_i + 2\rho\pi_1^2\Big[\varpi_1\Big(\sum_{i\in\mathcal{I}_{k_1}} p_i\ell_i\Big) \\ &+\varpi_2\Big(\sum_{i\in\mathcal{I}_{k_2}} p_i\ell_i\Big)\Big]p_i + \varpi_1(1+2\rho\gamma_1)\pi_1 p_i \\ &+ \varpi_2(1+2\rho\gamma_2)\pi_1 p_i - \mu\ln(2)2^{-\ell_i} = 0\end{aligned} \tag{12}$$

where $\varpi_1 := w_1\varrho_1$ and $\varpi_2 := w_2\varrho_2$ with $\varrho_1, \varrho_2 \in \{0,1\}$ being indicator parameters. Let us split set $\mathcal{I}_k$ into three disjoint sets of $\mathcal{A} \triangleq \mathcal{I}_{k_1} - \mathcal{B}$, $\mathcal{B} \triangleq \mathcal{I}_{k_1} \cap \mathcal{I}_{k_2}$, and $\mathcal{C} \triangleq \mathcal{I}_{k_2} - \mathcal{B}$. Thus, we have $(\varpi_1 = w_1, \varpi_2 = 0)$, $(\varpi_1 = w_1, \varpi_2 = w_2)$, or $(\varpi_1 = 0, \varpi_2 = w_2)$ if x_i belongs to $\mathcal{A}$, $\mathcal{B}$, or $\mathcal{C}$, respectively.

The complementary slackness condition is

$$\mu\Big(\sum_{i\in\mathcal{I}_k} 2^{-\ell_i} - 1\Big) = 0. \tag{13}$$

There exist two conditions, one of which meets (13): (i) $\mu = 0$, $\sum_{i\in\mathcal{I}_k} 2^{-\ell_i} < 1$; or (ii) $\mu \neq 0$, $\sum_{i\in\mathcal{I}_k} 2^{-\ell_i} = 1$. Under (i), the right hand side of (12), in which all terms are non-negative, equals to zero, resulting in $\ell_i = 0$, $\forall i \in \mathcal{I}_k$. Since negative codeword lengths are not meaningful, by contradiction, condition (ii) must satisfy (13). Summing (12) over all codeword indices, it is hard to find a closed-form expression based on $\mathbb{E}[L]$, $\mathbb{E}[L]_1$ and $\mathbb{E}[L]_2$. To this end, we introduce two parameters $\chi_1 = \frac{\mathbb{E}[L]_1}{\mathbb{E}[L]}$ and $\chi_2 = \frac{\mathbb{E}[L]_2}{\mathbb{E}[L]}$, with $0 \leq \chi_1 \leq 1$ and $0 \leq \chi_2 \leq 1$, which capture the *nonlinear* relationship with k_1 and k_2, respectively. For known χ_1 and χ_2, we can find $\mathbb{E}[L]$, $\mathbb{E}[L]_1$, and $\mathbb{E}[L]_2$. After some calculations, we reach

$$\frac{\mu(\ln(2))^2}{\alpha p_i}2^{-\ell_i}\exp\Big(\frac{\mu(\ln(2))^2}{\alpha p_i}2^{-\ell_i}\Big) = \frac{\mu(\ln(2))^2}{\alpha p_i}2^{\frac{\beta}{\alpha}} \tag{14}$$

where $\alpha = \rho\pi_2(\varpi_1+\varpi_2)$, and β is given in (10). The form of (14) is equal to $x\exp(x) = y$ for which the solution is $x = W_m(y)$, where $m = 0$ for $y \geq 0$. ■

Algorithm 1: Solution for deriving χ_1 and χ_2

Input: Fixed parameters $\mathcal{I}_k$, $\mathcal{I}_{k_1}$, $\mathcal{I}_{k_2}$, and p_i, $\forall i \in \mathcal{I}_k$. Stopping accuracy ε. Initial parameters $\mu^{(0)}$, $\chi_1^{(0)}$, $\chi_2^{(0)}$, $\beta^{(0)}$, and $\ell_i^{(0)}$, $\forall i$.

