Semantic Filtering and Source Coding in Distributed Wireless Monitoring Systems

Pouya Agheli, *Student Member, IEEE*, Nikolaos Pappas, *Senior Member, IEEE*, and Marios Kountouris, *Fellow, IEEE*.

Abstract

The problem of goal-oriented semantic filtering and timely source coding in multiuser communication systems is considered here. We study a distributed monitoring system in which multiple information sources, each observing a physical process, provide status update packets to multiple monitors having heterogeneous goals. Two semantic filtering schemes are first proposed as a means to admit or drop arrival packets based on their goal-dependent importance, which is a function of the intrinsic and extrinsic attributes of information and the probability of occurrence of each realization. Admitted packets at each sensor are then encoded and transmitted over block fading wireless channels so that served monitors can timely fulfill their goals. A truncated error control scheme is derived, which allows transmitters to drop or retransmit undelivered packets based on their significance. Then, we formulate the timely source encoding optimization problem and analytically derive the optimal codeword lengths assigned to the admitted packets which maximize a weighted sum of semantic utility functions for all pairs of communicating sensors and 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 the proposed semantic filtering, source coding, and error control schemes.

Index Terms

Goal-oriented semantic communication, semantic filtering, timely source coding, distributed monitoring systems.

P. Agheli and M. Kountouris are with the Communication Systems Department, EURECOM, Sophia-Antipolis, France, email: {`pouya.agheli, marios.kountouris`}`@eurecom.fr`. N. Pappas is with the Department of Computer and Information Science, Linköping University, Sweden, email: `nikolaos.pappas@liu.se`. Part of this work is presented in [1], [2].

I. INTRODUCTION

GOAL-ORIENTED semantic communication has recently attracted considerable attention and constitutes an information handling paradigm that has the potential to render various network processes more efficient and effective through a parsimonious usage of communication and computation resources. The design and the evolution of communication systems to date have been mainly driven by a maximalistic approach, which sets audacious yet often hard to achieve goals and comes with inflated requirements in terms of resources, network over-provisioning, and ineffective scalability. In sharp contrast, goal-oriented semantics communication could be seen as a minimalist design approach ("less is more"), advocating a paradigm shift from extreme performance to sustainable performance, where the effective performance is maximized while significantly improving network resource usage, energy consumption, and computational efficiency. This vision has a long history dating back to Weaver's introduction of Shannon model of communication [3]. Various attempts, from different angles and using diverse tools, have been made in the past towards a semantic theory of communication [4]–[11]; the vast majority of these endeavors remained at a conceptual level and did not lead to an elegant and/or insightful theory with immediate practical applications. Nonetheless, the quest for such theory has recently gained new impetus [12]–[14], fueled by the emergence of connected intelligence systems, real-time cyber-physical systems, and interactive, autonomous multi-agent systems.

An indispensable element to unlock the potential of goal-oriented communication is a concise, operational, and universal definition of the *semantics of information* (SoI), i.e., a set of measures for the significance and the usefulness of messages with respect to the goal of data exchange. Going beyond surrogate metrics, such as age of information (AoI) [15]–[17], quality of information (QoI) [18], and value of information (VoI) [19]–[21], SoI can be leveraged so that the communication process, together with key associated functionalities (e.g., sensing, learning, processing), are adapted to the end-user goals/requirements. Empowered by networked intelligence, in semantic communication, only valuable and relevant content with respect to a goal is acquired, transported and reconstructed, leading to a drastic reduction in the number of unnecessary bits processed and sent. Otherwise stated, following the mantra that not all bits are equal, semantic communication may boost the "information efficiency" of future communication systems, meaning that it could maximize the number of bits of useful information extracted and delivered per resource consumed. This new communication paradigm has the potential to redefine

importance, timing, and effectiveness in future networked intelligent systems.

In this paper, we study a distributed monitoring system (DMS), in which multiple remote monitors receive status updates from multiple smart devices (e.g., sensors), each observing an information source. The updates generated by an information source may correspond to observations or measurements of a random physical phenomenon (event) and are taken from a known discrete distribution with finite support. Each status update is assigned a value reflecting its importance based on its intrinsic features, such as probability of occurrence, and on its extrinsic attributes, related to the sensor (source) at which it is generated. Semantic filtering is first performed at the transmitter, as a means to admit or drop only the most relevant or useful packets according to the associated monitor's application-dependent goal. Admitted status updates are then encoded and sent to connected monitors over orthogonal block fading channels. Different error control protocols are employed to harness packet transmission failures due to fading. The objective of this paper is to design a semantic source coding scheme for a multiuser system with heterogeneous goals, considering the probability of occurrence of a realization at a sensor side, the probability that a monitor successfully receives an update from its connected sensor, and the rationale for which update packets are sent to the destination. Specifically, we consider that only a fraction of the source realizations is important for the different monitors. A set of realizations becomes more significant or relevant than the others for a certain application, where its elements could potentially vary for different goals or over different time spans. A simple instance of this model is a scenario where one decision maker 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). In that case, only "most" ("least") frequent source realizations are important for the first (second) monitor, treating the remaining ones as not informative or irrelevant. The SoI is captured here via a metric of timeliness for the received updates at the monitor(s), which in turn is a nonlinear function of AoI [21].

A. Related Work

This work falls within the realm of timely source coding problem [1], [2], [22]–[27]. These works study the design of lossless source codes and block codes that minimize the average AoI in status update systems under different queuing theoretic considerations. References [25], [27] consider a selective encoding mechanism at the transmitter for timely updates. The optimal real

codeword lengths that minimize the average age at the receiver are derived therein, whereas in [26] an empty symbol is used to reset (or not) the age. In [22], the authors consider a zero-wait update policy and find optimal source codes which achieve the minimum average age up to a constant gap, using Shannon codes based on a tilted version of the original symbol generating probability mass function (pmf). Semantic source coding is studied in [1], where optimal real codeword lengths that maximize a semantics-aware utility function and minimize a quadratic average length cost [28] are determined. The semantic encoding problem is extended to a two-user system with heterogeneous, possibly conflicting or diverging, goals in [2]. This paper extends our prior work into distributed monitoring systems (DMS) with multiple sensors at the transmitter side and multiple monitors at the destination, which have heterogeneous and dissimilar goals. We propose simple semantic value assessment schemes for status update packets and an adaptive semantic filtering method. We then determine codeword lengths that maximize a weighted sum of semantics-aware utility functions for all pairs of communicating sensors and monitors, highlighting the performance gains of semantic filtering and source coding.

B. Contributions

The main contributions of this work can be summarized as follows.

- We perform semantic value assessment for the status update packets in two levels. At the source level, importance of an arrival takes on the form of a *meta-value* and captures in a flexible manner the interdependence between different intrinsic and extrinsic attributes. At the link level, SoI is measured using a nonlinear function of AoI.

- We propose two semantic filtering mechanisms to control the flow of arrival packets, drop unimportant arrivals, and reserve the channel for informative packets. In the first (fixed-length) scheme, a set of unimportant arrivals is filtered using selective encoding. In the second (adaptive-length) scheme, a smaller fraction of the least important packets (considering the occurrence probability) are immediately dropped upon arrival, while the remaining ones may pass the filter depending on their relative importance for a predefined goal and its evolution. We show that the latter method performs better for any range of arrival rates at the expense of a small increase in the channel load.

- For packet error control, we employ truncated forms of simple and hybrid automatic repeat request (ARQ and HARQ) protocols. In contrast to conventional importance-agnostic protocols with a fixed number of retransmissions, we propose a scheme in which the maximum
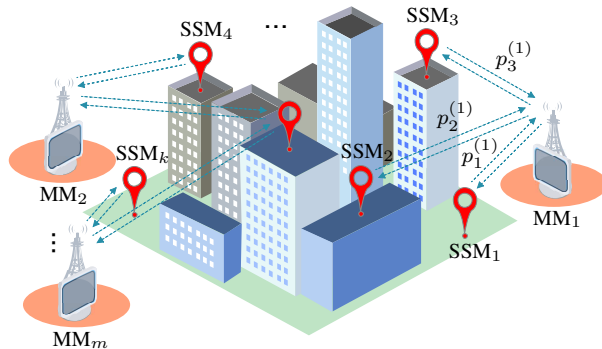
Fig. 1: A goal-oriented, semantics-empowered distributed monitoring system.

number of a packet's retransmissions is adapted to its importance with respect to the specific goal for which it is generated. Evidently, a packet with higher importance is assigned with more retransmission rounds in the event of multiple failures.

- We determine the optimal real codeword length of each admitted arrival packet, based on its probability of occurrence, the observation probability related to the sensor tracking the realization, and the served monitor, as well as its meta-value. For that, we cast an optimization problem that maximizes a weighted sum of semantics-aware utility functions for all pairs of sensors and monitors, and obtain the solution both analytically and using a proposed simple algorithm.

Notations: $\mathbb{R}$, $\mathbb{R}^+$, $\mathbb{R}_0^+$, and $\mathbb{Z}^+$ denote the set of real, positive real, non-negative real, and positive integer numbers, respectively. $\mathbb{E}[\cdot]$ is the expectation operator, $\mathcal{W}_0(.)$ denotes the principal branch of Lambert $\mathcal{W}$ function, and $\mathcal{O}(\cdot)$ denotes the growth rate of a function.

II. SYSTEM MODEL

We consider a DMS, in which a set $\mathcal{K}$ of smart sensing modules (SSMs), with cardinality $|\mathcal{K}| = K$, provides timely status updates to a set $\mathcal{M}$ of M monitor modules (MMs) (see Figure 1). An SSM consists of a sensor, a semantic filtering module, a source encoder, and a module for information transmission (PHY - physical layer) operations, such as modulation and channel coding. An MM could be a *display device* or an *actuator*, which serves a specific goal and performs tasks based on status updates and commands received from the SSMs.

We consider a model where the k-th SSM, $k = 1, 2, \ldots, K$, is connected to a subset of MMs, denoted by $\mathcal{M}(k)$, based on a fixed topology. Therefore, the m-th MM, where $m = 1, 2, \ldots, M$, receives timely status updates only from its serving SSMs, denoted as $\mathcal{K}(m) = \{k : m \in \mathcal{M}(k)\}$.

An SSM tracks a physical phenomenon (event) with finite realizations and generates independent and identically distributed (i.i.d.) status update packets. The realization set for the k-th SSM is defined as $\mathcal{X}^{(k)} = \{x_i^{(k)} \mid i \in \mathcal{I}_{n_k}\}$, $\mathcal{I}_{n_k} = \{1, 2, ..., n_k\}$, where each element has a probability of occurrence $\bar{p}_i^{(k)} = P_X(x_i^{(k)})$ with $P_X(\cdot)$ being a known pmf. Depending on the task(s) at the MM side or the goal specified by the application or the end-user, a realization $x_i^{(k)}$ at the k-th SSM connected to the m-th MM is attributed a certain importance (value) $v_i^{(k,m)}$ at its source level. The feature-based value assessment of each packet at an SSM is discussed in Section III. In this work, we consider a continuous-time system where packet generation follows a Poisson process with input rate λ_k for the k-th SSM. Thus, the input arrival rate at one layer of an MM connected to a set of identical sensors is equal to sum of the input rates from those SSMs.

The probability that the m-th MM observes an update packet from the k-th serving SSM among all the sensors it is connected to, i.e., $\mathcal{K}(m)$, is denoted by $p_k^{(m)}$, where $\sum_{k \in \mathcal{K}(m)} p_k^{(m)} = 1, \forall m$. Since the observation probability of an SSM from each of its connected MMs can be different, an arbitrary realization at that SSM is supposed to occur with different probabilities for those served MMs. In this regard, information processing transmission is performed in multiple layers and different communication channels, respectively. For instance, semantic filtering, source encoding, and PHY operations at the k-th SSM are performed in $|\mathcal{M}(k)|$ parallel layers, each of which is reserved for packets transmitted to one served MM. From the monitoring perspective, the sensors with the similar physical features, probabilities of realizations $p_i^{(k,m)}$, and observation probabilities $p_k^{(m)}$ are considered *identical*[1], while the rest are called *dissimilar* smart sensors. Like the SSMs, multiple layers are utilized at each MM to process packets arriving from its dissimilar sensors, whereas packets from identical sensors are processed in the same layers. A functional diagram for the k-th SSM and the m-th MM, $\forall m \in \mathcal{M}(k)$, is depicted in Figure 2.

A. Packet Transmission

Assuming no buffer is employed at the transmitter of every SSM, a status update packet is blocked when the channel is busy. *Semantics-aware packet filtering* is employed at every SSM for each of the connected MMs as a means to transfer only the most *valuable and important* packets for effectively serving the goal or purpose of the data exchange, as well as having the least possible blockage of valuable arrivals due to heavy packet load. Specifically, at the k-th

---

[1]In practice, sensors with identical characteristics can monitor the same phenomenon or process from different angles, locations, or time instants. In our analysis, a set of identical sensors can be treated as a unique sensor with shared properties.
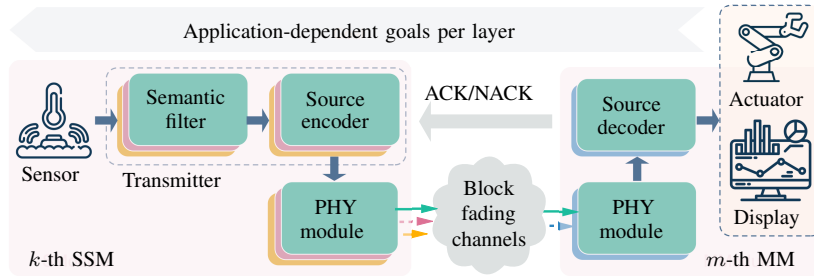
Fig. 2: The functional diagram of a communication link between the k-th SSM and the m-th MM.

