

A Spatio-Temporal Model for Information Freshness in Massive Random Access

Andrea Munari, *Senior Member, IEEE*, Alessandro Buratto, *Student Member, IEEE*, Federico Chiariotti, *Senior Member, IEEE*, Leonardo Badia, *Senior Member, IEEE*, and Petar Popovski, *Fellow, IEEE*

Abstract

Massive connectivity, a key building block of 5G, is expected to play an important role in the next generation of wireless systems, and its expected requirements are being revolutionized through the modeling of the information dynamics related to the vast numbers of Internet of things (IoT) devices. Motivated by this, the present paper introduces a model that captures the spatio-temporal nature of freshness of information sent via random access channel policies from an extremely large set of IoT devices via simple scalar parameters, i.e., the probability of success and accuracy of received updates. There are many information freshness metrics, starting from the age of information (AoI), all of which are proxies for the actual application performance, characterized over the temporal dimension. Our model adds the spatial dimension to this picture, observing that sensors distributed over the same area may have a strong correlation, and information from multiple close-by sensors may improve the overall accuracy of the receiver. We focus on characterizing the uncertainty of the receiver, expressed through the conditional entropy, considering a network of partially reliable, spatially distributed sensors observing the same process and reporting their measurements over a slotted ALOHA channel. We consider a simple forgetful receiver and a more complete model which accounts for the full history of past observations, deriving their performance, and optimizing the transmission probability of nodes to minimize the expected uncertainty.

Index Terms

A. Munari is with the Institute of Communications and Navigation, German Aerospace Center (DLR), Weßling, Germany. email: andrea.munari@dlr.de. A. Buratto, F. Chiariotti, and L. Badia are with the Dept. of Information Engineering, University of Padova, Italy. email: {alessandro.buratto.1@studenti.unipd.it, federico.chiariotti@unipd.it, leonardo.badia@unipd.it}. P. Popovski is with the Dept. of Electronic Systems, Aalborg University, Denmark. email: petarp@es.aau.dk.

This work was supported in part by the Italian National Recovery and Resilience Plan (NRRP), as part of the RESTART partnership (PE0000001), under the European Union NextGenerationEU Project, by the the German Federal Ministry of Research, Technology, and Space (BMFTR) with the xG-RIC project as part of the research program Communication Systems “Souverän. Digital. Vernetzt.” (grant number 16KIS2429K), and by the Villum Investigator Grant “WATER” from the Velux Foundations, Denmark.

A preliminary version of this work [1] was presented at IEEE SPAWC, Surrey (UK), 7–10 July, 2025.

Information freshness, wireless sensor networks, random access, remote source monitoring, Internet of Things.

I. INTRODUCTION

The rapid development of the Internet of Things (IoT) over the past decade has led to the deployment of tens of billions of distributed sensors, enabling a wide range of applications, from predictive maintenance in manufacturing plants to environmental monitoring [2], [3]. However, managing a massive number of low-power sensors, with limited communication capabilities and severe computational and energy constraints due to their small batteries, poses several challenges captured by novel key performance indicators and constraints [4], which have shaped and oriented recent research on medium access design.

In 5G, massive machine-type communication has already been recognized as a cornerstone service category, with random access protocols playing a central role in enabling scalable connectivity under sporadic traffic patterns. Yet, the requirements of 6G push this problem to a new level. Beyond sheer connectivity, 6G IoT scenarios are tightly linked to the concept of goal-oriented communications [5], where communication protocols shall be designed and optimized to convey information that serves a concrete task, e.g., monitoring, and where information freshness and accuracy become as important as throughput.

In this context, age of information (AoI) has emerged as a key metric of timeliness [6]: unlike traditional network latency, which focuses on a single packet, AoI measures the freshness of the most recent update available to the receiver, thus acting as a much closer proxy for the actual misalignment between the real condition of the measured process and the best available estimate. Since its introduction in 2011, AoI has been an active research field, and a significant amount of scientific literature has been dedicated to its characterization [7] and its use in network optimization [8]. Several extensions of AoI have also been developed, providing even closer proxies of the application performance, at the cost of higher complexity and application-specific definitions. The age of incorrect information (AoII) [9] and value of information (VoI) [10] are particularly interesting, as they consider the estimation error directly in their calculation, penalizing scenarios in which the estimate is misaligned with the actual state of the system.

However, while the temporal dimension of information freshness has been thoroughly explored and characterized, the spatially distributed nature of IoT networks has been mostly neglected so far, or only considered as part of the medium access model [11]. In real IoT networks, the physical distance of a sensor from the target event to be measured is one

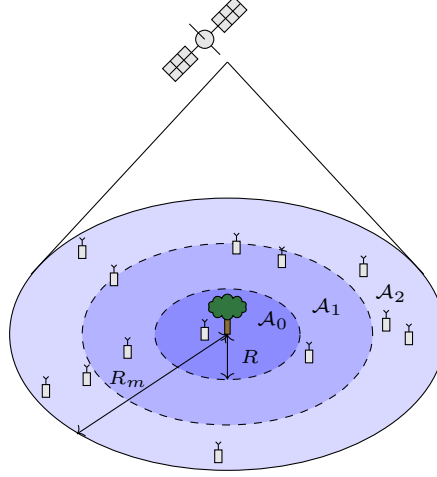


Figure 1. A schematic of the considered example scenario.

of the factors determining the quality of individual measurements, and fresher, lower-quality measurements may be less informative than older, better measurements. Consequently, metrics such as AoI become less useful as proxies for the actual knowledge accuracy at the receiver.

Corrections to the definitions of AoI or VoI that account for this phenomenon have been proposed [12]–[14], but these are still indirect measurements of the knowledge uncertainty. We argue that the entropy of the *a posteriori* distribution of the state, conditioned on the information available to the receiver [15], [16], provides a direct characterization of the performance of the monitoring system, without the need for such proxies, and is thus the ideal target for massive access optimization. This work is inspired by the use of entropy as an estimation and control theoretical metric [17], which has shown significant advantages over traditional metrics such as minimum mean square error under non-Gaussian errors [18]. Additionally, the receiver entropy can capture more than strictly spatial and temporal relations, as it is tightly coupled with system dynamics and structural factors.

A. Motivating Example

Consider the scenario represented in Fig. 1, in which a phenomenon of interest (represented by the tree) is observed by a set of spatially distributed sensors that can perform an uplink transmission towards an IoT gateway. This could be instantiated by a terrestrial base station, or by a non-terrestrial receiver, e.g., a low Earth orbit (LEO) satellite in the example. Closer sensors will be able to observe the process with a high accuracy, while farther ones will still obtain measurements, albeit with a reduced quality. Sensors beyond a certain distance will not observe the process at all. The IoT application, fed with the data collected by the gateway, monitors the system through sensor updates.

If observations can be polled at will, picking the closest sensor is a natural choice, so as to maximize reliability. However, such an approach is hardly viable in practical IoT systems, as (1) it would quickly deplete the energy of such node, and (2) sensors might be able to obtain and report new measurements only sporadically due to duty-cycle and normative constraints. On top of this, data collection is likely to be performed relying on random access protocols over a shared channel, as scheduled schemes often have signaling requirements that might prove too costly for low-power sensors. In addition, data collection can be hindered by wireless channel impairments, which can prevent communication. Updates from sensors placed farther away then become beneficial: although they may have a higher error rate, they still provide new information to the gateway, which can be integrated to the current belief over the state of the process.