Output: Final form parameters $\chi_1 = \chi_1^{(n)}$, $\chi_2 = \chi_2^{(n)}$, $\beta = \beta^{(n)}$, $\ell_i = \ell_i^{(n)}$, $\forall i$, and $\mu = \mu^{(m)}$.

1 *Iteration* m:
2 *Iteration* n:
3 Update $\beta^{(n)}$ and $\ell_i^{(n)}$ using (10) and (9), respectively.
4 Compute $\mathbb{E}[L] = \sum_{i\in\mathcal{I}_k} p_i\ell_i^{(n)}$, $\mathbb{E}[L]_1 = \sum_{i\in\mathcal{I}_{k_1}} p_i\ell_i^{(n)}$, and $\mathbb{E}[L]_2 = \sum_{i\in\mathcal{I}_{k_2}} p_i\ell_i^{(n)}$.
5 Update $\chi_1^{(n)}$ and $\chi_2^{(n)}$ based on **4**.
6 **if** *Criterion* $\left|\chi_1^{(n)} - \chi_1^{(n-1)}\right| > \varepsilon$ *or* $\left|\chi_2^{(n)} - \chi_2^{(n-1)}\right| > \varepsilon$ **then** set $n = n+1$, and **goto 2**.
7 Compute $\beta^{(n)}$ from (10), and derive $\ell_i^{(n)}$ from (9).
8 **if** $\sum_{i\in\mathcal{I}_k} 2^{-\ell_i^{(n)}} = 1$ **then** stop the process, and **goto 11**.
9 **else if** $\sum_{i\in\mathcal{I}_k} 2^{-\ell_i^{(n)}} < 1$ **then** decrease $\mu^{(m)}$, set $m = m+1$, and **goto 1**.
10 **else** increase $\mu^{(m)}$, set $m = m+1$, and **goto 1**.
11 Save $\chi_1^{(n)}$, $\chi_2^{(n)}$, $\beta^{(n)}$, $\ell_i^{(n)}$, $\forall i$, and $\mu^{(m)}$.

B. LDT Case

In this case, the area $Q_{r,j}$ for $j \geq 2$ yields

$$\begin{aligned}Q_{r,j} &= \int_{t_{r,j-1}}^{t_{r,j}+S_{r,j+1}} \ln(\rho(t - t_{r,j-1}))\mathrm{d}t \\ &- \int_{t_{r,j}}^{t_{r,j}+S_{r,j+1}} \ln(\rho(t - t_{r,j}))\mathrm{d}t \\ &\approx \rho Y_{r,j}^2 + 2\rho S_{r,j+1}Y_{r,j} - 2Y_{r,j},\end{aligned} \tag{15}$$

which results in

$$\begin{aligned}\mathbb{E}[Q_r] &\approx \rho\pi_2\mathbb{E}[L^2]_r + 2\rho\pi_1^2(\mathbb{E}[L]_r)^2 \\ &+ 2(2\rho\gamma_r - 1)\pi_1\mathbb{E}[L]_r + 2\rho\gamma_r^2 - 2\gamma_r.\end{aligned} \tag{16}$$

Consequently, we obtain the following objective function

$$\begin{aligned}\mathcal{J}_{\mathrm{SoI}} &= \rho\pi_2\Big(w_1\mathbb{E}[L^2]_1 + w_2\mathbb{E}[L^2]_2\Big) \\ &+ 2\rho\pi_1^2\Big(w_1(\mathbb{E}[L]_1)^2 + w_2(\mathbb{E}[L]_2)^2\Big) \\ &+ 2w_1(2\rho\gamma_1 - 1)\pi_1\mathbb{E}[L]_1 + 2w_1\rho\gamma_1^2 - 2w_1\gamma_1 \\ &+ 2w_2(2\rho\gamma_2 - 1)\pi_1\mathbb{E}[L]_2 + 2w_2\rho\gamma_2^2 - 2w_2\gamma_2.\end{aligned} \tag{17}$$