SSM, a semantic filter admits the $l_k^{(m)}$ most important realizations via a flow controller for the m-th MM and discards the rest, i.e., $n_k - l_k^{(m)}$. Thus, the index set of the most valuable arrivals at the k-th SSM from the perspective of the m-th MM is denoted by $\mathcal{I}_{l_k^{(m)}}$ where $\mathcal{I}_{l_k^{(m)}} \subseteq \mathcal{I}_{n_k}$. Once semantic filtering is performed, the source encoder at the k-th SSM assigns instantaneous (prefix-free) codewords with lengths $\ell_i^{(k,m)}$ to the status packets $x_i^{(k)}$ admitted for the m-th MM, $\forall m \in \mathcal{M}(k)$, based on the following truncated distribution

$$
p_i^{(k,m)} = \begin{cases} p_k^{(m)} \dfrac{\bar{p}_i^{(k)}}{q_{l_k^{(m)}}}, & \forall i \in \mathcal{I}_{l_k^{(m)}} \\ 0, & \forall i \notin \mathcal{I}_{l_k^{(m)}} \end{cases}
\tag{1}
$$

where $q_{l_k^{(m)}} := \sum_{i \in \mathcal{I}_{l_k^{(m)}}} \bar{p}_i^{(k)}$.

Processed packets are then forwarded to the PHY module, mapped using a binary modulation scheme, and transmitted over a *noisy orthogonal* channel subject to block fading. Thus, channel gains remain constant during each packet transmission period and change in an i.i.d. manner among different periods.

B. Packet Reception

We define $t_j^{(k,m)}$ the time instant that the j-th packet is received at the m-th MM from the serving k-th SSM. The update interval between the j-th successive arrival and the next one at the same layer is then modeled as random variable (r.v.) $Y_j^{(k,m)} = t_j^{(k,m)} - t_{j-1}^{(k,m)}$. Alternatively, this interval is formed as $Y_j^{(k,m)} = W_j^{(k,m)} + S_j^{(k,m)}$, where $W_j^{(k,m)}$ and $S_j^{(k,m)}$ indicate the waiting and service time variables, respectively. The waiting time denotes the span between the j-th packet and the previously delivered one at the same layer, which is written as $W_j^{(k,m)} = \sum_{a=1}^{A_j^{(k,m)}} Z_a^{(k)}$. In this definition, $A_j^{(k,m)}$ follows a geometric distribution with success probability $(1-\psi_k^{(m)})q_{l_k^{(m)}}$

and indicates the number of packets discarded until the arrival of the j-th packet from the selected set. Herein, $0 \leq \psi_k^{(m)} < 1$ denotes the semantics-aware drop factor, which is analyzed in Section IV-A. Furthermore, $Z_a^{(k)}$ is the time between two arrivals, which is exponentially distributed with rate λ_k. Therefore, the admitted arrivals are generated under a Poisson process with rate $\lambda_k(1-\psi_k^{(m)})q_{l_k^{(m)}}$. Besides, the service time corresponds to the duration an update packet spends in the DMS until being completely decoded at its destination. The service time analysis and channel error control protocols are investigated in Section IV-B.

III. SEMANTIC VALUE ASSESSMENT OF UPDATE PACKETS

In this section, we present the semantic value extraction and assessment of both arrival and successfully decoded packets. This is performed at two different scales, namely a *microscopic* one, which is related to the importance of the arrivals at the information source, and a *mesoscopic* one, which captures the importance of received packets at a link level [13].

A. Microscopic Scale

At the source level, the relative importance of an arrival packet depends on *extrinsic* and *intrinsic* features, which in turn could be either *time-sensitive* or *time-tolerant*. Extrinsic features are related to the smart sensor characteristics, such as spatial location, resolution and measurement quality, battery level, and reliability. Intrinsic features depend on the properties of a packet related to goal/application requirements, e.g., probability of occurrence, urgency, and loss risk.

1) Meta-value assignment: The goal-dependent importance of each packet can take on the form of a *meta-value*, which is characterized by a function of the above features. In general, the meta-value of each packet results from *knowledge fusion* or *aggregation* of A intrinsic and B extrinsic features. To overcome the limitations of aggregation functions in the form of weighted sums, we resort to knowledge fusion [29], which takes into account possible interdependence between different features/criteria and provides commensurate scales representing the attributes using the Choquet integral [30]. Specifically, using the discrete Choquet integral, a general meta-value formula at one arrival interval for the i-th realization of the k-th SSM connected to the m-th MM is given by

$$v_i^{(k,m)} = \underbrace{\prod_{a=1}^{A}(g_{a,i}^{(k,m)})^{\alpha_a^{(k,m)}}}_{\text{Intrinsic features}} \underbrace{\left(\sum_{b=1}^{B}\left(U_b^{(k,m)}-U_{b-1}^{(k,m)}\right)(h_b^{(k,m)})^{\bar{\alpha}_b^{(k,m)}}\right)}_{\text{Extrinsic features}} \qquad (2)$$

where $\alpha_a^{(k,m)}$ and $\bar{\alpha}_b^{(k,m)}$ are constant factors. Furthermore, defining $w_b^{(k,m)} \geq 0$ as the weight of the b-th extrinsic feature, $U_b^{(k,m)}$ as the weight of the b-th subset of extrinsic features is given by

$$U_b^{(k,m)} = \left(1 + \lambda_g^{(k,m)} w_b^{(k,m)}\right) U_{b-1}^{(k,m)} + w_b^{(k,m)} \tag{3}$$

where $U_0^{(k,m)} = 0$, $U_1^{(k,m)} \leq U_2^{(k,m)} \leq ... \leq 1$, $U_B^{(k,m)} = 1$, and $\lambda_g^{(k,m)} \geq -1$ comes from the Sugeno fuzzy measure [31], as $1 + \lambda_g^{(k,m)} = \prod_{b=1}^{B}\left(1 + \lambda_g^{(k,m)} w_b^{(k,m)}\right)$. For $\lambda_g^{(k,m)} = 0$, we have $U_b^{(k,m)} = U_{b-1}^{(k,m)} + w_b^{(k,m)}$, and the extrinsic part's fusion form in (2) becomes a weighted sum. Also, $g_{a,i}^{(k,m)} : \mathbb{R} \to \mathbb{R}_0^+$ and $h_b^{(k,m)} : \mathbb{R} \to \mathbb{R}_0^+$ in (2) denote value functions (VFs) of intrinsic and extrinsic features, respectively. Here, every member of $g_{a,i}^{(k,m)}$, $\forall a, i, k, m$, or $h_b^{(k,m)}$, $\forall b, k, m$, with known parameters is presented via a general function $\mathrm{VF} : \mathbb{R} \to \mathbb{R}_0^+$. Thus, for A intrinsic and B extrinsic features, we have $A + B$ different versions of VF. Conventionally, we model VF as a sum of Gaussian functions with predefined critical points z_{n_c}, for $n_c = 1, 2, ..., N_c$, relative criticality $\varpi_{n_c} \geq 0$, and minimum importance $\mathrm{VF}_{\min} \geq 0$. For a sample point z, we can write

$$\mathrm{VF}(z) = \frac{\widehat{\mathrm{VF}}(z)}{\max_z\left\{\widehat{\mathrm{VF}}(z)\right\}} \tag{4}$$

where, if $z_{\min} \leq z \leq z_{\max}$, we have

$$\widehat{\mathrm{VF}}(z) = \sum_{n_c=1}^{N_c} \varpi_{n_c} e^{-\frac{(z - z_{n_c})^2}{2\sigma^2}}. \tag{5}$$

In (5), the standard deviation σ is derived such that all sample points of VF get higher importance than a given threshold $\mathrm{VF}_{\min}$. For that, with given N_c, z_{n_c}, and ϖ_{n_c}, σ starts from a small value and increases with a fixed step over a finite number of iterations until $\mathrm{VF}(z) \geq \mathrm{VF}_{\min}$, $\forall z$. An illustrative example how to compute VF according to (4) and (5) is given in Appendix A.

The proposed value assessment scheme can operate under both pull-based and push-based communication models. Assume a packet from realization $x_i^{(k)}$, $\forall i \in \mathcal{I}_{l_k^{(m)}}$ arrives at the j-th interval and is assigned importance value $v_j^{(k,m)}$. In the push-based model, the value of the generated packet comes from (2), i.e., $v_j^{(k,m)} = v_i^{(k,m)}$, whereas the pull-based policy adds a constraint such that $v_j^{(k,m)} = v_i^{(k,m)}$ only if the m-th MM has requested an update from the k-th SSM at the j-th interval; otherwise, $v_j^{(k,m)} = 0$.

B. Mesoscopic Scale

At the link level, the importance of a received packet at a destination is measured using the *semantics of information* (SoI) [13], [32]. In this regard, we consider *timeliness* as a contextual attribute of information, which is a non-increasing function $f_k : \mathbb{R}_0^+ \to \mathbb{R}$ of the freshness of information from the k-th sensor. The instantaneous SoI provided by the k-th SSM, $\forall k \in \mathcal{K}(m)$, for the m-th MM at time t is modeled as

$$\mathcal{S}_k^{(m)}(t) = f_k(\Delta_k^{(m)}(t)) \tag{6}$$

where $\Delta_k^{(m)}(t) = t - u(t)$ denotes the instantaneous AoI at the m-th MM, which is defined as the difference between the current time instant t and the timestamp $u(t)$ of the most recently arrived update from the considered k-th SSM or another identical sensor connected to that monitor.

The average SoI over an observation interval $(0, T)$, assuming a stationary ergodic process, is given by

$$\bar{\mathcal{S}}_k^{(m)} = \lim_{T \to \infty} \frac{1}{T} \int_0^T f_k(\Delta_k^{(m)}(t)) \mathrm{d}t. \tag{7}$$

From a monitor's perspective, the SoI at a layer reserved for the k-th SSM decreases according to $f_k(\cdot)$ until a valuable status update for that layer arrives. Then, the SoI rises to the value of the new update at that time. Therefore, dissimilar monitors potentially attain different SoIs over similar time spans.

In this paper, we study the following forms for the SoI, namely *exponential* (E-), *logarithmic* (L-), and *reciprocal* utility of timeliness (RUT) cases, i.e.,

$$f_k(\Delta_k^{(m)}(t)) = \begin{cases} \exp(-\rho_k^{(m)}(t)\Delta_k^{(m)}(t)) + \beta_k^{(m)} & \text{EUT case} \\ \ln(-\rho_k^{(m)}(t)\Delta_k^{(m)}(t)) + \beta_k^{(m)} & \text{LUT case} \\ (\rho_k^{(m)}(t)\Delta_k^{(m)}(t))^{-\kappa} + \beta_k^{(m)} & \text{RUT case} \end{cases} \tag{8}$$

where $\kappa \in \mathbb{Z}^+$ and $\beta_k^{(m)}$ are constant parameters, and $\rho_k^{(m)}(t) > 0$ is an attenuation factor that is fixed during each update interval of the k-th SSM and is reinitialized for a new arrival.[2] The value of the attenuation factor at an update interval comes from the importance of that admitted

---

[2] The analysis is extendable for any (single or composite) non-increasing forms for the utility of timeliness.

arrival at the source level. Hence, we can define

$$\rho_k^{(m)}(t) := \rho_j^{(k,m)} \mathbb{1}\left\{t_{j-1}^{(k,m)} + S_j^{(k,m)} \leq t \leq t_j^{(k,m)} + S_{j+1}^{(k,m)}\right\} \tag{9}$$

where $\rho_j^{(k,m)} = \rho_{\min} + \left(\max_{i \in \mathcal{I}_{l_k^{(m)}}}\left\{v_i^{(k,m)}\right\} - v_j^{(k,m)}\right)\left(\rho_{\max} - \rho_{\min}\right)$ for the j-th admitted arrival. $\rho_{\min}$ and $\rho_{\max}$ are the minimum and maximum values that the attenuation factors may attain, respectively. Evidently, the higher the importance of an arrival, the lower attenuation factor it is assigned.

IV. SEMANTIC FILTERING AND CHANNEL ERROR CONTROL

In this section, we introduce a semantics-aware packet filtering mechanism and analyze two asynchronous error control schemes.

A. Semantics-Aware Filtering

We consider two semantics-aware packet filtering schemes, which control the flow of update packet arrivals through different levels of each SSM, namely *fixed-* and *adaptive-length* filtering. In the light of assessed meta-values and freshness, the rationale of these filtering mechanisms is to feed the transmitter with important arrivals, while keeping the blockage rate as low as possible. The blockage rate quantifies the number of blockages occurring to the admitted packets due to heavy channel load over the total number of arrival packets.