This setting presents two major challenges. The first is to determine the accuracy of the knowledge at the receiver's end and its uncertainty over the state of the system. Metrics derived from the AoI do not fully capture the scenario, or need significant adjustments to account for the sensors' different levels of accuracy. The second is medium access optimization, as the difference in performance between a simple scheme in which sensors within a given radius transmit with a set rate and a more complex one in which sensors closer to the event of interest transmit more often is an open question. It is possible that preventing farther nodes from transmitting, reducing the frequency of received updates, i.e., the system throughput, but increasing their quality, may have a beneficial effect on the receiver uncertainty. Furthermore, if sensors have some knowledge of their position, it might be useful to tag messages with the distance, but this incurs additional overhead. A complete characterization of the system also allows for optimization, improved performance, and reduction of unnecessary signaling. We argue that understanding the trade-offs that emerge in these settings is paramount to derive solid design principles of IoT massive connectivity in 6G [3], and we aim to provide useful insights in this paper. *As we show in the following, this does not impose a complexity burden on the system, but requires a careful optimization of key parameters such as the coverage radius. Notably, simple configurations in which all nodes transmit with the same probability are proven to minimize the uncertainty.*

B. Related Work

As the focus of the research community has moved from characterizing freshness to a more direct consideration of application performance in remote monitoring, several studies have analyzed the effect of different transmission policies in this scenario. In particular, AoII has

been the subject of a significant amount of work, as it combines the freshness and timeliness properties of AoI with the consideration of the mismatch between the physical process and the beliefs of the receiver. In a single transmitter scenario, policies that can optimize AoII under rate limitations and communication delays have been designed for slotted [19] and continuous time [20]. In both cases, the optimal transmission policies are based on state-dependent thresholds, which depend on the transition probabilities of the Markov chain that describes the monitored process. A recent work [21] has also considered pull-based strategies, i.e., systems in which a receiver might poll one of several sensors, finding that a characterization of the belief over the state, along with the system AoI, represent a sufficient statistic to obtain the AoII-optimal policy.

Another possibility that has been explored in the past few years has been the direct consideration of the estimator error at the receiver. Naturally, this requires the specification of the estimator and loss function, as different errors might have different impacts on performance. The specification of a cost for false alarms and missed detections was considered for an anomaly reporting system in [22], and the effect of different policies in the same scenario was analyzed in [23] using a hold estimator that maintains the last received value as its best estimate of the current state. A more advanced class of estimators was proposed in [24]: as the complexity of a maximum a posteriori (MAP) approach makes analysis difficult, and martingale-based solutions that exploit the Markov property only represent a limited set of systems, structured estimators that can still prove tractable while representing a wider class of systems are useful in the characterization of this problem. Finally, [25] considered a joint design of communication and estimation policies by modeling the problem as a two-agent system, showing that a joint design can significantly outperform fixed estimators and pull-based schemes.

All the aforementioned works consider systems in which sensors have a perfect observation of the process, and the only bottleneck is communication. Imperfect transmitters have been considered in [26], which proposed a pull-based reinforcement learning-based scheduler for a receiver that implements a Kalman filter and decides which sensor to poll, and in [27], which defined a sampling cost for the individual sensor in a push-based system, effectively reducing the observability of the process.

In this work, we are interested in a special case of imperfect observation, in which the spatial distribution of the sensor network is crucial to determine the accuracy of information. While older works only considered the spatial dimension's effect on medium access [11], a recent paper [12] proposed a spatio-temporal definition of AoI, which models spatial corre-

lation as a proportional decrease of the AoI of a sensor if another close-by sensor transmits. A similar definition has been given in [13], considering the sum of AoI contributions from multiple sensors, weighted by their proximity to the sensor to be scheduled. Another recent work [14] applied a distance-based penalty in determining the freshness for a given point in space, choosing the freshest sample, which might not be from the sensor that is physically closest, depending on its relative age.

However, adapting AoI presents some drawbacks, since its value as a proxy for the actual monitoring performance requires a precise modeling of the sensor network and the effect of distance on the accuracy of the observation. The entropy of full-sequence [16] and forgetful [28] receivers was considered in our previous work under different transmission policies. Unlike the present contribution, these papers consider a case in which observations are perfectly reliable and each node independently monitors a different process. While the analytical tools are similar to the ones used in this paper, [16], [28] relied on the renewal property of successful transmissions, which does not hold under the present settings, requiring a more complex characterization. The entropy rate at the receiver was tackled in [15], which considered the tracking of a binary Markov process through a binary channel: this is a complementary scenario with respect to [16], [28], as it does not include medium access aspects, but focuses on a single node transmitting at every slot over an unreliable channel that may yield incorrect observations. The present work combines both the effect of multiple nodes attempting to access the channel and of their observations being unreliable following a spatial model, requiring a more complex analysis. The receiver entropy may prove more meaningful than proxy metrics [29], as it directly represents the performance of the monitoring application [30]. Particularly relevant to the present paper is the uncertainty of information (UoI), studied in [31], [32] for scheduled monitoring of Markov processes, and defined as the entropy at the monitor on the current state of a source conditioned on the last received sample. The metric is akin to what considered here in one of the cases we target, i.e., when the receiver only has access to the last received value and to the time elapsed since its retrieval. The novelty of our work is to consider imperfect observations, whose accuracy depends on spatial aspects, i.e., the distance from the event of interest, and the interactions between accuracy and success rate over a collision channel. To properly characterize these, we consider the case in which the whole history of observations is available at the receiver up to the present time, and explore how this can significantly improve performance.

C. Contribution

The goal of this work is to move beyond purely temporal notions of freshness, as captured by metrics such as AoI or VoI, by explicitly incorporating a spatial dimension. These two aspects jointly influence the uncertainty at the receiver, which we quantify through the conditional entropy of the system state, given the history of received messages.

This setting is particularly relevant in IoT data collection scenarios, where a gateway asynchronously gathers updates from spatially distributed sensors. We consider a model in which sensors monitor a Markov process and access the channel via slotted ALOHA, providing a tractable yet representative framework for more complex systems. We leverage a general distance-based error model for sensor observations, and formulate an analytically tractable framework that preserves the spatio-temporal characteristics of the scenario, which can be characterized through a hidden Markov model (HMM). The resulting model has a single-sensor equivalent, whose activity rate and accuracy reflect the spatial distribution of nodes and their activation probabilities. We thus optimize the medium access parameters, analyzing the effect of the coverage radius and characterizing the optimal access probability for different regions as well as the effect of including geographical knowledge into transmitted packets. Furthermore, we analyze a *forgetful* receiver that only keeps the last received packet and a *full-sequence based (FS)* receiver that considers the full history of received data up to the present moment.¹

To the best of our knowledge, our work is the first to fully characterize this type of scenario in terms of receiver uncertainty, providing a useful tool for system optimization. In particular, we show that restricting the coverage range to a smaller value may reduce said uncertainty at the expense of system throughput. For most of our work, we concentrate on binary Markov processes. Aside from being relevant in and of itself for applications such as anomaly detection, the binary case allows us to analyze and understand trends and obtain insights that can carry over to more complex models. We then extend our results to a general Markov model with $M > 2$ states. Our main contributions can then be summarized as follows:

- We observe that medium access is memoryless across slots, which can be captured by two parameters, the success probability and accuracy of updates. The problem could also be modeled as a single sensor with those aggregate properties. We characterize