Putting (17) into $\mathcal{P}_2$ and following the same procedure as (11)–(14), we obtain the optimal codeword length ℓ_i, $\forall i \in \mathcal{I}_k$, as

$$\ell_i = -\log_2\Big(\frac{\alpha' p_i}{\mu'(\ln(2))^2}W_0\Big(\frac{\mu'(\ln(2))^2}{\alpha' p_i}2^{\frac{\beta'}{\alpha'}}\Big)\Big) \tag{18}$$

where $\mu' \geq 0$, $\alpha' = 2\rho\pi_2(\varpi_1 + \varpi_2)$,

$$\beta' = \frac{4\mu'\rho\ln(2)\pi_1^2(\varpi_1\chi_1 + \varpi_2\chi_2) + 2\alpha'\theta'\pi_1}{\alpha' + 4\rho\pi_1^2(\varpi_1\chi_1 + \varpi_2\chi_2)}, \tag{19}$$

and $\theta' = \varpi_1(2\rho\gamma_1 - 1) + \varpi_2(2\rho\gamma_2 - 1)$. For $i \in \mathcal{A}$ and $i \in \mathcal{C}$, we set $\varrho_2 = 0$ and $\varrho_1 = 0$, respectively. Otherwise, $\varrho_1 = \varrho_2 = 1$. The values of μ', χ_1, and χ_2 are found via **Algorithm 1** and substituting the parameters of ℓ_i, such as β with β'.

C. PDT Case

Setting $\kappa = 1$ for exposition convenience, we obtain

$$Q_{r,j} \approx \frac{\rho}{2} Y_{r,j}^2 + \rho S_{r,j+1} Y_{r,j}, \tag{20}$$

for which the expected value is derived as

$$\begin{aligned}\mathbb{E}[Q_r] \approx \frac{\rho}{2}\pi_2 \mathbb{E}[L^2]_r + \rho\pi_1^2(\mathbb{E}[L]_r)^2 \\ + 2\rho\gamma_r\pi_1\mathbb{E}[L]_r + \rho\gamma_r^2.\end{aligned} \tag{21}$$

Similar to IV-A and IV-B, we get $\mathcal{J}_{\text{SoI}}$ and solve $\mathcal{P}_2$. The solution for optimal ℓ_i, $\forall i \in \mathcal{I}_k$, is

$$\ell_i = -\log_2\left(\frac{\alpha p_i}{\mu''(\ln(2))^2} W_0\left(\frac{\mu''(\ln(2))^2}{\alpha p_i} 2^{\frac{\beta''}{\alpha}}\right)\right) \tag{22}$$

where $\mu'' \geq 0$, and if $\theta'' = 2\varpi_1\rho\gamma_1 + 2\varpi_2\rho\gamma_2$, we have

$$\beta'' = \frac{2\mu''\rho\ln(2)\pi_1^2(\varpi_1\chi_1 + \varpi_2\chi_2) + \alpha\theta''\pi_1}{\alpha + 2\rho\pi_1^2(\varpi_1\chi_1 + \varpi_2\chi_2)}. \tag{23}$$

We set $\varrho_2 = 0$ and $\varrho_1 = 0$ for $i \in \mathcal{A}$ and $i \in \mathcal{C}$, respectively, or otherwise $\varrho_1 = \varrho_2 = 1$. The values of μ'', χ_1, and χ_2 are obtained from **Algorithm 1** by replacing the parameters of ℓ_i.