1) Fixed-length filtering: A fixed-length filter merely admits packets generated from the $l_k^{(m)}$ most important realizations among all n_k realizations the k-th SSM shares with the m-th MM. Packets of less important realizations, based on the given index set $\mathcal{I}_{l_k^{(m)}}$, are directly discarded, independently of their freshness (e.g., most recent realization) or the channel idleness. Therefore, even the most recent (highest freshness) yet less important packets are blocked.

2) Adaptive-length filtering: In the proposed adaptive-length filtering mechanism, packets' importance at both microscopic and mesoscopic scales is taken into account, which gives a chance to a subset of the least important packets, in addition to the most important ones, to be carried over the DMS. Differently from the fixed-length filtering, only a fraction of the least important packets are directly discarded in the adaptive-length filtering scheme, and a subset of the fresh yet of lower importance arrivals may pass the filter in an adaptive manner. This means that the number of admitted realizations in adaptive-length filtering is larger or equal to that of the fixed-length one. Specifically, upon the arrival of a packet generated from one of the

admitted realizations when the channel is idle, the adaptive-length filter inspects whether that packet can increase the SoI at the link level after being decoded, thanks to its freshness or not. If the inspection outcome is positive, the packet passes the filter. Thus, adaptive-length filtering brings flexibility owing to which packets with lower importance but higher freshness have a chance to pass the filter.

3) Acceptance probability analysis: A key element of adaptive-length filtering is a parameter that quantifies the probability that a packet generated from the selected realizations is rejected (or accepted), considering previously observed updates at the link level. As mentioned in Section II-B, we coin this as semantics-aware drop factor. Using this, a new arrival (status update) that does not increase the SoI upon receipt at the monitor, despite being the most recent/fresh, is rejected, hence not transmitted. To make it clear, Figure 3 illustrates the shape of packet drop by adaptive-length filtering at the k-th SSM connected to the m-th MM under three possible conditions listed below, considering a non-increasing $f_k(\cdot)$, and convex for the purpose of semantic filtering.

I. The first and second arrival are of comparable importance, hence attenuation factors. Therefore, the SoI curves provided by these arrivals intersect at infinity. In this case, the filter does not discard the new arrival since it may offer a higher value upon receipt at its monitor.

II. The j-th arrival has higher importance than the $(j+1)$-th one, where the primary arrival crosses the latter one before both reach $\beta_k^{(m)}$. However, the cross point is after the service time of the new arrival, i.e., $t_j^{(k,m)} + S_{j+1}^{(k,m)}$. Therefore, the filter keeps the $(j+1)$-th packet since it can increase $\mathcal{S}_k^{(m)}(t)$ after its successful decoding.

III. The j'-th arrival has higher importance compared to the $(j'+1)$-th one; thus, the primary arrival crosses the latter one at a point before the delivery of that new packet. In that case, the filter discards the $(j'+1)$-th packet since it cannot bring a higher value at the link level even after its successful delivery. Thanks to this mechanism, the system obtains better $\mathcal{S}_k^{(m)}(t)$, as highlighted in the figure.

The semantics-aware drop factor is denoted by $\psi_k^{(m)}$ for the link between the k-th SSM and m-th MM. Here, $\psi_k^{(m)} = 1$ for the fixed-length filtering, while we can define

$$\psi_k^{(m)} = \lim_{T \to \infty} \frac{1}{\mathcal{N}_k^{(m)}(T)} \sum_{j=1}^{\mathcal{N}_k^{(m)}(T)} \sum_{d=1}^{d_{\max}} \mathbb{1}\left\{ \frac{\rho_{j+d}^{(k,m)}}{\rho_j^{(k,m)}} > \tau_d^{(k,m)} \right\}$$
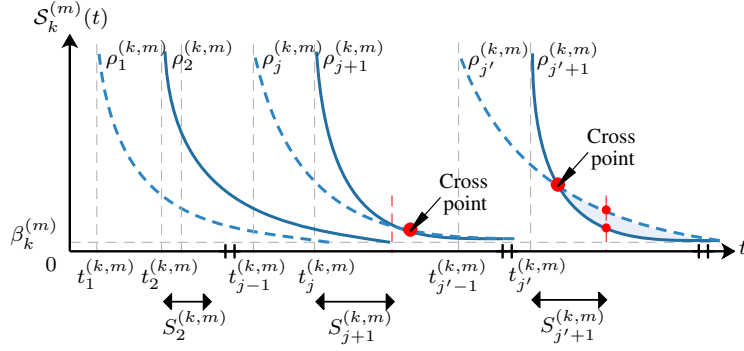
Fig. 3: Three conditions for packet drop via adaptive-length filtering.

$$\simeq \underset{i \in \mathcal{I}_{n_k}}{\mathbb{E}} \left[\frac{1}{d_{\max}} \sum_{d=1}^{d_{\max}} \Pr \left\{ i' \in \mathcal{I}_{l_k^{(m)}} \, \middle| \, \frac{\rho_{i'}^{(k,m)}}{\rho_i^{(k,m)}} > \tau_d^{(k,m)} \right\} \right] \tag{10}$$

for the adaptive-length filtering scheme. In (10), $\mathcal{N}_k^{(m)}(T)$ is the number of all admitted packets at the k-th SSM for the m-th MM by time T, and $d_{\max} \ll \mathcal{N}_k^{(m)}(T)$ for $T \to \infty$. Also, $\tau_d^{(k,m)}$ denotes a threshold for the drop of order d, which is computed in the following lemma.

Lemma 1. *The threshold $\tau_d^{(k,m)}$ in (10) is derived as*

$$\tau_d^{(k,m)} = (d+1) + \frac{\widehat{W}_d^{(k,m)}}{\ell_{\max}} \tag{11}$$

where $\widehat{W}_d^{(k,m)}$ is an Erlang r.v. of order d with rate $\lambda_k q_{l_k^{(m)}}$, and $\ell_{\max}$ indicates an upper bound to which the size of a codeword length converges. Notably, we reach $\tau_d^{(k,m)} = d+1$ if $\ell_{\max} \to \infty$.

Proof. See Appendix B.

B. Channel Error Control and Service Time Analysis

We utilize two non-adaptive asynchronous error control schemes for handling transmission errors over block fading channels, namely ARQ and truncated HARQ based on chase combing. In this regard, an SSM connected to one of the served MMs is equipped with an individual buffer to support ARQ and HARQ protocols. The successful delivery of each packet to an MM is declared by an instantaneous and error-free acknowledgment (ACK) feedback to the serving SSM (see Figure 2). However, in the event of failure at each monitor, the serving SSM retransmits the packet only to the MM from which a negative ACK (NACK) message is received. Packets are retransmitted either until successful reception or up to the maximum allowable number of

transmission rounds. After successful delivery to its destination or reaching the retransmission limit (in which case the packet is dropped), the transmitter waits for a new admitted arrival.[3]

Consequently, the service time of packet $x_i^{(k)}$, which is tagged important for the m-th MM during the j-th arrival, is a function of its codeword length, the channel conditions, and the number of transmission rounds according to the error control protocol. Thus, we consider $\mathbb{E}[(S^{(k,m)})^c] = \mathbb{E}\big[\varphi^{(k,m)}(L^{(k,m)})^c\big]$ the c-th moment of the service time for packets being delivered to the m-th MM from the k-th SSM, where

$$\mathbb{E}\big[\varphi^{(k,m)}(L^{(k,m)})^c\big] = \lim_{T\to\infty} \frac{1}{\mathcal{N}_k^{(m)}(T)} \sum_{j=1}^{\mathcal{N}_k^{(m)}(T)} \sum_{r=1}^{r_{\max,j}^{(k,m)}} \varphi_{j,r}^{(k,m)}(\ell_j^{(k,m)})^c \tag{12}$$

since $f_k(\cdot)$ is non-increasing. Herein, r is the order of transmission and truncated by $r_{\max,j}^{(k,m)} \ll \mathcal{N}_k^{(m)}(T)$, where $r_{\max,j}^{(k,m)}$ depends on the j-th arrival's meta-value, i.e., $v_j^{(k,m)}$. As a simple arbitrary form, we can write

$$r_{\max,j}^{(k,m)} = \left(\frac{v_j^{(k,m)}}{\frac{1}{l_k^{(m)}} \sum_{i \in \mathcal{I}_{l_k^{(m)}}} v_i^{(k,m)}} \right) r_{\max} \tag{13}$$

for the j-th arrival. According to (13), higher sample/packet importance results in larger $r_{\max,j}^{(k,m)}$. Besides, $\varphi_{j,r}^{(k,m)}$ is relevant to error control processes for the j-th arrival packet that can be transmitted up to r times before reaching $r_{\max,j}^{(k,m)}$. In this regard, $\varphi_{j,r}^{(k,m)}$ is given by [33]

$$\varphi_{j,r}^{(k,m)} = \frac{c_j^{(k,m)}}{1 - \theta_{r,j}^{(k,m)}} \left(r\theta_{r,j}^{(k,m)} + \sum_{r'=1}^{r} r'\left(\theta_{r'-1,j}^{(k,m)} - \theta_{r',j}^{(k,m)} \right) \right) \tag{14}$$

where $c_j^{(k,m)} \geq 1$ indicates the reverse of the channel coding rate with $c_j^{(k,m)} = 1$ for the ARQ protocol, and $\theta_{r,j}^{(k,m)}$ denotes the probability that the first r transmissions of the j-th packet from the k-th SSM to the m-th MM are performed with error. For the chase-combining HARQ, $\theta_{r,j}^{(k,m)}$, $\forall r, j, k, m \in \mathcal{M}(k)$, is given by [34]

$$\theta_{r,j}^{(k,m)} = e^{-\frac{\gamma_{M,j}^{(k,m)}}{\bar{\gamma}_k^{(m)}}} \times \left(\sum_{i=r}^{\infty} \frac{\left(\frac{\gamma_{M,j}^{(k,m)}}{\bar{\gamma}_k^{(m)}} \right)^i}{i!} + \sum_{i=0}^{r-1} \frac{\left(\frac{\gamma_{M,j}^{(k,m)}}{\bar{\gamma}_k^{(m)}} \right)^i}{i!} \prod_{i'=1}^{r-i} \frac{1}{1 + i'g_j^{(k,m)}\bar{\gamma}_k^{(m)}} \right), \tag{15}$$

---

[3]The computation and propagation delays are assumed negligible for both packet transmission and feedback processes.

while for ARQ we have

$$\theta_{r,j}^{(k,m)} = \left(1 - \frac{g_j^{(k,m)}\bar{\gamma}_k^{(m)}}{1 + g_j^{(k,m)}\bar{\gamma}_k^{(m)}}e^{-\frac{\gamma_{M,j}^{(k,m)}}{\bar{\gamma}_k^{(m)}}}\right)^r. \tag{16}$$

Expressions (15) and (16) are obtained by applying an approximated expression for packet error rates as a function of the signal-to-noise ratio (SNR) for Rayleigh fading channels [34], [35], where $\bar{\gamma}_k^{(m)}$ is the average received SNR at the m-th MM from the k-th SSM. Fitting the approximated expression to the exact formula, we find $\gamma_{M,j}^{(k,m)}$ and $g_j^{(k,m)}$ under a considered modulation scheme.

V. SEMANTIC SOURCE CODING

In this section, we formulate the problem of semantics-aware packet encoding, which aims to maximize the semantic value of packets delivered at the served monitors, and whose solution provides the optimal codeword lengths assigned to admitted packets at each sensor.

A. Problem Statement

The objective of optimal codeword length assignment is to maximize the weighted sum of the average SoIs provided at all layers of MMs, subject to the following two constraints: (i) codeword lengths should be positive integers, i.e., $\ell_i^{(k,m)} \in \mathbb{Z}^+$ (feasibility); (ii) existence of a prefix-free (or uniquely decodable) code for a given set of codeword lengths at each SSM (Kraft-McMillan inequality [36]). Therefore, the optimization problem is formulated as

$$\mathcal{P}_1: \underset{\{\ell_i^{(k,m)}\}}{\text{maximize}} \quad \sum_{m=1}^{M} w_m \sum_{k\in\mathcal{K}(m)} p_k^{(m)} \bar{\mathcal{S}}_k^{(m)} := \sum_{m=1}^{M} \sum_{k\in\mathcal{K}(m)} \mathcal{J}_{\text{SoI}}^{(k,m)}$$

$$\text{subject to} \begin{cases} \mathcal{C}_1 : \sum_{i\in\mathcal{I}_{l_k^{(m)}}} 2^{-\ell_i^{(k,m)}} \leq 1, \forall k,m \\ \mathcal{C}_2 : \ell_i^{(k,m)} \in \mathbb{Z}^+, \forall i,k,m \end{cases} \tag{17}$$

where w_m denotes a weight parameter.