¹These two receivers are often colloquially referred to as *stateless* and *stateful*, respectively. We avoid this terminology in the present work, as the belief of the forgetful receiver also evolves over time following the Markov chain that represents the process, and the use of “stateless” would thus be formally incorrect.

the spatio-temporal observation process and provide the parameters corresponding to different medium access settings;

- We *analyze* the uncertainty of a forgetful and full-sequence based receiver for a binary Markov process by computing the conditional entropy of the source state, given the observations available at the receiver. The forgetful approach can be of practical relevance in view of its simplicity and allows us to obtain an exact, closed-form expression of the metrics of interest. For the full-sequence based case, the uncertainty is captured by formalizing an HMM that describes the observations collected by the receiver;
- For both solutions, we study in detail the effect of the asymmetry of the binary Markov process and prove the existence of an optimal coverage range to minimize the average uncertainty at the receiver. Notably, the solution differs from the system configuration that would be attained by minimizing AoI, which also maximizes the system throughput, clarifying the critical role of considering the spatial dimension in information freshness;
- The initial model is extended to capture the case in which devices are aware of their location, and make this information available whenever sending an update. The receiver can then leverage such geographical knowledge to weigh the reliability of the incoming readings. Moreover, we also determine how the transmission probabilities should be adapted based on the nodes' distance from the sensed process, in order to minimize the receiver uncertainty. Our study provides non-trivial insights, revealing that both cases lead to negligible performance gains, hinting at the design of simpler IoT systems;
- We finally extend our main results to a Markov process with M states, showing that the main considerations we draw are not due to the simplistic nature of the binary process, but represent more general properties of spatio-temporal monitoring systems.

A preliminary version of this work was presented in [1], which contained only the characterization of the forgetful receiver. The rest of this paper is organized as follows: first, Sec. II presents the system model and metrics. The analytical results for the forgetful and full-sequence receiver are then derived in Sec. III and Sec. IV, respectively. Sec. V presents the optimization of the communication systems, and Sec. VI describes the results when extending the monitoring to an M -state process. Finally, Sec. VII concludes the paper.

II. SYSTEM MODEL AND PRELIMINARIES

A. Notation

We denote a discrete random variable (r.v.) and its realization by upper- and lowercase letters, respectively, e.g., X and x . The probability mass function of the r.v. X is indicated as

$P[X = x] = p(x)$, and the conditional distribution of X given Y is $p_{X|Y}(x|y)$. Subscripts are omitted when no ambiguity arises. The Shannon entropy of a r.v. X is denoted by $H(X) = -\sum_x p(x) \log_2 p(x)$, with the usual extensions to conditional entropy. When n denotes the current time index, X_n is the corresponding value of a random process, whereas we indicate by $X^n = [X_0, X_1, \dots, X_n]$ the vector of all the values of the process up to n . Finally, we denote vectors in lowercase boldface, e.g., $\mathbf{a}$. In particular, a vector denoting the position of a point in $\mathbb{R}^2$ is indicated as $\mathbf{x}$.

B. System model

Consider a physical process of interest, modeled as a finite-state, time-homogeneous discrete time Markov chain X_n , $n \in \mathbb{N}$, taking values in $\mathcal{X}$. We denote the one-step transition probability matrix for X_n as T . For the initial part of our study, we focus on the case of a two-state chain with $\mathcal{X} = \{0, 1\}$, so that

$$T = \begin{pmatrix} 1-q & q \\ \eta q & 1-\eta q \end{pmatrix} \quad (1)$$

with $q > 0$, leading to the stationary distribution of the ergodic chain $\pi_0 = \eta/(1+\eta)$, $\pi_1 = 1 - \pi_0$. In (1), $\eta \in (0, 1/q)$ is the *asymmetry factor*, denoting the ratio of the time spent in state 0 to the time spent in state 1. Without loss of generality, we will consider in the remainder $\eta \geq 1$. The case $\eta = 1$ identifies symmetric source, whereas, for $\eta > 1$, state 0 is more frequently visited. The case $\eta < 1$ means that state 1 is more frequent than 0, so we can simply swap the indices of the states. The general setup with $|\mathcal{X}| > 2$ will be tackled in Sec. VI, introducing the required modeling details.

We study a network topology composed of a receiver and a population of $m \in \mathbb{N}$ IoT devices, which are spread over a circular coverage area $\mathcal{A}$ of radius R_m , centered without loss of generality in the origin of the plane. Nodes are then randomly distributed over the coverage area according to the uniform distribution probability density function (PDF)

$$f(x, y) = \begin{cases} \pi^{-1} R_m^{-2}, & \text{if } (x, y) \in \mathcal{A}; \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

At any time slot, each device within coverage independently decides whether to take a reading of the process and transmit a message containing the value to the receiver.²

²Nodes outside the coverage area are inactive and do not attempt any communication with the receiver.

Due to the displacement of devices within $\mathcal{A}$, the reliability of the observations produced by a node depends on its position. In particular, we assume that devices closer to the origin of the plane provide more precise readings, whereas the accuracy of messages coming from farther away progressively reduces, as the longer the information about the event propagates physically, the higher the probability that it may be distorted. Such a setting captures, for instance, an application tracking the value of a physical quantity at the specific coordinates of interest (plane origin), or a monitoring system that has to detect an event (e.g., presence/absence of an object) in the corresponding surroundings.

We partition the coverage area $\mathcal{A}$ in K regions:

$$\mathcal{A}_i := \{\mathbf{x} \in \mathcal{A} : iR \leq \|\mathbf{x}\|_2 < (i+1)R\} \quad (3)$$

for $i \in \mathcal{K} = \{0, \dots, K-1\}$, and where $R := R_m/K$. Region $\mathcal{A}_i$ is thus an annulus with internal radius iR , external radius $(i+1)R$, and area $A_i = (2i+1)\pi R^2$. The overall area of the coverage region is $A = \pi R_m^2 = \pi K^2 R^2$. We can thus consider the number of nodes in area $\mathcal{A}_i$, denoted as m_i , as the outcome of an urn model [33]. Specifically, the portion of the m total sensors falling into that specific annulus follows a binomial distribution, i.e., $m_i \sim \text{Bin}(m; \frac{A_i}{A}) = \text{Bin}(m; \frac{2i+1}{K^2})$, as the probability of any given node falling into $\mathcal{A}_i$ is A_i/A , due to the uniform distribution across the coverage area following (2). The expected number of nodes in $\mathcal{A}_i$ is thus $\mathbb{E}[m_i] = \frac{(2i+1)m}{K^2}$.

Whenever a device in $\mathcal{A}_i$ performs a reading of the process X_n , it obtains a value Z_n according to the conditional distribution $p(z_n | x_n, i)$. In the case of the two-state process under study, the PMF simplifies to having a correct ($z_n = x_n$) or wrong ($z_n \neq x_n$) reading. For the time being, we assume that the reliability of Z_n does not depend on the current value of the source, and we model

$$\lambda_i := P[Z_n = X_n | i] = \max \left\{ \frac{1}{(1+iR)^\alpha}, \frac{1}{2} \right\} \quad (4)$$

with $\alpha > 0$, so that $P[Z_n \neq X_n | i] = 1 - \lambda_i$. The extension to the case in which the accuracy of a sensor is a continuous, monotonically decreasing function of the distance is relatively straightforward, replacing probability mass functions (PMFs) with PDFs and sums with integrals in the following. This power-law characterization has been shown to accurately capture spatial correlation among readings provided by sensor nodes for physical processes of interest [34], although the analysis presented can be easily adapted to different reliability functions.