D. Asymptotic Expansions

Using $\left[\frac{W_0(y)}{y}\right]^\xi = \exp(-\xi W_0(y))$, for $y \geq e$, and $W_0(y) = \ln(y) - \ln(\ln(y)) + \mathcal{O}(1)$ for large y, we can write $\ell_i \propto c/p_i - \ln(c/p_i)$, for $p_i \leq \frac{c}{e} \leq 1$, with $c := \frac{\mu(\ln(2))^2}{\alpha} 2^{\frac{\beta}{\alpha}}$. This expression, which is monotonically decreasing, corroborates that the higher the probability of occurrence of a realization, the shorter the assigned codeword length, and vice versa. $p_i \to 0$ yields $\ell_i \to \ell_{\max}$, where $\ell_{\max}$ is an upper bound to which the size of a codeword length converges.

Furthermore, consider a uniform pmf for the realizations, i.e., $p_i = 1/n$, $\forall i$. In that case, we have codewords of equal size, i.e., $\ell_1 = \ell_2 = ... = \ell_n$, with $\ell_i \propto n - \ln(n)$. Moreover, $n \gg 1$ results in $\ell_i = \ell_{\max}$, $\forall i \in \mathcal{I}_k$, and $\ell_{\max}$ remains almost fixed for the large enough n. As the number of realizations increases, the assigned codeword lengths become larger at the expense of longer service time.

V. Numerical Results

In this section, we provide numerical results for the optimal number of selected status updates (packets) in different scenarios, highlighting the advantages of semantic filtering and source coding. We use Zipf(n, s) distribution with pmf $P_X(x_i) = \frac{1/i^s}{\sum_{j=1}^n 1/j^s}$, $n = |\mathcal{X}| = 100$, and $s = 0.4$. The parameter s allows us to vary from a uniform distribution ($s = 0$) to a "peaky distribution". The packet error rate ϵ_0 is initialized according to its upper bound $\Phi\left(\frac{\delta}{\sqrt{\delta(1-\delta)}}\right)$ [18], where δ and $\Phi(.)$ denote the erasure probability and the cumulative distribution function of the standard normal Gaussian distribution, respectively. Unless otherwise stated, we set $\rho = 0.2$, $\delta = 0.5$, $w_1 = w_2 = 1$, and $T = 10$ [sec].

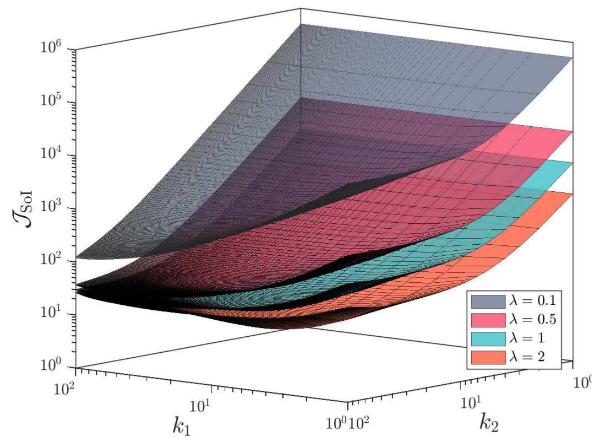

Fig. 3. The interplay between $\mathcal{J}_{\text{SoI}}$ and the number of selected realizations k_1 and k_2 for the EDT case with $n = 100$ and $w_1 = w_2 = 1$.

TABLE I
Optimal number of selected packets for $n = 100$.

	EDT case		LDT case		PDT case	
Arrival rate	k_1	k_2	k_1	k_2	k_1	k_2
$\lambda = 0.1$	100	94	100	91	100	89
$\lambda = 0.5$	100	45	100	42	100	41
$\lambda = 1$	45	14	46	13	46	12
$\lambda = 2$	45	10	46	8	46	8

Fig. 3 depicts the value of $\mathcal{J}_{\text{SoI}}$ versus the number of selected (important) realizations for the EDT case. We observe that increasing the arrival rate reduces $\mathcal{J}_{\text{SoI}}$, as well as the optimal k_1 and k_2 values. In addition, the transmitter filters a considerably higher number of frequent arrivals than infrequent ones. However, no filtering ($k_1, k_2 \to n$) results in performance degradation due to spending time for sending insignificant realizations. Likewise, the derived optimal values of k_1 and k_2 for the all cases with different arrivals rates are listed in Table I. We see that the exponential penalty results in the lowest and highest values for optimal k_1 and k_2, respectively.