To derive $\bar{\mathcal{S}}_k^{(m)}$, we use different forms of $f_k(\cdot)$ defined in (8), considering the type and functionality of the k-th sensor. Thus, from (7) and (8), the average SoI can be computed for the proposed cases. To do so, we divide the non-negative area below the curve of $f_k(\Delta_k^{(m)}(t))$
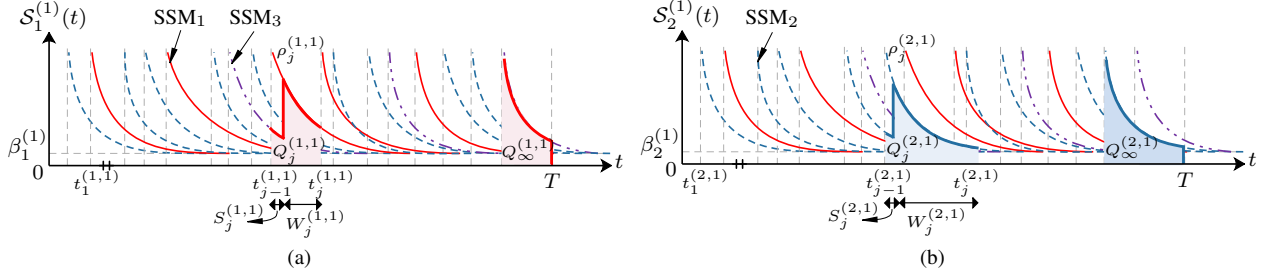
Fig. 4: Sample evolution for the EUT case, where (a) MM_1 receives updates from identical SSM_1 and SSM_3, and (b) MM_2 receives packets from SSM_2.

over interval $(0, T)$ into polygons of $Q_j^{(k,m)}$, $j = 1, 2, ..., \mathcal{N}_k^{(m)}(T)$, and $Q_\infty^{(k,m)}$, as depicted in Figure 4 for the EUT case. Thus, we can rewrite (7) as

$$\bar{\mathcal{S}}_k^{(m)} = \lim_{T\to\infty} \frac{1}{T} \left(\sum_{j=1}^{\mathcal{N}_k^{(m)}(T)} Q_j^{(k,m)} + Q_\infty^{(k,m)} \right) = \eta_k^{(m)} \mathbb{E}[Q^{(k,m)}] \tag{18}$$

where $\eta_k^{(m)} = \lim_{T\to\infty} \frac{1}{T} \left(\mathcal{N}_k^{(m)}(T) - 1 \right)$ denotes the steady-state time average arrival rate. Importing (18) into (17), we reach the following problem.

$$\mathcal{P}_2: \underset{\{\ell_i^{(k,m)}\}}{\text{maximize}} \quad \sum_{m=1}^{M} w_m \sum_{k\in\mathcal{K}(m)} p_k^{(m)} \eta_k^{(m)} \mathbb{E}[Q^{(k,m)}]$$

$$\text{subject to} \begin{cases} \mathcal{C}_1 : \sum_{i\in\mathcal{I}_{l_k^{(m)}}} 2^{-\ell_i^{(k,m)}} \leq 1, \forall k, m \\ \hat{\mathcal{C}}_2 : \ell_i^{(k,m)} \in \mathbb{R}^+, \forall i, k, m. \end{cases} \tag{19}$$

where constraint $\hat{\mathcal{C}}_2$ is a relaxation of $\mathcal{C}_2$ in (17) to allow for non-negative real-valued codeword lengths. The solution of $\mathcal{P}_2$ gives real-valued codeword lengths, whose corresponding integer values can be found using a rounded-off operation.

B. Semantics-Aware Encoding Design

To solve the problem $\mathcal{P}_2$, in what follows, we find $\mathbb{E}[Q^{(k,m)}]$ based on different forms of $f_k(\Delta_k^{(m)}(t))$ defined in (8).

1) EUT case: Considering the exponential utility function of timeliness, we propose the following lemma.

Lemma 2. *The expected form of polygons $Q_j^{(k,m)}$ for the EUT case is approximately derived as*

$$\mathbb{E}[Q^{(k,m)}] \approx \left(1 + \beta_k^{(m)} - \bar{\rho}_k^{(m)}\gamma_k^{(m)} - \mathbb{E}\big[\rho^{(k,m)}\varphi^{(k,m)}L^{(k,m)}\big]\right)\mathbb{E}\big[\varphi^{(k,m)}L^{(k,m)}\big] - \frac{1}{2}\bar{\rho}_k^{(m)}\mathbb{E}\big[\varphi^{(k,m)}(L^{(k,m)})^2\big]$$

$$- \gamma_k^{(m)}\mathbb{E}\big[\rho^{(k,m)}\varphi^{(k,m)}L^{(k,m)}\big] - \bar{\rho}_k^{(m)}(\gamma_k^{(m)})^2 + (1+\beta_k^{(m)})\gamma_k^{(m)} \tag{20}$$

where $\gamma_k^{(m)} := 1/(\lambda_k \psi_k^{(m)} q_{l_k^{(m)}})$ and $\bar{\rho}_k^{(m)} := \mathbb{E}[\rho^{(k,m)}] = \lim_{T\to\infty}\frac{1}{T}\sum_{j=1}^{\mathcal{N}_k^{(m)}(T)} \rho_j^{(k,m)} = \sum_{i\in\mathcal{I}_{l_k^{(m)}}} p_i^{(k,m)}\rho_i^{(k,m)}$.

Proof. See Appendix C.

Importing (20) into (19), $\mathcal{P}_2$ becomes a convex problem and can be solved using standard solvers, e.g., MOSEK in CVX. Alternatively, we propose a heuristic solution to compute closed-form expressions for the codeword length, as follows.

Proposition 1. *The codeword lengths $\ell_i^{(k,m)}$, $\forall i \in \mathcal{I}_{l_k^{(m)}}$, which maximize (19) in $\mathcal{P}_2$ for the EUT case are computed as*

$$\ell_i^{(k,m)} = -\log_2\left(\frac{\bar{\rho}_k^{(m)}p_i^{(k,m)}\varphi_i^{(k,m)}}{\mu_k^{(m)}(\ln(2))^2}\mathcal{W}_0\left(\frac{\mu_k^{(m)}(\ln(2))^2}{\bar{\rho}_k^{(m)}p_i^{(k,m)}\varphi_i^{(k,m)}}2^{\xi_k^{(m)}}\right)\right) \tag{21}$$

where $\mu_k^{(m)} \geq 0$ is a constant multiplier,

$$\xi_k^{(m)} := \frac{2\chi_k^{(m)}\mu_k^{(m)}\ln(2) + \bar{\rho}_k^{(m)}\gamma_k^{(m)}\big(1+\chi_k^{(m)}\big) - \big(1+\beta_k^{(m)}\big)}{\bar{\rho}_k^{(m)}\big(1+2\chi_k^{(m)}\bar{\varphi}_k^{(m)}\big)}, \tag{22}$$

$\bar{\varphi}_k^{(m)} := \mathbb{E}[\varphi^{(k,m)}] = \lim_{T\to\infty}\frac{1}{T}\sum_{j=1}^{\mathcal{N}_k^{(m)}(T)}\varphi_j^{(k,m)}$, *and*

$$\chi_k^{(m)} = \frac{\mathbb{E}\big[\rho^{(k,m)}\varphi^{(k,m)}L^{(k,m)}\big]}{\bar{\rho}_k^{(m)}\mathbb{E}[\varphi^{(k,m)}L^{(k,m)}]}. \tag{23}$$

Proof. See Appendix D.

The optimal codeword lengths of delivered packets from the k-th SSM to the m-th MM can be found via (21) for $(\mu_k^{(m)}, \chi_k^{(m)})$ pair, subject to $\mu_k^{(m)} > 0$ and $\sum_{i\in\mathcal{I}_{l_k^{(m)}}} 2^{-\ell_i^{(k,m)}} = 1$. The values of $\chi_k^{(m)}$ and $\mu_k^{(m)}$ can be obtained using Algorithm 1, through its *inner* and *outer* loops, respectively. We first assume similar importance, hence attenuation factors, for all packets and initialize $\chi_k^{(m)} = 1$. Then, given a small value of $\mu_k^{(m)}$, we compute $\ell_i^{(k,m)}$ and new $\chi_k^{(m)}$ through the inner loop. Thereafter, based on the found $\chi_k^{(m)}$, we derive new values for $\ell_i^{(k,m)}$ in the

---

Algorithm 1: Solution for deriving $\mu_k^{(m)}$ and $\chi_k^{(m)}$

    Input: Fixed parameters $\mathcal{I}_{l_k^{(m)}}$, $p_i^{(k,m)}$, $\rho_i^{(k,m)}$, and $\varphi_i^{(k,m)}$, $\forall i \in \mathcal{I}_{l_k^{(m)}}$, and $\beta_k^{(m)}$. Stopping
           accuracy ε. Initial parameters $(\mu_k^{(m)})^{(0)}$, $(\chi_k^{(m)})^{(0)}$, $(\xi_k^{(m)})^{(0)}$, and $(\ell_i^{(k,m)})^{(0)}$,
           $\forall i \in \mathcal{I}_{l_k^{(m)}}$.

    Output: Computed parameters $\chi_k^{(m)} = (\chi_k^{(m)})^{(b)}$, $\xi_k^{(m)} = (\xi_k^{(m)})^{(b)}$, $\ell_i^{(k,m)} = (\ell_i^{(k,m)})^{(b)}$,
           $\forall i \in \mathcal{I}_{l_k^{(m)}}$, and $\mu_k^{(m)} = (\mu_k^{(m)})^{(a)}$.

1 *Iteration a*:   ▷ Outer loop

2 *Iteration b*:   ▷ Inner loop

3 Compute $(\xi_k^{(m)})^{(b)}$ and $(\ell_i^{(k,m)})^{(b)}, \forall i \in \mathcal{I}_{l_k^{(m)}}$, using (22) and (21), respectively.

4 Calculate $\mathbb{E}\big[\rho^{(k,m)}\varphi^{(k,m)}L^{(k,m)}\big]$ and $\bar{\rho}_k^{(m)}\mathbb{E}\big[\varphi^{(k,m)}L^{(k,m)}\big]$.

5 Update $(\chi_k^{(m)})^{(b)}$ from (23) based on **4**.

6 **if** *Criterion* $\big|(\chi_k^{(m)})^{(b)} - (\chi_k^{(m)})^{(b-1)}\big| > \varepsilon$ **then** set $b = b + 1$, and **goto 2**.

7 Update $(\xi_k^{(m)})^{(b)}$ and $(\ell_i^{(k,m)})^{(b)}$ according to **5**.

8 **if** $\sum_{i \in \mathcal{I}_{l_k^{(m)}}} 2^{-(\ell_i^{(k,m)})^{(b)}} = 1$ **then** stop the process, and **goto 11**.

9 **else if** $\sum_{i \in \mathcal{I}_{l_k^{(m)}}} 2^{-(\ell_i^{(k,m)})^{(b)}} < 1$ **then** decrease $(\mu_k^{(m)})^{(a)}$, set $a = a + 1$, and **goto 1**.

10 **else** increase $(\mu_k^{(m)})^{(a)}$, set $a = a + 1$, and **goto 1**.

11 Save $(\chi_k^{(m)})^{(b)}$, $(\xi_k^{(m)})^{(b)}$, $(\ell_i^{(k,m)})^{(b)}$, $\forall i \in \mathcal{I}_{l_k^{(m)}}$, and $(\mu_k^{(m)})^{(a)}$.

---

next iteration. This process continues until we reach the stopping accuracy ε. Once reached, the outer loop checks the Kraft-McMillan condition and resets $\mu_k^{(m)}$ if the condition is not satisfied. Subsequently, the inner loop starts again, considering the new value of the constant multiplier. Finally, the algorithm converges to the final amounts of $\chi_k^{(m)}$ and $\mu_k^{(m)}$ with the rate of $\mathcal{O}((N_a N_b)^{-1})$ in which N_a and N_b denote the maximum numbers of outer and inner iterations, respectively.

   2) LUT case: Under a logarithmic utility function of timeliness, $\mathbb{E}[Q^{(k,m)}]$ is found as follows.

Lemma 3. *The expected value of $Q_j^{(k,m)}$ for the LUT case is approximately derived as*

$$
\begin{aligned}
\mathbb{E}[Q^{(k,m)}] \approx & -2\mathbb{E}\big[\rho^{(k,m)}\varphi^{(k,m)}L^{(k,m)}\big]\,\mathbb{E}\big[\varphi^{(k,m)}L^{(k,m)}\big] - \bar{\rho}_k^{(m)}\mathbb{E}\big[\varphi^{(k,m)}(L^{(k,m)})^2\big] \\
& -2\gamma_k^{(m)}\mathbb{E}\big[\rho^{(k,m)}\varphi^{(k,m)}L^{(k,m)}\big] + \Big(\beta_k^{(m)} - 1 - 2\bar{\rho}_k^{(m)}\gamma_k^{(m)}\Big)\mathbb{E}\big[\varphi^{(k,m)}L^{(k,m)}\big] \\
& -2\bar{\rho}_k^{(m)}(\gamma_k^{(m)})^2 + (\beta_k^{(m)} - 1)\gamma_k^{(m)}.
\end{aligned}
\tag{24}
$$

Proof. See Appendix E.