The nodes send updates to the receiver over a shared wireless channel, following a slotted ALOHA policy [35]. Accordingly, at the beginning of a slot, each node independently decides

with probability $\zeta \in (0, 1)$ to sense the process and transmit a packet containing the obtained reading. *As the transmission decision does not depend on the state of the underlying process or on its observation, the transmission process is memoryless.* Following the on-off fading channel model [36]–[38], a packet is erased with probability $p_e < 1$, bringing no power contribution to the receiver, or arrives unfaded with probability $1 - p_e$. We further regard collisions to be destructive, i.e., a slot in which two or more packets arrive non-erased at the receiver prevents decoding of any of them, while singleton packets arriving unfaded over a slot are successfully received [35]. For the considered model, a slot sees the delivery of a message *as an i.i.d. Bernoulli random variable with success* probability

$$p_s = m \zeta (1 - p_e) \cdot [1 - \zeta(1 - p_e)]^{m-1} \quad (5)$$

corresponding to the average throughput of the system. At the receiver side, no knowledge can be gathered on the area out of which a received message was generated, e.g., due to location-unaware devices, or to the fact that no geographical information is piggybacked in transmitted packets. This assumption will be relaxed in Sec. V-A.

Remark 1. *The considered model captures setups in which packets coming from different nodes are equally likely to be retrieved. For instance, this is the case in settings where the relative displacement of nodes is not predominant on the overall path-loss. This includes ground nodes communicating with a satellite [39] or long-range IoT scenarios (e.g., LoRa [40]) in which the event of interest is not co-located with the base station. The model has been widely used in the literature [41]–[43] to abstract the specific type of physical layer being considered, aiming at results that can be applied to a wide class of systems, and has been proven to correctly identify the fundamental trade-offs that arise in the presence of random access contention.*

Interestingly, one could construct a single-sensor model whose performance is equivalent to the full system: this single-sensor equivalent summarizes the spatial aspect into the accuracy of individual observations, which results from the measurement distortion function described above, and the temporal aspect in the throughput p_s . Once the decentralized transmission strategy has been established, this equivalent model allows for an abstract performance analysis, where the parameters depend on spatial patterns and medium access design.

Remark 2. *The existence of a single-sensor equivalent depends only on the memoryless nature of the transmission process, and would also hold in the presence of distance-dependent reception or more complex channel models. Since the focus of the present work is on the*

spatial accuracy of the readings, the model we target allows us to factorize the effect of p_s and of the accuracy of the received measurements. In the general case, the joint distribution of accuracy and reception probability would need to be considered.

In the remainder of the analysis, we will often focus on a generic slot n over which the receiver decodes an incoming message (singleton). Denote this event as $\mathcal{R}_n$ and note that it occurs with probability p_s . For the sake of compactness, and to simplify notation, we will indicate the distribution of a r.v. A_n conditioned on $\mathcal{R}_n$ as $\bar{p}(a_n) = P[A_n = a_n | \mathcal{R}_n]$.

In this context, our aim is to characterize the uncertainty the receiver has about the current state of the source. The metric is relevant for a number of IoT applications, e.g., dealing with actuation and decision-making, and, as will be illustrated later, is an interesting proxy to capture both the temporal and spatial components of information freshness for the system model under study. In the remainder, we will tackle two possible implementations at the receiver, dubbed *forgetful* and *full-sequence based* (FS). In both cases, we will capture the uncertainty considering information available at the application layer and assume that only the content of successfully decoded messages is forwarded from the lower layers of the receiver. Moreover, we assume that the number of users in the coverage area, the channel access probability, and the source statistics, as well as the reliability law in (4), are known at the receiver.³

Remark 3. *The modeling approach presented is inspired by IoT applications, in which control decisions are taken at the application layer, and little to no cross-layer information is exchanged, e.g., presence of idle vs. collided slots. To leverage this additional knowledge, which could help reduce receiver uncertainty by defining transmission policies linked to the source evolution [16], [28], specific interfaces across layers have to be implemented. As we target practical system implementations, we do not tackle this aspect in the present work.*

C. Receiver entropy

Let Y_n denote the channel output that the receiver observes at time n . The r.v. has alphabet $\mathcal{X} \cup \{\ast\}$, as Y_n belongs to the Markov process state space $\mathcal{X}$ when the receiver decodes a packet containing a reading of the tracked process, whereas the special symbol $\ast$ is used to

³Assuming knowledge of m , p_e and ζ is reasonable in practical systems. The possibility of remote source monitoring without prior knowledge of source statistics has also been recently studied, e.g., in [44].

represent the output of a slot in which no packet is decoded. In this case, the uncertainty at time n given the current realizations of the involved processes is

$$h(y^n) := H(X_n | Y^n = y^n) \quad (6)$$

where we recall that Y^n is the sequence of all past observations up to and including slot n .

To capture average system performance, we consider the conditional entropy on the source state given the random vector Y^n , $H(X_n | Y^n)$. Specifically, we are interested in its limiting behavior as n grows, denoted as

$$\tilde{H} := \lim_{n \rightarrow \infty} H(X_n | Y^n). \quad (7)$$

The existence of the limit is proven in Appendix A.

Remark 4. *The considered metric characterizes the uncertainty on the true state of the process conditioned on the available observations. As such, it is not directly tied to any specific estimator, and shall not be confused with the uncertainty on the estimate $\hat{X}_n$ the receiver may extract (i.e., with $H(\hat{X}_n)$). On the other hand, $H(X_n | Y^n)$ is by definition driven by the conditional distribution $p(x_n | y^n)$, which can be used to implement a maximum a posteriori (MAP) estimator, minimizing the probability of error. The value of $h(y^n)$ can then be used to gauge the level of uncertainty at the receiver, e.g., when deciding whether to take an action as response to the source state based on the MAP estimate. It shall also be noted that the metric provides broader insights on the problem of remote monitoring and estimation, giving bounds on the average error probability P_e . Specifically, for the binary source X_n under study, we have*

$$H_b^{-1}(H(X_n | Y^n)) \stackrel{(a)}{\leq} P_e \stackrel{(b)}{\leq} \frac{1}{2} H(X_n | Y^n). \quad (8)$$

where $H_b(p) = -p \log p - (1-p) \log(1-p)$ is the binary entropy function. Within (8), (a) follows from Fano's inequality [45], characterizing a lower bound on P_e for any estimator. Even more relevantly, Kovalevsky's result in (b) bounds from above the average error probability of a MAP estimator [46], [47].⁴ The inequalities in (8) corroborate the use of the conditional entropy $H(X_n | Y^n)$ as metric for protocol design in remote monitoring applications, as its reduction directly improves the attainable performance in terms of error probability.

⁴The bound is valid for $P_e \leq 1/2$, which is always true for the considered binary X_n .