Fig. 4 shows the effects of weight parameters in the objective function $\mathcal{J}_{\text{SoI}}$ and the number of selected realizations for the EDT case, with $\lambda = 1$, and $\delta = 0.5$. Varying each weight parameter alters the optimal values of k_1 and k_2, hence $\mathcal{J}_{\text{SoI}}$. Notably, giving ten times more weight to the arrivals of monitor 2 compared to those of monitor 1 equalizes the optimal k_1 and k_2. In this case, the transmitter equally filters around 33% of frequent and infrequent arrivals. The obtained information from Fig. 4 and from its extension for $\delta = 0.25$ is given in Table II. We observe that higher erasure probability results in fewer selected packets since the transmitter spends more time to retransmit the unsuccessful packets. The same results hold for the LDT and PDT cases.

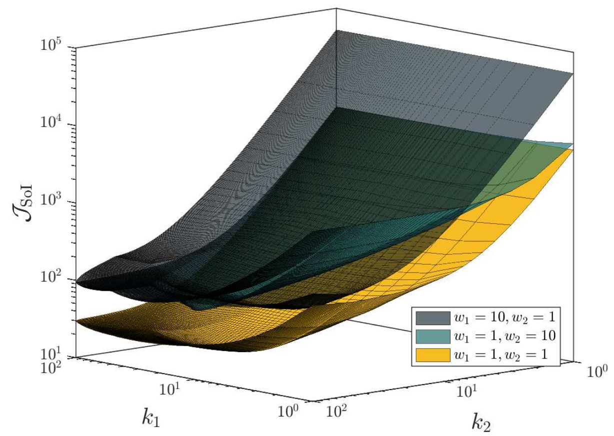

Fig. 4. The interplay between $\mathcal{J}_{\text{SoI}}$ and the number of selected realizations k_1 and k_2 for the EDT case with $n = 100$, $\lambda = 1$, and $\delta = 0.5$.

TABLE II
OPTIMAL NUMBER OF SELECTED PACKETS FOR $n = 100$ AND $\lambda = 1$.

	Weight parameters (w_1, w_2)					
	$(1,1)$		$(10,1)$		$(1,10)$	
Erasure probability	k_1	k_2	k_1	k_2	k_1	k_2
$\delta = 0.25$	47	17	100	60	33	34
$\delta = 0.5$	45	14	50	5	33	33

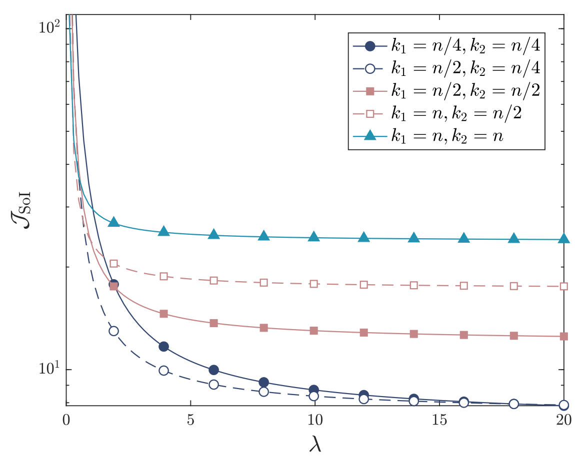

Fig. 5. The objective function $\mathcal{J}_{\text{SoI}}$ versus arrival rate λ for the EDT case with $n = 100$ and $w_1 = w_2 = 1$.