Inserting (24) into (19), $\mathcal{P}_2$ is a convex problem, and the optimal codeword lengths $\ell_i^{(k,m)}$ can be obtained using either standard solvers or the following expression

$$\ell_i^{(k,m)} = -\log_2\left(\frac{\bar{\rho}_k^{(m)} p_i^{(k,m)} \varphi_i^{(k,m)}}{\hat{\mu}_k^{(m)}(\ln(2))^2} \mathcal{W}_0\left(\frac{\hat{\mu}_k^{(m)}(\ln(2))^2}{\bar{\rho}_k^{(m)} p_i^{(k,m)} \varphi_i^{(k,m)}} 2^{\hat{\xi}_k^{(m)}}\right)\right), \tag{25}$$

which is derived with the same method as in (21). Herein, $\hat{\mu}_k^{(m)} \geq 0$ indicates a constant. We also have $\hat{\xi}_k^{(m)} := \dfrac{4\hat{\chi}_k^{(m)} \hat{\mu}_k^{(m)} \ln(2) + 2\bar{\rho}_k^{(m)} \gamma_k^{(m)}\left(1+\hat{\chi}_k^{(m)}\right) - \left(\beta_k^{(m)}-1\right)}{2\bar{\rho}_k^{(m)}\left(1+2\hat{\chi}_k^{(m)} \bar{\varphi}_k^{(m)}\right)}$. The optimal codeword lengths are computed in (25) based on the known pair of $(\hat{\mu}_k^{(m)}, \hat{\chi}_k^{(m)})$, subject to $\hat{\mu}_k^{(m)} > 0$ and $\sum_{i \in \mathcal{I}_{l_k^{(m)}}} 2^{-\ell_i^{(k,m)}} = 1$. Similar to the EUT case, we can find the values of $\hat{\mu}_k^{(m)}$ and $\hat{\chi}_k^{(m)}$ using Algorithm 1 replacing its parameters with those for the LUT case.

3) RUT case: Following the same steps as for deriving (20) and (24), the expected form of polygons $Q_j^{(k,m)}$ for the RUT case is obtained as

$$\begin{aligned}
\mathbb{E}[Q^{(k,m)}] \approx &-\kappa\mathbb{E}\left[\rho^{(k,m)}\varphi^{(k,m)}L^{(k,m)}\right]\mathbb{E}\left[\varphi^{(k,m)}L^{(k,m)}\right] - \frac{\kappa}{2}\mathbb{E}\left[\rho^{(k,m)}\varphi^{(k,m)}(L^{(k,m)})^2\right] \\
&- 2\kappa\gamma_k^{(m)}\mathbb{E}\left[\rho^{(k,m)}\varphi^{(k,m)}L^{(k,m)}\right] + \left(\frac{\kappa}{\kappa+1}+\beta_k^{(m)}-\kappa\bar{\rho}_k^{(m)}\gamma_k^{(m)}\right)\mathbb{E}\left[\varphi^{(k,m)}L^{(k,m)}\right] \\
&- 2\kappa\bar{\rho}_k^{(m)}(\gamma_k^{(m)})^2 + \left(\frac{\kappa}{\kappa+1}+\beta_k^{(m)}\right)\gamma_k^{(m)}.
\end{aligned} \tag{26}$$

Consequently, the optimal codeword lengths are given as

$$\ell_i^{(k,m)} = -\log_2\left(\frac{\bar{\rho}_k^{(m)} p_i^{(k,m)} \varphi_i^{(k,m)}}{\hat{\hat{\mu}}_k^{(m)}(\ln(2))^2} \mathcal{W}_0\left(\frac{\hat{\hat{\mu}}_k^{(m)}(\ln(2))^2}{\bar{\rho}_k^{(m)} p_i^{(k,m)} \varphi_i^{(k,m)}} 2^{\hat{\hat{\xi}}_k^{(m)}}\right)\right) \tag{27}$$

where $\hat{\hat{\mu}}_k^{(m)} \geq 0$, and $\hat{\hat{\xi}}_k^{(m)} := \dfrac{2\kappa\hat{\hat{\chi}}_k^{(m)} \hat{\hat{\mu}}_k^{(m)} \ln(2) + \kappa\bar{\rho}_k^{(m)} \gamma_k^{(m)}\left(1+2\hat{\hat{\chi}}_k^{(m)}\right) - \left(\frac{\kappa}{\kappa+1}+\beta_k^{(m)}\right)}{\kappa\bar{\rho}_k^{(m)}\left(1+2\hat{\hat{\chi}}_k^{(m)} \bar{\varphi}_k^{(m)}\right)}$. The values of $\hat{\hat{\mu}}_k^{(m)}$ and $\hat{\hat{\chi}}_k^{(m)}$ are calculated using Algorithm 1 with parameters for the RUT case.

C. Asymptotic Expansions for Codeword Lengths

We provide here two asymptotic expansions for the derived closed-form codeword lengths in (21), (25), and (27), for the EUT, LUT, and RUT cases, respectively.

1) Dependency on occurrence probability and transmission rounds: As shown in Section V-B, the codeword length assigned to a realization (update packet) depends - among others - on its probability of occurrence and the number of transmission rounds. Using

$$\left[\frac{\mathcal{W}_0(y)}{y}\right]^z = \exp\left(-z\mathcal{W}_0(y)\right), \forall y \geq e, \tag{28}$$

from the Laurent series of order z, and $\mathcal{W}_0(y) = \ln(y) - \ln(\ln(y)) + \mathcal{O}(1)$ from the second-order Taylor expansion for large enough y, we can write

$$\ell_i^{(k,m)} \propto \frac{\max_i\left\{p_i^{(k,m)}\varphi_i^{(k,m)}\right\}}{p_i^{(k,m)}\varphi_i^{(k,m)}} - \ln\left(\frac{\max_i\left\{p_i^{(k,m)}\varphi_i^{(k,m)}\right\}}{p_i^{(k,m)}\varphi_i^{(k,m)}}\right). \tag{29}$$

From the asymptotic expression in (29), we deduce that the codeword length monotonically decreases (increases) by an increase in the occurrence probability, the number of transmissions, or both when $p_i^{(k,m)}\varphi_i^{(k,m)} \leq 1$ ($p_i^{(k,m)}\varphi_i^{(k,m)} > 1$). Besides, $p_i^{(k,m)}\varphi_i^{(k,m)} \to 0$ yields $\ell_i^{(k,m)} \to \ell_{\max}$, $\forall i \in \mathcal{I}_{l_k^{(m)}}$, where $\ell_{\max}$ denotes the upper bound for the size of a codeword length. Furthermore, if $\varphi_i^{(k,m)} = \varphi_{i'}^{(k,m)}$, $\forall i \neq i'$, the codeword length of a packet becomes proportional to the reverse of its probability. The *higher* the probability of occurrence of a realization, the *shorter* its assigned codeword length, and vice versa.

2) Uniform sources and equal semantic value: For a uniform source, i.e., $p_i^{(k,m)} = p_k^{(m)}/n_k$, $\forall i$, with equal goal-oriented importance, e.g., $v_i^{(k,m)} = 1$, and the same number of transmissions, i.e., $\varphi_i^{(k,m)} = \varphi_{i'}^{(k,m)}$, $\forall i \neq i'$, we end up with codewords of equal size, i.e., $\ell_i^{(k,m)} = \ell_{i'}^{(k,m)}$, where

$$\ell_i^{(k,m)} \propto \frac{n_k}{p_k^{(m)}} - \ln\left(\frac{n_k}{p_k^{(m)}}\right), \forall i \in \mathcal{I}_{l_k^{(m)}}. \tag{30}$$

Thus, $n_k \gg 1$ results in $\ell_i^{(k,m)} \to \ell_{\max}$, which remains almost fixed for very large n_k. Therefore, as the number of realizations *increases*, the assigned codeword lengths become *longer* to satisfy the Kraft-McMillan condition at the expense of *higher* service time. On the other hand, (30) indicates that the codeword lengths of packets transferred to a monitor with a *higher* probability of observation from a serving sensor are *shorter* than those with a lower probability of observation from the same sensor. That is, $\ell_i^{(k,m)}$ of packets sent to the m-th SSM from the k-th SSM become shorter with the increase of $p_k^{(m)}$.

VI. SIMULATION RESULTS

In this section, we present simulation results that corroborate our analysis and show the performance gains by properly designing semantic filtering and source coding for timely status update delivery in DMSs.
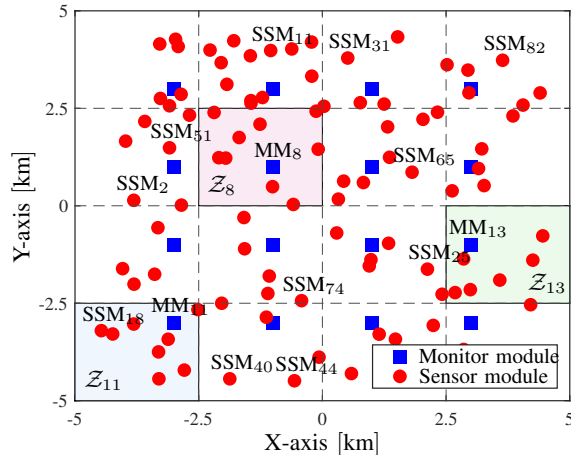
Fig. 5: The distribution of SSMs and MMs in the assumed area.

A. Setup and Assumptions

We consider a DMS with $K = 100$ randomly distributed SSMs and $M = 16$ fixedly positioned MMs in an area of $10 \times 10\,[\text{km}^2]$ divided into sixteen identical subareas (see Figure 5). The k-th SSM transmits update packets to its four closest MMs, i.e., $|\mathcal{M}(k)| = 4$, $\forall k$, over Rayleigh block fading channels. The m-th MM observes packets from its serving SSMs with the same probabilities, i.e., $p_k^{(m)} = \frac{1}{|\mathcal{K}(m)|}$, $\forall m, k \in \mathcal{I}_{l_k^{(m)}}$. The information source follows a $\text{Zipf}(n^{(k)}, s)$ distribution with pmf

$$P_X(x_i^{(k)}) = \frac{1/i^s}{\sum_{l=1}^{n^{(k)}} 1/l^s} \tag{31}$$

for $n_k = |\mathcal{X}^{(k)}|$ realizations at the k-th sensor. The parameter s is an exponent characterizing $P_X(\cdot)$, varying from uniform distribution ($s = 0$) to *peaky* ones.

The intrinsic and extrinsic features considered in our simulation scenario are shown in Table I for a fictitious *air pollution* monitoring platform that collects status packets from sixteen zones ($\mathcal{Z}_1$ to $\mathcal{Z}_{16}$), with different levels of importance. Without loss of generality, we consider $\mathcal{Z}_8$, $\mathcal{Z}_{11}$, and $\mathcal{Z}_{13}$ the most crucial zones, and the average Euclidean distance of a sensor from the centers of these three zones indicates its spatial importance. Furthermore, we assume that rare occurrences are monitored, where the importance of an update packet increases as its carrying pollution rate approaches 100%. From Table I, we first find $\lambda_g^{(k,m)} = 0.37$, and then, the corresponding $U_n^{(k,m)}$ for each $w_b^{(k,m)}$ is calculated. The parameters used in the simulations are summarized in Table II. The values of $\gamma_{M,j}^{(k,m)}$ and $g_j^{(k,m)}$ depend on the codeword length of the j-th arrival according to

TABLE I: Intrinsic and Extrinsic features with their proposed properties.

Feature	$w_b^{(k,m)}$	$U_n^{(k,m)}$	Limits	Critical points
Probability	–	–	$(0, 1]$	$\{0\}$
Usefulness	–	–	$[1, n_k]$	Ten random $x_i^{(k)}$
Loss risk*	–	–	$[0, 100]\%$	$\{100\}\%$
Average distance	0.4	0.4	$[3.66, 8.68)$ [km]	$\{3.66\}$ [km]
Resolution	0.2	0.6	$(0, R_{\max}]$	$\{R_{\max}\}$
Tolerance	0.2	1	$[0, 100]\%$	$\{100\}\%$
Circuit power	0.2	0.63	$[P_{\min}, \infty)$	$\{P_{\min}\}$
Battery state	0.3	1	$[0, 100]\%$	$\{0\}\%$
*The criticality for all critical points is equal to 1.				

TABLE II: Parameters for simulation results.