D. Age of Information (AoI)

The metric defined in (7) aims to capture how both the spatial and temporal freshness components of the information available to the receiver influence its uncertainty. In the remainder, it will also be useful to consider a measure of the sole temporal dimension. This is provided by the instantaneous AoI Δ_n , defined as the number of slots since the reception of the latest packet, and by its average $\bar{\Delta} := \mathbb{E}[\Delta_n]$. Recall that in the system model considered: (1) a message is generated at the time of transmission; and, (2) the slotted ALOHA channel access probability is the same for all nodes and independent of the collected reading. Thus, the average AoI can be characterized in terms of the success probability in (5) as [48]:

$$\bar{\Delta} = \frac{1}{2} + \frac{1}{p_s}. \quad (9)$$

III. ANALYTICAL CHARACTERIZATION OF THE RECEIVER UNCERTAINTY: FORGETFUL CASE

The first receiver type we consider monitors the source solely based on the last received message, i.e., it is oblivious of previously received information. Following this approach, the receiver in time slot n has knowledge about: (1), the time elapsed since the last reception; and (2), the content of the last decoded packet. Recalling that we assume that transmitting nodes sample the process at the beginning of the slot used to attempt delivery, (1) also corresponds to the current AoI at the receiver Δ_n . Moreover, we indicate the reading contained in the last decoded message as of time n as $W_n \in \mathcal{X}$. In other words, if a new reading is received over a slot, the quantity is reset to the obtained update, whereas it retains the value of the last retrieved message if no packet is decoded over the current time unit. The uncertainty of the *forgetful* receiver in slot n , given that a specific realization is observed, is then captured by the entropy

$$h_f(w_n, \delta_n) := H(X_n | W_n = w_n, \Delta_n = \delta_n). \quad (10)$$

The quantity h_f is relevant from an operational point of view, as it can serve as a basis for properly gauging current knowledge or making actuation decisions. Note indeed that, in the forgetful case, the posterior distribution $p(x_n | w_n, \delta_n)$ suffices to derive a MAP estimator, and the entropy metric can provide a measure of the uncertainty on the true source state to be used in combination with the available estimate. We also note that h_f reduces to the uncertainty of information introduced in [31], [32] in the case of perfectly reliable readings. In this setting, we are interested in the metric $\tilde{H}_f$, capturing the limiting behavior of the conditional entropy $H(X_n | W_n, \Delta_n)$. We characterize it with the following result.

Lemma 1. *For the forgetful receiver, we have*

$$\tilde{H}_f := \lim_{n \rightarrow \infty} H(X_n | W_n, \Delta_n) = \sum_{w_n, \delta_n} h_f(w_n, \delta_n) p(w_n, \delta_n) \quad (11)$$

where $p(w_n, \delta_n)$ is the joint stationary distribution of W_n and Δ_n .

Proof: See Appendix B. ■

While clearly suboptimal, this type of receiver is of interest for a number of reasons. Firstly, it allows for the derivation of compact closed-form expressions for the metrics of interest. Moreover, its discussion will be instrumental to introduce the fundamental trade-offs that characterize the system, such as the effect of tuning the coverage radius by preventing nodes farther than a certain distance from transmitting, which reduces the system throughput, but may increase the receiver's accuracy. Secondly, the solution epitomizes the behavior of practical IoT systems, in which low-complexity approaches are often preferred. In this perspective, the values of $h_f(w_n, \delta_n)$ could easily be pre-computed and stored in a look-up table, allowing the receiver to gauge the current level of the entropy in (10) without any computational effort. Finally, the forgetful receiver provides a meaningful benchmark when readings are reliable, as clarified by the following result.

Remark 5. *Following the data processing inequality, we have $h_f(w_n, \delta_n) \geq h(y^n)$, as (w_n, δ_n) can be deterministically derived from y^n . In this case, packet receptions represent renewals from the receiver's perspective, as all information about previous steps is forgotten. If sensor readings are always reliable, the forgetful receiver minimizes the entropy, in the sense that $h_f(w_n, \delta_n) = h(y^n)$, as the next steps of the process depend only on the last observed state thanks to the Markov property.*

Example 1. *Consider the tracking of a symmetric source ($\eta = 1$), with transition probability $q = 0.05$. Two examples of how $h_f(w_n, \delta_n) = H(X_n | W_n = w_n, \Delta_n = \delta_n)$ may evolve over time are reported in Fig. 2, assuming a reliability power-law coefficient $\alpha = 0.02$, $p_e = 0.1$, and setting a transmission probability $\zeta = 10^{-4}$. In the first case, Fig. 2a, a smaller radius of $R_m = 6R$ was considered, whereas the rightmost plot, Fig. 2b, was obtained for a wider coverage of $R_m = 15R$, with $R = 10$ m. The total number of nodes was scaled proportionally to the area, maintaining the same average number of nodes per unit area, which we denote as $\rho = \frac{m}{\pi R_m^2}$. In both examples, $\rho \simeq 0.05$, up to a negligible error due to the integer number of nodes. The example hints at a key trade-off. On the one hand, selecting a smaller radius leads to more sporadic update deliveries (corresponding to resets of $h_f(w_n, \delta_n)$), and the*

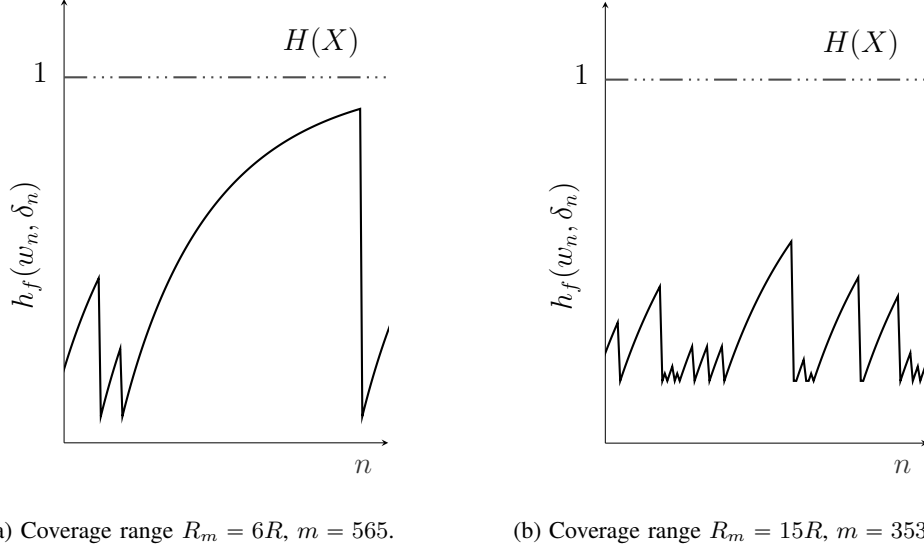


Figure 2. Examples of time evolution of $h_f(w_n, \delta_n)$ when tracking a symmetric source. The horizontal line reports the value of the stationary source entropy, $H(X) = -\pi_0 \log_2 \pi_0 - \pi_1 \log_2 \pi_1$. The plot was generated for $\alpha = 0.02$, $R = 10\text{m}$, $q = 0.005$, $\rho \simeq 0.05$, $\zeta = 10^{-4}$, $p_e = 0.1$.

uncertainty tends to grow for longer periods of time. Whenever the receiver does not retrieve messages for some time, such as in the central part of Fig. 2(a), its uncertainty tends to converge to the stationary entropy of the source, i.e., $H(X) = -\pi_0 \log_2 \pi_0 - \pi_1 \log_2 \pi_1 = 1$ [bit]. This effect is much more subdued in Fig. 2b, as the presence of more nodes in the system results in an improvement of the frequency of message delivery, with a lower AoI and an uncertainty that is more often brought back to its minimum. On the other hand, such reset value gets higher as R_m increases. In fact, incoming messages are more likely to be generated by devices farther away from the origin, containing less reliable readings as per (4), leaving the receiver with a higher residual uncertainty. The critical balance of these two factors will determine system performance, epitomizing the role of spatial and temporal information freshness.