Fig. 5 presents the objective function $\mathcal{J}_{\text{SoI}}$ versus the arrival rate λ for different numbers of selected packets and the EDT case. Increasing the arrival rate decreases $\mathcal{J}_{\text{SoI}}$, which in turn diminishes and saturates at higher rates. Furthermore, increasing both or one of k_1 or k_2 means that the arrival rate required to decrease the objective function diminishes. For instance, the minimum values of arrival rates (hence, $\mathcal{J}_{\text{SoI}}$) for $k_1 = k_2 = n/4$, $k_1 = k_2 = n/2$, and $k_1 = k_2 = n$ are around $11(14)$, $17(11)$, $20(2)$, respectively. Thus, for large k_1 and k_2, the objective function takes high values for any arrival rates. According to the analytical expressions derived for the EDT case, we find the global optimal values of $\lambda^* \simeq 10.27, 17.63,$ and 19.12 sequentially for $k_1 = k_2 = n/4$, $k_1 = k_2 = n/2$, and $k_1 = k_2 = n$. Likewise, Fig. 5 can be plotted for the LDT and PDT cases, resulting in similar interpretations.

VI. Conclusion

We studied the timely source coding problem in a two-user status update system, where observations of an information source are filtered and sent to two monitors depending on their importance for achieving each user's goal. Optimizing the codeword length according to semantics-aware utility functions, the amount of status updates communicated is significantly reduced. Our analytical and numerical results show an optimal value of realizations to send, which depends on the source distribution, the arrival rate, and the weight of each monitor to maximize the value of transferred information in the network.

References

[1] C. E. Shannon and W. Weaver, *The mathematical theory of communication*. University of Illinois Press, Urbana, 1949.

[2] Y. Bar-Hillel and R. Carnap, "Semantic information," *British Journal for the Philosophy of Science*, vol. 4, no. 14, 1953.

[3] F. I. Dretske, *Knowledge and the flow of information*, 1st ed. MIT Press Cambridge, Mass., 1981.

[4] B. Juba and M. Sudan, "Universal semantic communication I," in *Proc. of the 40th Annual ACM Symposium on Theory of Computing*, ser. STOC '08, New York, NY, USA, 2008, p. 123–132.

[5] B. Güler, A. Yener, and A. Swami, "The semantic communication game," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 787–802, 2018.

[6] 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, pp. 435–443, 2020.

[7] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 96–102, 2021.

[8] 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.

[9] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE Journal on Sel. Areas in Commun.*, vol. 39, no. 5, pp. 1183–1210, 2021.

[10] R. L. Stratonovich, "On the value of information," *Izv. USSR Acad. Sci. Tech. Cybern.*, no. 5, 1965.

[11] R. A. Howard, "Information value theory," *IEEE Trans. on Systems Science and Cybernetics*, vol. 2, no. 1, 1966.

[12] P. Mayekar, P. Parag, and H. Tyagi, "Optimal source codes for timely updates," *IEEE Trans. on Inform. Theory*, vol. 66, no. 6, pp. 3714–3731, 2020.

[13] J. Zhong and R. D. Yates, "Timeliness in lossless block coding," in *Data Compression Conference*, 2016, pp. 339–348.

[14] J. Zhong, R. D. Yates, and E. Soljanin, "Timely lossless source coding for randomly arriving symbols," in *IEEE Inform. Theory Workshop*, 2018.

[15] M. Bastopcu, B. Buyukates, and S. Ulukus, "Optimal selective encoding for timely updates," in *Annual Conference on Information Sciences and Systems*, 2020, pp. 1–6.

[16] P. Agheli, N. Pappas, and M. Kountouris, "Semantics-aware source coding in status update systems," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, 2022.

[17] N. Pappas and M. Kountouris, "Goal-oriented communication for real-time tracking in autonomous systems," in *IEEE International Conference on Autonomous Systems (ICAS)*, 2021, pp. 1–5.

[18] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.