Parameter	Symbol	Value	Parameter	Symbol	Value
Number of the SSMs	K	100	Minimum attenuation factors	$\rho_{\min}$	0.1
Number of the MMs	M	16	Maximum attenuation factors	$\rho_{\max}$	5
Size of realizations set	$n_k, \forall k$	100	Maximum packet drop times	$d_{\max}$	10
Constant exponent for Zipf$(.,.)$	s	0.4	Maximum transmission rounds	$r_{\max}$	3 [35]
Update packet input rate	$\lambda_k, \forall k$	0.5	Reverse of channel coding rate	$c_j^{(k,m)}, \forall k, b, j$	2 [35]
Number of intrinsic features	A	3	Constant weight parameter	$w_m, \forall m$	1
Number of extrinsic features	B	3	Upper bound of codeword lengths	$\ell_{\max}$	100
Exponent for intrinsic features	$\alpha_a^{(k,m)}, \forall a, k, m$	0.5	Monitoring time length of arrivals	T	1000 [sec]
Exponent for extrinsic features	$\bar{\alpha}_b^{(k,m)}, \forall b, k, m$	1	Minimum circuit power	$P_{\min}$	0.1 [W]
Minimum importance	$\mathrm{VF}_{\min}$	0.1	Maximum sensing resolution	$R_{\max}$	5 [m]
Bias value for the utility forms	$\beta_k^{(m)}, \forall k, m$	5	Average received SNR	$\bar{\gamma}_k^{(m)}, \forall k, b$	12 [dB]
Exponent for the RUT case	κ	2	Packet error rate's approx. factors for the ARQ (HARQ) protocol	$\gamma_{M,j}^{(k,m)}, \forall k, b, j$	17.19 (10.1) [dB]
				$g_j^{(k,m)}, \forall k, b, j$	0.1 (0.96)

(15) and (16). We consider a binary phase shift keying (BPSK) modulation scheme at the PHY module. Initializing the codeword length ℓ_j from zero to $\ell_{\max}$ as the worst case, the value of $\gamma_{M,j}^{(k,m)}$ varies from 5.53 to 5.63 with almost constant $g_j^{(k,m)} \approx 0.1$ under the ARQ protocol. We consider the average value of $\gamma_{M,j}^{(k,m)}$ for all codeword lengths, as it results in negligible difference. Applying (2)–(5), and Tables I and II, meta-values are computed. Unless otherwise specified, the EUT case, fixed-length filtering, ARQ protocol, and Algorithm 1 are used to obtain the results.
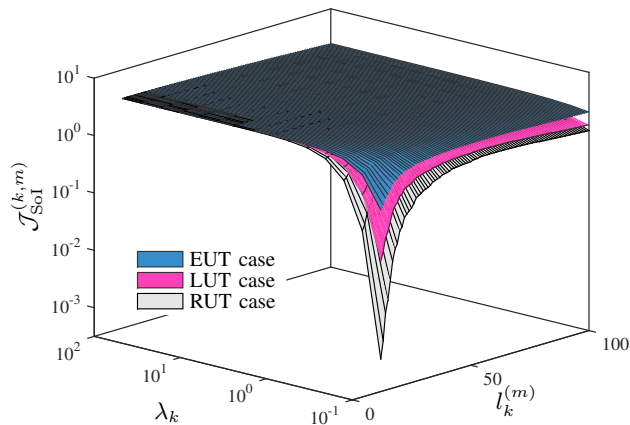
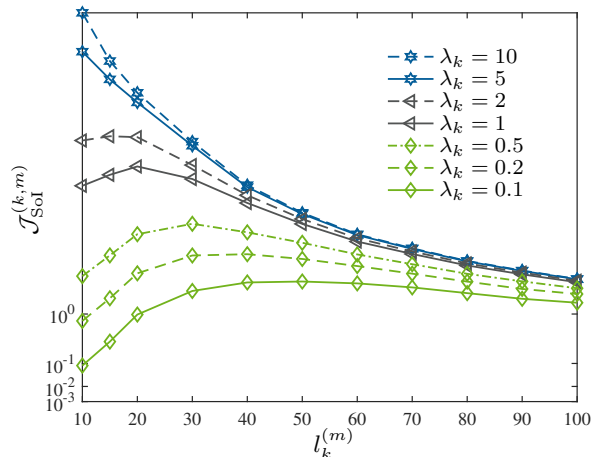Fig. 6: The interplay between the SoI, the arrival rate, and the admission size for the EUT, LUT, and RUT cases.

Fig. 7: SoI versus the size of the admitted realizations for the EUT case and different arrival rates.

B. Results and Discussion

Figure 6 shows the objective function $\mathcal{J}_{\mathrm{SoI}}^{(k,m)}$ versus the input rate λ_k and the number of selected realizations $l_k^{(m)}$ at the k-th sensor connected to the m-th MM. We observe that there is an optimal size of selected realizations for any input rate, which maximizes the SoI over the network. Otherwise stated, every random size of the selected realization set requires a specific input rate to maximize the SoI at the link level. However, the SoI saturates for large sizes of realization sets and high arrival rates. The convergence of the SoI to its maximum value with an optimal value of $l_k^{(m)} \leq n_k$ implies that the most valuable status packets can be efficiently transferred with a lower channel load. Considering the three defined forms of the utility function in (8), the EUT and RUT offer the highest and lowest transferred SoI, respectively. Likewise, one can plot the SoI objective function versus the arrival rate and the size of the realization set for a system employing adaptive-length filtering, reaching to similar conclusions as in Figure 6. Optimal values of $l_k^{(m)}$ based on different input rates are shown in Table III for fixed- and adaptive-length filtering under all three utility forms. Increasing the input rate results in smaller selection size to compensate a high blockage rate. Figure 7, a two-dimensional depiction of Figure 6 for the EUT case, shows the objective function $\mathcal{J}_{\mathrm{SoI}}^{(k,m)}$ versus the selection size $l_k^{(m)}$ for different values of $\lambda_k^{(m)}$. Expectedly, higher arrival rates lead to smaller sets of selected realizations, as fewer packets are selected to avoid a high blockage rate, whereas for low arrival rates, a large number of packets can pass the filter to harness high lateness.
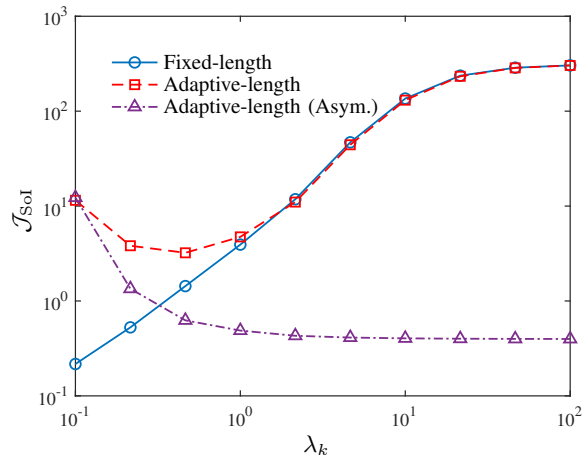
Fig. 8: SoI performance of semantics-aware filtering versus the arrival rate.

Initializing the arrival rate from 10^{-1} to 10^2 and finding optimal values for $l_k^{(m)}$, we assess the performance of the proposed semantics-aware filtering techniques. For that, Figure 8 depicts the sum SoI objective function $\mathcal{J}_{\mathrm{SoI}} \coloneqq \sum_{m=1}^M \sum_{k \in \mathcal{K}(m)} \mathcal{J}_{\mathrm{SoI}}^{(k,m)}$ versus input rate $\lambda_1 = ... = \lambda_K = \lambda_k$ for the fixed-length, adaptive-length, and an asymptotic variant of adaptive-length with $l_k^{(k)} = n_k$, $\forall k$. We see that the asymptotic adaptive-length method outperforms the fixed-length one for low arrival rates. However, the performance of the asymptotic adaptive-length filtering decreases with increasing the arrival rate. The performance of the fixed-length filtering gradually improves by increasing the input rate since the size of admitted packets shrinks to lower the blockage rate. Thanks to its flexibility, adaptive-length filtering outperforms its asymptotic variant and fixed-length filtering for any arrival rate.

The normalized SoI of the transmitted and monitored update packets are plotted in Figure 9, simulated over an interval of $[0, 1000]$ [sec.] for the rates of $\lambda_k = 0.5$ and 5, which validates the analytical results used for plotting Figure 8.

Using the simulations in Figure 9, Figure 10 depicts the average percentages (rates) of filtered, blocked, and transmitted packets to the m-th MM, out of all packets arrived at the k-th SSM within the interval of $[0, 1000]$ [sec.], averaged over 500 iterations. The adaptive-length method has clearly higher filtering rates, i.e., 70.47% (51.13%), and provides lower blockage rates, i.e., 19.93% (46.45%), compared to those of the fixed-length filtering, i.e., the filtering rates of 65.68% (45.66%) and the blockage rates of 24.78% (52.17%), for $\lambda_k = 0.5$ (5), thanks to its flexibility and semantics-aware decisions. The advantages of adaptive-length filtering come at
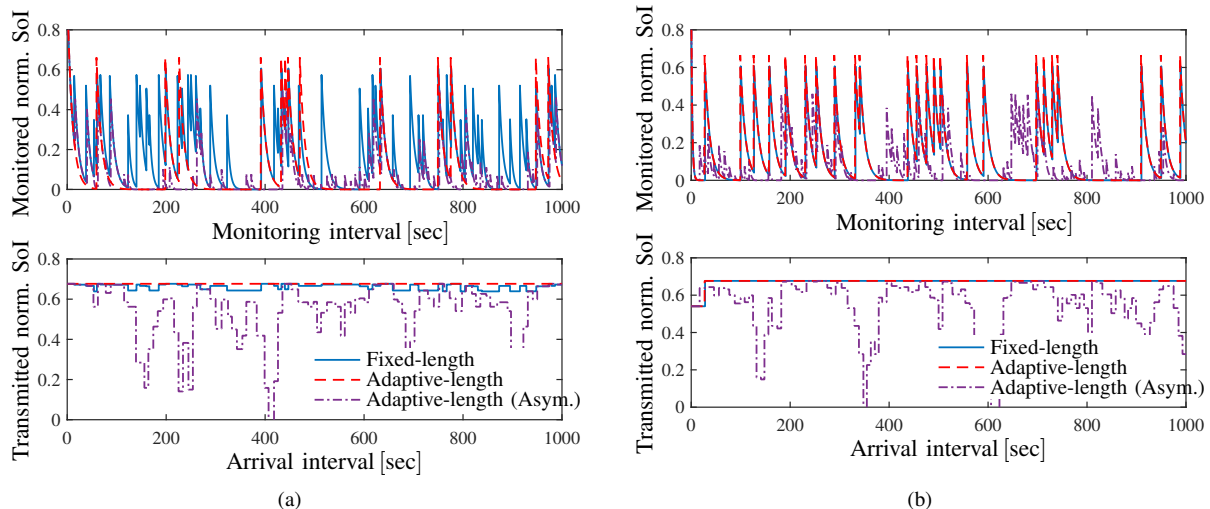
Fig. 9: Comparison of the proposed semantics-aware filtering from the transmitted and monitored SoI perspectives for (a) $\lambda_k = 0.5$ and (b) $\lambda_k = 5$.

Fig. 11: SoI versus the arrival rate and the admission size for Algorithm 1 and the MOSEK solver.
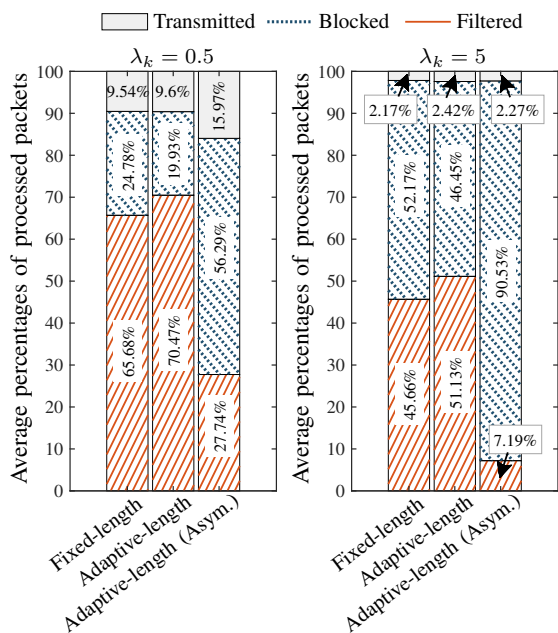
Fig. 10: Average percentage of filtered, blocked, and transmitted packets arrived within the interval of $[0, 1000]$ [sec.] under different filtering techniques.
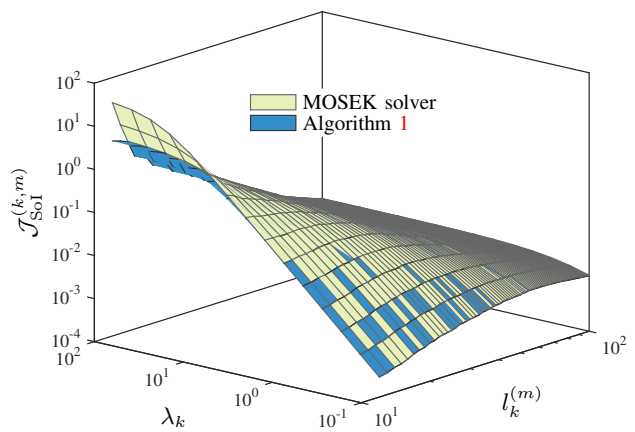
the cost of requiring slightly higher channel capacity, i.e., 0.06% and 0.25% for $\lambda_k = 0.5$ and 5, respectively, compared to a fixed-length filter. It is worth mentioning that the drops of the transmitted SoI for the asymptotic adaptive-length filtering in Figure 9 occur due to elevated blockage rate and number of consecutive blockages. The blockage and transmission load rates

TABLE III: Optimal number of admitted packets ($n_k = 100$), together with blockage and transmission load rates.