A. Derivation of the average uncertainty, $\tilde{H}_f$

To characterize the receiver uncertainty, we focus first on a generic time n_0 over which a message is successfully decoded. Accordingly, the slot sees the reset of: (1), the AoI value to $\Delta_{n_0} = 0$, and (2), W_{n_0} to the value contained in the incoming reading. Applying Bayes' rule, the distribution of the state of the tracked process conditioned on the received reading can be expressed as

$$\bar{p}(x_{n_0} | w_{n_0}) = \frac{\bar{p}(w_{n_0} | x_{n_0}) \pi_{x_{n_0}}}{\sum_{x'_{n_0} \in \mathcal{X}} \bar{p}(w_{n_0} | x'_{n_0}) \pi_{x'_{n_0}}} \quad (12)$$

where we recall that the notation $\bar{p}(\cdot)$ was introduced in Sec. II-A to denote a distribution conditioned on having decoded a packet during the current slot. Within (12), π_{x_n} is the stationary distribution of X_n , introduced in Sec. II, and we have used the fact that the event of successfully receiving a packet is independent of X_n for the transmission policies under study. In turn, $\bar{p}(w_{n_0} | x_{n_0})$ can be conveniently expressed in terms of the r.v. $D_{n_0} \in \mathcal{K}$, which captures the region from which the message originated. Leaning on the law of total probability, and on the fact that transmissions are independent on the state of the source:

$$\bar{p}(w_{n_0} | x_{n_0}) = \sum_{d_{n_0}=0}^{K-1} \bar{p}(w_{n_0} | x_{n_0}, d_{n_0}) \bar{p}(d_{n_0}). \quad (13)$$

Following the urn model from Sec. II-B, we have

$$\bar{p}(d_{n_0}) = \frac{\mathbb{E}[m_i]}{m} = \frac{2d_{n_0} + 1}{K^2}. \quad (14)$$

In turn, the first factor within the summation in (13) captures the probability that the decoded message contains value w_{n_0} , conditioned on the actual state of the process and the reliability of the sender. Accordingly, we can write in a compact form

$$\bar{p}(w_{n_0} | x_{n_0}, d_{n_0}) = \lambda_{d_{n_0}} \mathbb{1}(x_{n_0}, w_{n_0}) + (1 - \lambda_{d_{n_0}}) \mathbb{1}(\bar{x}_{n_0}, w_{n_0}) \quad (15)$$

where $\mathbb{1}(a, b)$ is the indicator function, taking value 1 if $a = b$ and 0 otherwise.

For the power-law in (4), combining (14) and (15), we obtain

$$\bar{p}(w_{n_0} | x_{n_0}) = \sum_{d_{n_0}=0}^{K-1} \frac{2d_{n_0} + 1}{K^2(1 + d_{n_0} R)^\alpha}, \quad \text{if } w_{n_0} = x_{n_0} \quad (16)$$

and its complementary value for the case $w_{n_0} \neq x_{n_0}$. Plugging (16) into (12) thus provides the statistics available at the receiver on the state of the process of interest upon reception of a packet. Recalling that $\bar{p}(x_{n_0} | w_{n_0}) = p_{X_n | W_n, \Delta_n}(x_{n_0} | w_{n_0}, 0)$, the result allows to derive the general probability $p(x_n | w_n, \delta_n)$ at any slot $n = n_0 + \delta_n$ between n_0 and the successive reception. This is obtained as the δ_n -step evolution of the transition matrix T with initial distribution $\bar{p}_{X_n | W_n}(0 | w_{n_0})$ and $\bar{p}_{X_n | W_n}(1 | w_{n_0})$.

By its definition in (10), the uncertainty of the forgetful receiver on the current state of the process given the currently available knowledge can then be computed as

$$h_f(w_n, \delta_n) = - \sum_{x_n \in \mathcal{X}} p(x_n | w_n, \delta_n) \log_2 p(x_n | w_n, \delta_n). \quad (17)$$

Finally, we derive the conditional entropy $\tilde{H}_f$. From (11), this can be done computing the PMF $p(w_n, \delta_n)$, capturing, for a generic time slot n under stationary conditions, the joint probability of having decoded the last message δ_n slots ago and of its content being w_n .

Observing that packet receptions are i.i.d. across slots with probability p_s , the current AoI value is independent of the collected reading, as each node's transmission probability does not depend on its age or position. Consequently, $p(w_n, \delta_n) = p(w_n)p(\delta_n)$. Moreover, $p(w_n)$ can simply be computed from (13) as

$$p(w_n) = \sum_{x_n \in \mathcal{X}} \bar{p}(w_n | x_n) \pi_{x_n}. \quad (18)$$

In turn, the receiver experiences an AoI $\delta_n \geq 0$ at the start of slot n if the last success was followed by δ_n failures, so that:

$$p(\delta_n) = p_s(1 - p_s)^{\delta_n}. \quad (19)$$

Combining (18)-(19) with (17) provides an exact formulation of $\tilde{H}_f$.

Remark 6. *The approach followed in the analysis above considers an i.i.d. distribution for the random variable D_n across slots. This implicitly corresponds to independently re-drawing the positions of all nodes at each time unit, and thus introduces an approximation of the behavior of a practical topology in which the locations of the devices are static. The validity of this assumption will be discussed in Sec. III-B.*

B. Numerical results and key trends

To gauge the impact of spatio-temporal information freshness on the forgetful receiver, we focus on a population of devices spread over a variable coverage area $\mathcal{A}$, scaling the number of nodes by maintaining a fixed average number of nodes per unit area $\rho \simeq 5 \cdot 10^{-2}$ [nodes/m²], up to a negligible approximation error given by the constraint that m must be an integer number, transmitting with probability $\zeta = 10^{-4}$ at every slot.⁵ A sent packet is erased with probability $p_e = 0.1$, and the tracked Markov process X_n transitions from state 0 to 1 with probability $q = 5 \cdot 10^{-3}$. Unless otherwise specified, we assume that the nodes produce readings following the reliability law in (4), where $\alpha = 0.02$ and $R = 5$ m.

Some first fundamental insights emerge from Fig. 3, where we report the conditional entropy $\tilde{H}_f$ against the coverage radius R_m . The solid line reports the case of monitoring a symmetric process ($\eta = 1$), whereas the dashed curve was obtained for an asymmetric case with $\eta = 5$. Circular markers show the results of detailed network simulations obtained at a subset of R_m values, and aimed at verifying the approximation introduced in our analysis

⁵The chosen transmission probability is representative for IoT applications, corresponding to the behavior of a LoRa device sending packets of duration 300 ms roughly once per hour.