	λ_k	Fixed-length			Adaptive-length		
		EUT	LUT	RUT	EUT	LUT	RUT
$l_k^{(m)}$	0.1	43	43	47	80	80	87
	0.5	23	23	28	38	38	68
	1	10	8	12	31	28	37
	5	3	1	1	3	1	1
	10	2	1	1	2	1	1
Blockage rate	0.1	3.47	3.93	4.99	2.6	2.81	3.29
	0.5	24.78	29.34	39.9	19.93	27.1	31.28
	1	38.31	43.04	47.25	35.68	39.2	42.71
	5	52.17	56.78	64.64	46.45	52.78	60.61
	10	73.8	79.23	83.61	69.52	70.27	79.45
Load rate	0.1	31.44	33.09	35.42	33.21	34.37	36.34
	0.5	9.54	10.57	11.89	9.6	10.98	12.16
	1	7.03	7.05	7.06	7.22	7.32	7.37
	5	2.17	2.44	3.01	2.42	2.9	3.15
	10	1.55	1.87	1.78	1.82	1.84	1.84
All values indicate percentages (%).							

of the fixed- and adaptive-length filtering are given in Table III by extending the simulations in Figure 10 for the three proposed utility forms of timeliness and their associated optimal $l_k^{(m)}$.

Figure 11 depicts the SoI versus λ_k and $l_k^{(m)}$ for two scenarios, in order to evaluate the accuracy of Algorithm 1. In the first scenario, the MOSEK solver is used to find the optimal codeword lengths; in the second scenario, Algorithm 1 is applied to determine the optimal codeword lengths from (21), (25), or (27). We observe that the proposed algorithm and the solver perform identically for low to moderate arrival rates, whereas for higher rates, the MOSEK solver slightly outperforms Algorithm 1, resulting in a difference in SoI as low as 0.32% on average. Nevertheless, the algorithm has the advantage of tractability, providing design insights and information about the effect of different system parameters on the codeword lengths.

In Figure 12 (a) and Figure 12 (b), we compare the SoI when ARQ and HARQ protocols are used respectively under two different truncation models. The first model has a fixed (predefined) maximum number of transmission rounds, i.e., $r_{\max,j}^{(k,m)} = r_{\max}$, while in the second model, the importance of the arrival affects the maximum possible number of its transmissions according to (13). It is shown that both models provide almost the same SoI for small admission sizes.
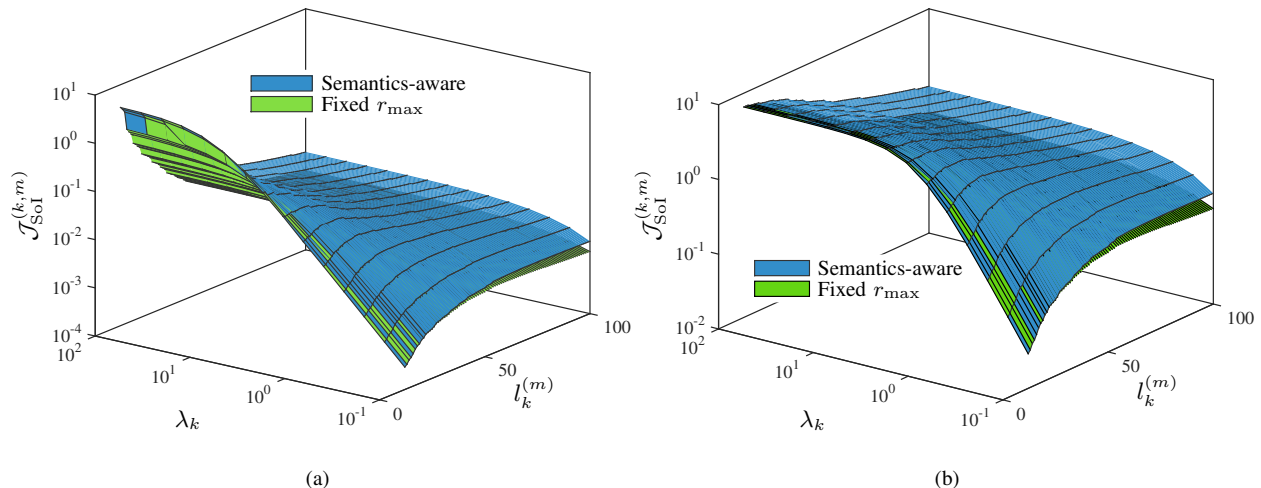
Fig. 12: SoI performance of (a) ARQ and (b) HARQ error control protocols truncated with fixed and semantics-aware transmission limits.

However, the semantics-aware truncated protocol exhibits higher performance for moderate to large admission sizes. Indeed, when the arrival traffic increases, reserving the channel for more important packets results in higher gain compared to giving an equal chance to all blocked packets to be retransmitted for a fixed number of rounds, regardless of their usefulness.

VII. CONCLUSION

We studied the problem of timely source coding in a distributed monitoring network with heterogeneous goals, where observations at each information source are filtered and transmitted to monitoring entities, depending on their importance (semantics) for achieving each monitor's goal. We analytically derived the optimal real codeword lengths that maximize a weighted sum of semantics-aware utility functions. Specifically, our analytical and numerical results show an optimal number of realizations to send for each link, which depends, among others, on the source statistics, the packet arrival rate, the semantic value of each arrival, and the weight assigned to each monitor's goal. We proposed two different semantic filtering methods and showed that adaptive-length filtering may provide higher SoI with a lower channel blockage rate at the cost of a minor increase in the transmission load. We also extended the conventional truncated error control protocols to a semantics-aware one that specifies the maximum number of retransmissions based on the importance of a packet. The main takeaway is that properly optimized semantic

filtering and source coding can significantly reduce the number of status updates required for the timely delivery of semantic information to monitors with heterogeneous goals.

VIII. ACKNOWLEDGEMENTS

APPENDIX A
VALUE FUNCTION DERIVATION: AN EXAMPLE

We provide an example on how to calculate the value function. For that, we study an air quality measurement system in which (a) the *usefulness* of each packet and (b) the *location* of its source are considered the intrinsic and extrinsic features of interest, respectively. For modeling $VF(z)$ of feature (a), large z shows a higher degree of contamination, hence higher usefulness of packets at all intervals. In this regard, there is globally one critical point $z_{n_1} = z_{max}$ with an arbitrary criticality. Thus, as the value of a realization is close to z_{max}, any packet from that realization gets high importance. In this case, from (5), we can write

$$\widehat{VF}(z) = e^{-\frac{(z - z_{max})^2}{2\sigma^2}} \tag{32}$$

where σ is iteratively found such that $VF(z) \geq VF_{min}, \forall z \in [z_{min}, z_{max}]$. According to (4) and (32), we plot Figure 13.

Besides, for feature (b), we consider $z_{n_1} = \frac{1}{4}(z_{max} + z_{min})$, $z_{n_2} = z_{max} + z_{min}$, and $z_{n_3} = \frac{3}{4}(z_{max} + z_{min})$ as critical spatial zones with the same criticality, i.e., $\varpi_{n_1} = \varpi_{n_2} = \varpi_{n_3} = 1$, from a served monitor's perspective. In this regard, we have

$$\widehat{VF}(z) = e^{-\frac{(4z - z_{max} - z_{min})^2}{32\sigma^2}}$$
$$+ e^{-\frac{(z - z_{max} - z_{min})^2}{2\sigma^2}}$$
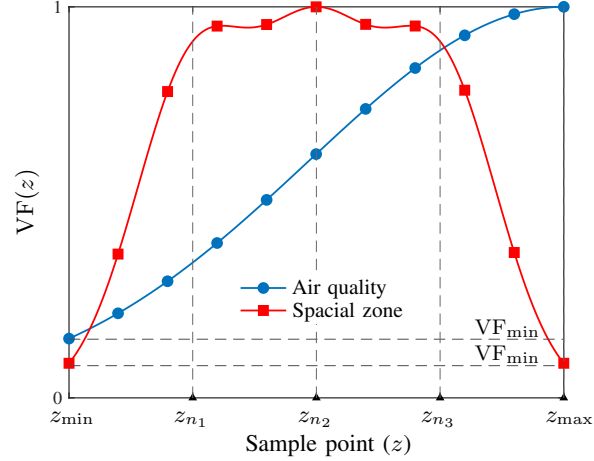$$+ e^{-\frac{(4z - 3z_{max} - 3z_{min}))^2}{32\sigma^2}}. \tag{33}$$

Fig. 13: The value function versus a sample point under two forms related to the assumed intrinsic and extrinsic features in the example.

Therefore, any packets from the sensors located in or at the adjacency of these zones are highly important (see Figure 13). It goes without saying that the sensors located around z_{n_2} take the highest $\text{VF}(z)$ since they are not only in a critical zone, but also under the effect of the two other critical zones.

APPENDIX B

PROOF OF LEMMA 1

At a layer of the k-th SSM serving the m-th MM, the timeliness utility of the $(j+d)$-th packet crosses the j-th one if

$$\rho_j^{(k,m)}\left(t_c^{j,d}-t_j^{(k,m)}\right) = \rho_{j+d}^{(k,m)}\left(t_c^{j,d}-t_{j+d}^{(k,m)}\right), \tag{34}$$

as in Figure 3, for every type of the defined forms in (8), where $t_c^{j,d}$ denotes the cross time of order d for the j-th arrival. From (34), we can write

$$\rho_j^{(k,m)}\left(t_c^{j,d}-t_j^{(k,m)}\right) = \rho_{j+d}^{(k,m)}\left(t_c^{j,d}-t_j^{(k,m)} - \sum_{k=j+1}^{j+d} Y_k^{(k,m)}\right) \tag{35}$$

which results in

$$t_c^{j,d} = t_j^{(k,m)} + \frac{\rho_{j+d}^{(k,m)}}{\rho_{j+d}^{(k,m)}-\rho_j^{(k,m)}} \sum_{k=j+1}^{j+d} Y_k^{(k,m)}. \tag{36}$$

Since $f_k(\cdot)$ is convex and non-increasing, if the $(j+d)$-th arrivals' utility plot crosses that of the j-th arrival, the j-th packet provides higher or equal SoI compared to the $(j+d)$-th packet, beyond that cross point. Therefore, the semantic filter discards the $(j+d)$-th status packet in the presence of the j-th one if $t_c^{j,d}$ is before the end of $S_{j+d}^{(k,m)}$. From (36), we reach

$$t_j^{(k,m)} + \frac{\rho_{j+d}^{(k,m)}}{\rho_{j+d}^{(k,m)} - \rho_j^{(k,m)}} \sum_{k=j+1}^{j+d} Y_k^{(k,m)} < t_j^{(k,m)} + \sum_{k=j+1}^{j+d} Y_k^{(k,m)} + S_{j+d}^{(k,m)}, \tag{37}$$

then we obtain

$$\frac{\rho_j^{(k,m)}}{\rho_{j+d}^{(k,m)} - \rho_j^{(k,m)}} < \frac{S_{j+d}^{(k,m)}}{\sum_{k=j+1}^{j+d}(S_k^{(k,m)} + W_{0,k}^{(k,m)})} \leq \frac{\ell_{\max}}{d\ell_{\max} + \sum_{k=j+1}^{j+d} W_{0,k}^{(k,m)}} \overset{(*)}{=} \frac{\ell_{\max}}{d\ell_{\max} + \widehat{W}_d^{(k,m)}} \tag{38}$$

where $\ell_{\max}$ denotes the upper bound for the size of a codeword length, and $W_{0,k}^{(k,m)}$ is the waiting time of the k-th arrival prior to filtering, which follows an exponential distribution with rate $\lambda_k q_{l_k^{(m)}}$. For $(*)$, we define $\widehat{W}_d^{(k,m)} := \sum_{k=j+1}^{j+d} W_{0,k}^{(k,m)}$. Since d is integer, $\widehat{W}_d^{(k,m)}$ becomes an Erlang distributed r.v. of order d and rate $\lambda_k q_{l_k^{(m)}}$. With regard to the definition of $\psi_k^{(m)}$ in (10), and after some manipulations, we obtain (11).

APPENDIX C
PROOF OF LEMMA 2

With regard to the form of $f_k(\Delta_k^{(m)}(t))$ for the EUT case in (8) and Figure 4, we compute polygons $Q_j^{(k,m)}$, $\forall j \geq 2$, as

$$Q_j^{(k,m)} = \int_{t_{j-1}^{(k,m)}}^{t_{j-1}^{(k,m)} + S_j^{(k,m)}} \left(e^{-\rho_{j-1}^{(k,m)}(t - t_{j-2}^{(k,m)})} + \beta_k^{(m)} \right) dt + \int_{t_{j-1}^{(k,m)} + S_j^{(k,m)}}^{t_j^{(k,m)}} \left(e^{-\rho_j^{(k,m)}(t - t_{j-1}^{(k,m)})} + \beta_k^{(m)} \right) dt$$

$$\approx -\rho_{j-1}^{(k,m)} Y_{j-1}^{(k,m)} S_j^{(k,m)} + \frac{1}{2}(\rho_j^{(k,m)} - \rho_{j-1}^{(k,m)})(S_j^{(k,m)})^2 - \frac{1}{2}\rho_j^{(k,m)}(Y_j^{(k,m)})^2 + (1 + \beta_k^{(m)})Y_j^{(k,m)} \tag{39}$$

where the approximation comes from second-order Taylor expansion for the exponential function. The expected value is then given by

$$\mathbb{E}[Q^{(k,m)}] \approx -\mathbb{E}[\rho^{(k,m)} Y^{(k,m)}] \mathbb{E}[S^{(k,m)}] + \frac{1}{2}\mathbb{E}[\rho^{(k,m)}(S^{(k,m)})^2]$$

$$- \frac{1}{2}\mathbb{E}[\rho^{(k,m)}]\mathbb{E}[(S^{(k,m)})^2] - \frac{1}{2}\mathbb{E}[\rho^{(k,m)}(Y^{(k,m)})^2] + (1 + \beta_k^{(m)})\mathbb{E}[Y^{(k,m)}]. \tag{40}$$

Note that large enough $\beta_k^{(m)}$ guarantees the positiveness of $\mathbb{E}[Q^{(k,m)}]$ in (40) after applying the expansion. By importing $Y_j^{(k,m)} = W_j^{(k,m)} + S_j^{(k,m)}$ and $\mathbb{E}[(S^{(k,m)})^m]$ from (12) into (40), and since $\rho_j^{(k,m)}$ and $W_j^{(k,m)}$ are independent, we obtain (20).