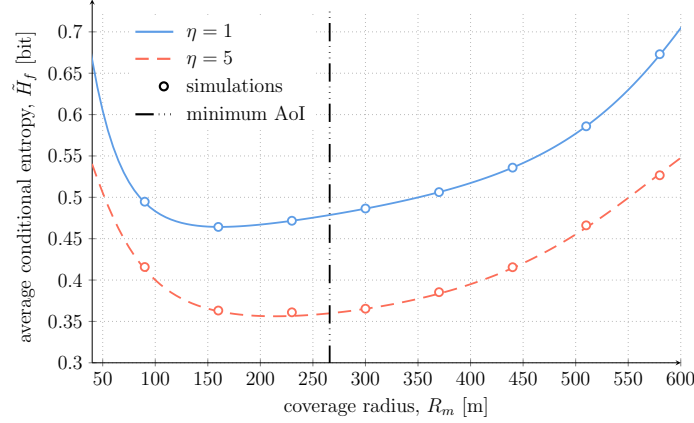


Figure 3. Conditional entropy $\tilde{H}_f$ vs. coverage radius R_m , when tracking a symmetric (dashed line) or an asymmetric (solid line) process, considering a forgetful receiver. The markers report the outcomes of simulation results, and the vertical line indicates the coverage radius that would minimize AoI.

(see Remark 6). Specifically, for each simulation, the positions of the m nodes were drawn uniformly within $\mathcal{A}$, and their location was kept fixed for the whole duration of the run ($5 \cdot 10^4$ slots). Throughout the run, the Markov process evolved according to the transition matrix T , and the time evolution of $h_f(w_n, \delta_n)$ was computed via (17). The value of $\tilde{H}_f$ was then obtained as the time average of h_f as per (10). Each of the reported marked points shows in turn the average of 50 independent topologies. The excellent match between analysis and simulations under all conditions confirms the tightness of the modeling assumptions.

More interestingly, Fig. 3 reveals the existence of an optimal coverage range to minimize $\tilde{H}_f$. When R_m is too small, the updates are sporadic as few nodes fall within $\mathcal{A}$. In this case, even if incoming messages are highly reliable in view of the proximity of active devices to the origin, the receiver's uncertainty is dominated by the lack of timely (fresh) updates. However, if the coverage radius is increased beyond a critical point R_m^* , the conditional entropy grows once more (rightmost region of the plot). In such conditions, even if messages are received more often, their reliability decreases sharply due to the higher probability of obtaining readings from nodes farther away from the origin, leading to a greater uncertainty. Striking a proper balance between the temporal and spatial components of information freshness is thus paramount and shall represent a key criterion in the design of the system.⁶ From this standpoint, it is relevant to note that the optimal radius R_m^* can be significantly smaller than

⁶The coverage radius could, for instance, be controlled in LEO IoT systems by tuning the receiver beam-width via beamforming, or in terrestrial systems by admitting at logon only users that are received with a high enough (average) power level.

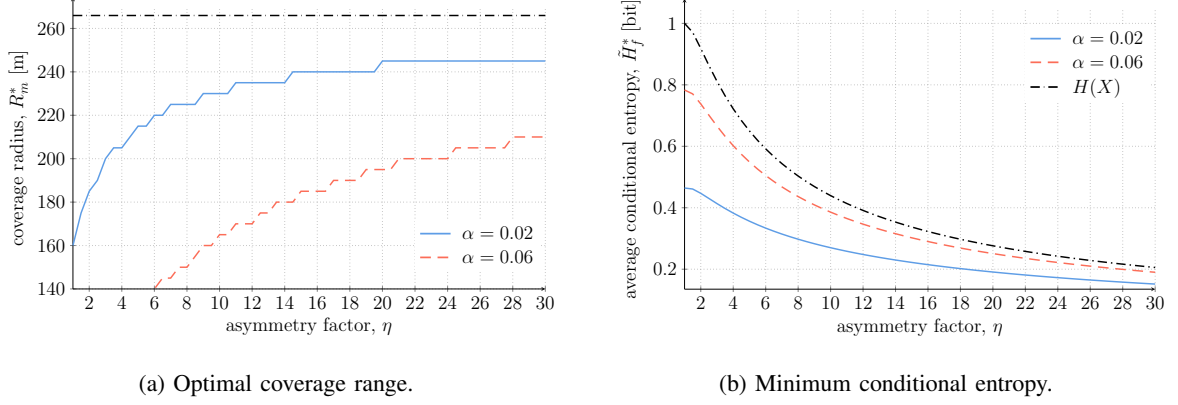


Figure 4. Optimal coverage range R_m^* and minimum average conditional entropy $\tilde{H}_f^*$ obtained with R_m^* vs asymmetry factor η for a forgetful receiver. The horizontal line corresponds to coverage for minimum AoI in (a) and to the source stationary entropy in (b).

the one that would be obtained considering AoI only. This is reported for reference by the vertical dash-dotted line in the figure. From (9), such a configuration is clearly obtained by maximizing the system throughput p_s [49], [50]. This is achieved when the number of users in coverage satisfies $m\zeta(1 - p_e) = 1$, that is, for an incoming channel load of 1 [pkt/slot], corresponding to an optimal radius $(\pi\rho\zeta(1 - p_e))^{-1/2}$.

Fig. 3 also highlights a difference in optimal coverage when tracking symmetric and asymmetric sources, with the latter case presenting a lower uncertainty and a larger R_m^* . This is due to the change in $p(x_n|w_n, \delta_n)$ in slots without an update: we know that $p(x_{n+1}|w_n, \delta_n + 1) = \sum_{x_n \in \mathcal{X}} T_{x_n, x_{n+1}} p(x_n|w_n, \delta_n)$. Consequently, highly asymmetric sources tend to quickly return to the most frequent state, thus reducing the entropy in the absence of updates. On the other hand, the uncertainty over the value of the state after several steps from the last update is very high if $\eta = 1$. This effect reduces the impact of the temporal component of information freshness, represented by the lower AoI, as the receiver can tolerate higher values of the age without a significant increase in the entropy. The relative importance of obtaining accurate updates, which depends on the location that successful packets come from and represents the spatial component of our problem, thus increases, decreasing the optimal value of R_m .

To further inspect this phenomenon, we consider the optimal coverage radius R_m^* , shown in Fig. 4a, and the corresponding minimum conditional entropy $\tilde{H}_f^*$, depicted in Fig. 4b, against the asymmetry factor. The optimal radius increases sharply as the asymmetry grows and, eventually, R_m^* tends to converge to the value minimizing AoI, i.e., $(\pi\rho\zeta(1 - p_e))^{-1/2}$,

since the temporal component of freshness dominates.⁷ A corresponding decrease in the attainable average uncertainty is experienced as η grows, obtained by admitting more nodes in the system. Fig. 4 reports the trends for two different values of the exponent α , determining the reliability of the messages incoming according to (4). It is interesting to note that for larger values of α the optimal coverage shrinks, as an effect of the faster degradation with distance of the quality of the updates. In this setting, trading off messages' timeliness for an increased reliability pays off. For instance, when $\eta = 5$ (the same setup considered in Fig. 3), the optimal coverage radius reduces to less than one half of what would be optimal in terms of AoI when $\alpha = 0.06$. Fig. 4b also reports the stationary entropy, i.e., the entropy of the steady-state distribution of the process. As η grows, the three values tend to converge, but the gains with $\alpha = 0.02$ are quite impressive, further showing the viability of a forgetful approach in systems with significant hardware constraints.