APPENDIX D

PROOF OF PROPOSITION 1

Using the parameter $\chi_k^{(m)}$ given in (23), we rewrite (20) as follows.

$$\mathbb{E}[Q^{(k,m)}] \approx -\chi_k^{(m)}\bar{\rho}_k^{(m)}\big(\mathbb{E}\big[\varphi^{(k,m)}L^{(k,m)}\big]\big)^2 - \frac{1}{2}\bar{\rho}_k^{(m)}\mathbb{E}\big[\varphi^{(k,m)}(L^{(k,m)})^2\big] - \chi_k^{(m)}\bar{\rho}_k^{(m)}\gamma_k^{(m)}\mathbb{E}\big[\varphi^{(k,m)}L^{(k,m)}\big]$$
$$+ \Big(1+\beta_k^{(m)} - \bar{\rho}_k^{(m)}\gamma_k^{(m)}\Big)\mathbb{E}\big[\varphi^{(k,m)}L^{(k,m)}\big] - \bar{\rho}_k^{(m)}(\gamma_k^{(m)})^2 + (1+\beta_k^{(m)})\gamma_k^{(m)}. \tag{41}$$

By (41), we define the Lagrange function $\mathcal{L}(\ell_i^{(k,m)}; \mu_k^{(m)})$ for $\mathcal{P}_2$ as

$$\mathcal{L}(\ell_i^{(k,m)}; \mu_k^{(m)}) = -\sum_{m=1}^{M} w_m \sum_{k\in\mathcal{K}(m)} p_k^{(m)}\eta_k^{(m)} \times \Bigg[\chi_k^{(m)}\bar{\rho}_k^{(m)}\Big(\sum_{i\in\mathcal{I}_{l_k^{(m)}}} p_i^{(k,m)}\varphi_i^{(k,m)}\ell_i^{(k,m)}\Big)^2$$

$$- \frac{1}{2}\bar{\rho}_k^{(m)}\sum_{i\in\mathcal{I}_{l_k^{(m)}}} p_i^{(k,m)}\varphi_i^{(k,m)}(\ell_i^{(k,m)})^2 - \chi_k^{(m)}\bar{\rho}_k^{(m)}\gamma_k^{(m)}\sum_{i\in\mathcal{I}_{l_k^{(m)}}} p_i^{(k,m)}\varphi_i^{(k,m)}\ell_i^{(k,m)}$$

$$+ \Big(1+\beta_k^{(m)} - \bar{\rho}_k^{(m)}\gamma_k^{(m)}\Big)\sum_{i\in\mathcal{I}_{l_k^{(m)}}} p_i^{(k,m)}\varphi_i^{(k,m)}\ell_i^{(k,m)}$$

$$- \bar{\rho}_k^{(m)}(\gamma_k^{(m)})^2 + (1+\beta_k^{(m)})\gamma_k^{(m)}\Bigg] - \mu_k^{(m)}\Big(\sum_{i\in\mathcal{I}_{l_k^{(m)}}} 2^{-\ell_i^{(k,m)}} - 1\Big) \tag{42}$$

where $\mu_k^{(m)} \geq 0$ is the Lagrange multiplier. Next, we can write Karush-Kuhn-Tucker (KKT) conditions for $i \in \mathcal{I}_{l_k^{(m)}}$, as what follows.

$$\frac{\partial\mathcal{L}(\ell_i^{(k,m)}; \mu_k^{(m)})}{\partial\ell_i^{(k,m)}} = -2\chi_k^{(m)}\bar{\rho}_k^{(m)}p_i^{(k,m)}\varphi_i^{(k,m)}\sum_{i\in\mathcal{I}_{l_k^{(m)}}} p_i^{(k,m)}\varphi_i^{(k,m)}\ell_i^{(k,m)} - \bar{\rho}_k^{(m)}p_i^{(k,m)}\varphi_i^{(k,m)}\ell_i^{(k,m)}$$

$$- \chi_k^{(m)}\bar{\rho}_k^{(m)}\gamma_k^{(m)}p_i^{(k,m)}\varphi_i^{(k,m)} + \Big(1+\beta_k^{(m)} - \bar{\rho}_k^{(m)}\gamma_k^{(m)}\Big)p_i^{(k,m)}\varphi_i^{(k,m)} + \mu_k^{(m)}\ln(2)2^{-\ell_i^{(k,m)}} = 0 \tag{43}$$

after combining $w_m p_k^{(m)}\eta_k^{(m)}$ with $\mu_k^{(m)}$, where $\mu_k^{(m)} \geq 0$ is the Lagrange multiplier. The complementary slackness condition is

$$\mu_k^{(m)}\Big(\sum_{i\in\mathcal{I}_{l_k^{(m)}}} 2^{-\ell_i^{(k,m)}} - 1\Big) = 0. \tag{44}$$

There exist two conditions, one of which meets (44), as below.

(i) $\mu_k^{(m)} = 0$, hence $\sum_{i \in \mathcal{I}_{l_k^{(m)}}} 2^{-\ell_i^{(k,m)}} < 1$; or

(ii) $\mu_k^{(m)} \neq 0$, hence $\sum_{i \in \mathcal{I}_{l_k^{(m)}}} 2^{-\ell_i^{(k,m)}} = 1$.

Under condition (i), the left side of (44) may potentially lead to non-positive codeword lengths, hence $\mathbb{E}\big[\varphi^{(k,m)} L^{(k,m)}\big] \leq 0$, for a given $\beta_k^{(m)}$. Since negative codeword lengths are not meaningful, by contradiction, condition (ii) must satisfy (44) where parameter $\mu_k^{(m)}$ offers a degree of flexibility. After some algebraic manipulations, we reach

$$\frac{\mu_k^{(m)} (\ln(2))^2}{\bar{\rho}_k^{(m)} p_i^{(k,m)} \varphi_i^{(k,m)}} 2^{-\ell_i^{(k,m)}} \exp\left(\frac{\mu_k^{(m)} (\ln(2))^2}{\bar{\rho}_k^{(m)} p_i^{(k,m)} \varphi_i^{(k,m)}} 2^{-\ell_i^{(k,m)}}\right) = \frac{\mu_k^{(m)} (\ln(2))^2}{\bar{\rho}_k^{(m)} p_i^{(k,m)} \varphi_i^{(k,m)}} 2^{\xi_k^{(m)}} \quad (45)$$

where $\xi_k^{(m)}$ is given in (22). Herein, the form of (45) is equal to $x \exp(x) = y$ for which the solution is $x = W_m(y)$, where $m = 0$ for $y \geq 0$.

APPENDIX E

PROOF OF LEMMA 3

Considering the form of $f_k(\Delta_k^{(m)}(t))$ for the LUT case in (8), polygons $Q_j^{(k,m)}$, $\forall j \geq 2$, and using Taylor expansion, yields

$$Q_j^{(k,m)} = \int_{t_{j-1}^{(k,m)}}^{t_{j-1}^{(k,m)}+S_j^{(k,m)}} \left(\ln\big(-\rho_{j-1}^{(k,m)}(t - t_{j-2}^{(k,m)})\big) + \beta_k^{(m)}\right) dt$$

$$+ \int_{t_{j-1}^{(k,m)}+S_j^{(k,m)}}^{t_j^{(k,m)}} \left(\ln\big(-\rho_j^{(k,m)}(t - t_{j-1}^{(k,m)})\big) + \beta_k^{(m)}\right) dt$$

$$\approx -2\rho_{j-1}^{(k,m)} Y_{j-1}^{(k,m)} S_j^{(k,m)} + (\rho_j^{(k,m)} - \rho_{j-1}^{(k,m)})(S_j^{(k,m)})^2 - \rho_j^{(k,m)}(Y_j^{(k,m)})^2 + (\beta_k^{(m)} - 1)Y_j^{(k,m)}, \quad (46)$$

whose expectation is given by

$$\mathbb{E}[Q^{(k,m)}] \approx -2\mathbb{E}\big[\rho^{(k,m)} Y^{(k,m)}\big] \mathbb{E}[S^{(k,m)}] + \mathbb{E}\big[\rho^{(k,m)}(S^{(k,m)})^2\big]$$

$$- \mathbb{E}[\rho^{(k,m)}]\mathbb{E}\big[(S^{(k,m)})^2\big] - \mathbb{E}\big[\rho^{(k,m)}(Y^{(k,m)})^2\big] + (\beta_k^{(m)} - 1)\mathbb{E}[Y^{(k,m)}]. \quad (47)$$

Using (47), and after some mathematical manipulations, we obtain (24).

REFERENCES

[1] P. Agheli, N. Pappas, and M. Kountouris, "Semantics-aware source coding in status update systems," in *Proc. IEEE Intern. Conf. on Commun. Workshops (ICC Workshops)*, 2022.

[2] ——, "Semantic source coding for two users with heterogeneous goals," in *Proc. IEEE Global Comm. Conf. (GLOBECOM)*, 2022.

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

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

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

[6] L. Floridi, "Semantic Conceptions of Information," in *The Stanford Encyclopedia of Philosophy*, Winter 2005.

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

[8] J. Bao *et al.*, "Towards a theory of semantic communication," in *IEEE Network Science Workshop*, 2011.

[9] F. M. Willems and T. Kalker, "Semantic coding: Partial transmission," in *Proc. IEEE Intern. Symp. on Inf. Theory (ISIT)*, 2008, pp. 1617–1621.

[10] 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.

[11] O. Goldreich, B. Juba, and M. Sudan, "A theory of goal-oriented communication," *J. ACM*, vol. 59, no. 2, 2012.

[12] 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.

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

[14] Z. Qin, X. Tao, J. Lu, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv preprint arXiv: 2201.01389v2*, 2022.

[15] 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.

[16] 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.

[17] N. Pappas, M. A. Abd-Elmagid, B. Zhou, W. Saad, and H. S. Dhillon, *Age of Information: Foundations and Applications*. Cambridge University Press, 2023.

[18] C. Bisdikian, L. M. Kaplan, and M. B. Srivastava, "On the quality and value of information in sensor networks," *ACM Trans. Sen. Netw.*, vol. 9, no. 4, July 2013.

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

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

[21] 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 Commun.*, vol. 68, no. 8, pp. 4905–4918, 2020.

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

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

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

[25] M. Bastopcu, B. Buyukates, and S. Ulukus, "Optimal selective encoding for timely updates," in *Proc. Annual Conf. on Inform. Sciences and Systems (CISS)*, 2020, pp. 1–6.

[26] B. Buyukates, M. Bastopcu, and S. Ulukus, "Optimal selective encoding for timely updates with empty symbol," in *Proc. IEEE Intern. Symp. on Inf. Theory (ISIT)*, 2020, pp. 1794–1799.

[27] M. Bastopcu, B. Buyukates, and S. Ulukus, "Selective encoding policies for maximizing information freshness," *IEEE Trans. on Commun.*, vol. 69, no. 9, pp. 5714–5726, 2021.

[28] M. B. Baer, "Source coding for quasiarithmetic penalties," *IEEE Trans. on Inf. Theory*, vol. 52, no. 10, pp. 4380–4393, 2006.

[29] L. H. Warren, "Strategic information synthesis by globular knowledge fusion," in *Proc. of IEEE Information, Decision and Control (IDC)*, 1999, pp. 407–412.

[30] G. Choquet, "Theory of capacities," in *Annales de l'institut Fourier*, vol. 5, 1954, pp. 131–295.

[31] M. Sugeno, "Theory of fuzzy integrals and its applications," *Doct. Thesis, Tokyo Institute of Technology*, 1974.

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

[33] P. Parag, A. Taghavi, and J.-F. Chamberland, "On real-time status updates over symbol erasure channels," in *Proc. IEEE Wir. Comm. and Net. Conf (WCNC)*, 2017, pp. 1–6.

[34] X. Lagrange, "Throughput of HARQ protocols on a block fading channel," *IEEE Comm. Letters*, vol. 14, no. 3, pp. 257–259, 2010.

[35] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. on Wir. Comm.*, vol. 3, no. 5, pp. 1746–1755, 2004.

[36] T. M. Cover, *Elements of Information Theory*. J. Wiley & Sons, 1999.