IV. FULL-SEQUENCE BASED RECEIVER

Consider now an FS receiver monitoring the state of the source relying on the whole history of past observations. In this case, the uncertainty is driven by the conditional distribution $p(x_n | y^n)$. To track this, observe that, for the transmission policy under study, $p(y^n | x^n, y^{n-1}) = p(y_n | x_n)$, so that (X_n, Y_n) is a HMM. The corresponding observation probabilities can be expressed as

$$p(y_n | x_n) = \begin{cases} p_s \bar{p}_{W_n | X_n}(y_n | x_n) & y_n \in \mathcal{X} \\ 1 - p_s & y_n = * \end{cases} \quad (20)$$

Specifically, when a message is decoded over the current slot, i.e., $y_n \in \mathcal{X}$, the r.v. W_n describing the last received value is reset to y_n , and the AoI is reset to 0. Thus, $p(y_n | x_n)$ is captured by $p_s \bar{p}_{W_n | X_n}$, where the latter PMF was obtained in (16). On the other hand, an observation $*$ is obtained whenever no message is retrieved due to either a collision or lack of transmissions, i.e., with overall probability $1 - p_s$.

Leaning on this, the joint distribution of the current source value and the overall vector of collected observations can be derived via the forward recursion of the HMM [51] as

$$p(x_n, y^n) = \sum_{x_{n-1} \in \mathcal{X}} p(x_{n-1}, y^{n-1}) p(x_n | x_{n-1}) p(y_n | x_n) \quad n > 1$$

⁷The stepwise behavior of the radius is due to the fact that only integer and even values were considering due to computational constraints.

with initial conditions at $n = 1$ set considering the stationary distribution π_x of the tracked source, i.e.,

$$p(x_1, y_1) = \sum_{x_1 \in \mathcal{X}} \pi_{x_1} p(y_1 | x_1).$$

The conditional distribution of the source is then obtained as

$$p(x_n | y^n) = \frac{p(x_n, y^n)}{\sum_{x_n \in \mathcal{X}} p(x_n, y^n)}$$

allowing us to derive the receiver uncertainty $h(y^n)$ via (6).

Before engaging in the discussion of the performance of the FS receiver, we report in the following two examples, which are instrumental to understanding the behavior of the two receivers for both symmetric and asymmetric sources.

Example 2. A first example of the evolution of $h(y^n) = H(X_n | Y^n = y^n)$ over time is reported in Fig. 5a, considering a symmetric source and using an FS receiver. The snapshot was obtained for a coverage range of $R_m = 50$ m, with $\alpha = 0.02$, $m = 392$ (corresponding to $\rho \simeq 0.05$), $\zeta = 10^{-4}$, $p_e = 0.1$. In the plot, the reception of a message containing reading 0 is highlighted with blue circle markers, whereas reception of a 1 is indicated with red square markers. In the early phase of the example, after collecting several 0 messages in a row, the receiver has a strong belief about the current state of the source, with $P[X_n = 0 | Y^n = y^n]$ reaching 0.948. The subsequent reception of a (possibly unreliable) reading with value 1 then leads to a rapid spike in the level of uncertainty, which then progressively reduces as further messages corroborate a change in the state of the source. The effect of taking all history into account is apparent even with a qualitative comparison with the simpler, more regular trend discussed in Example 1 for the forgetful receiver. On the one hand, uncertainty can suddenly rise when a message contradicting a strong belief arrives. On the other hand, the reset values of entropy upon decoding are not always the same as in the forgetful case, but rather strongly depend on the past collected messages.

Example 3. Figs. 5b and 5c present the time evolution of $h(y^n)$ for a FS receiver, when tracking an asymmetric source, with $\eta = 5$ and $\eta = 20$, respectively. The same parameters discussed in Example 2 were used and the receiver experiences the same sequence of incoming readings at the same time. Clearly, as η increases, the stationary source entropy $H(X)$, i.e., the uncertainty at which the receiver tends when not receiving updates for long time, decreases (horizontal dashed lines in the plots). Note that, while $h(y^n)$ for the tracked binary source can never exceed the stationary entropy $H(X) = 1$ in the symmetric case, the instantaneous

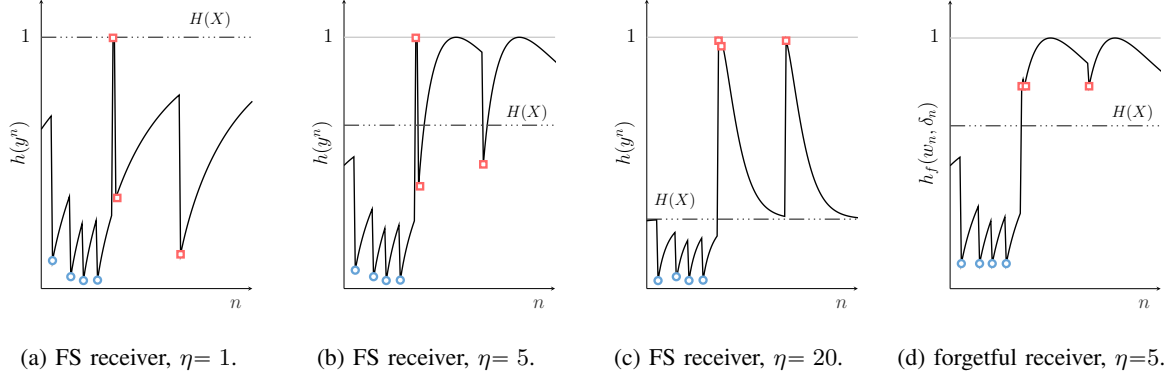


Figure 5. Instances of time evolution of $h(y^n) = H(X_n|Y^n = y^n)$ and $h_f(w_n, \delta_n) = H(X_n|W_n = w_n, \Delta_n = \delta_n)$, discussed in Examples 2 and 3. In all cases, the receiver observes the same set of channel outputs at the same time instants. Blue circle markers denote decoding of a message containing reading 0, whereas red square markers report decoding a message with reading 1. The dash-dotted line shows the source stationary entropy $H(X)$. Results obtained for $q = 0.005$, $m = 392$ ($\rho \simeq 0.05$), $R = 10m$, $q = 0.01$, $\alpha = 0.02$, $\zeta = 10^{-4}$, $p_e = 0.1$.

uncertainty of the receiver can exceed the value $H(X) < 1$ for larger values of η . By definition, the entropy is limited to 1 regardless of the value of η , as X_n is a binary variable. Let us first consider the case of a slight asymmetry in the source, as in Fig. 5b. Here, the reception of a single reading with value 1 leads to a high uncertainty, and while the second consecutive packet reduces the uncertainty, the overall entropy is higher than in the case with $\eta = 1$, since state 1 is generally less likely. We can observe that, after receiving packets reporting the less frequent value, the entropy gradually grows as the probability that the system is actually in that state decreases over time. When the probability flips, i.e., there is a higher chance that the system is in state 0, the entropy reaches 1 and then decreases, asymptotically converging from above to $H(X)$. On the other hand, when the reported value is 0, the entropy immediately after a packet is much lower, gradually increasing with time and converging to $H(X)$ from below. If the asymmetry grows significantly, as in Fig. 5c, which shows the trend for $\eta = 20$, even multiple consecutive readings confirming that the system is in the less frequent state lead to an entropy close to 1, as the low probability of being in state 1 leads the FS receiver to maintain a high uncertainty. The convergence towards $H(X)$ is then much faster from state 1 and correspondingly slower from state 0.

For completeness, Fig. 5d reports the uncertainty $h_f(w_n, \delta_n)$ of a forgetful receiver exposed to the same sequence of observations, considering a source with $\eta = 5$. As discussed in Example 1, the effect of ignoring previously received readings results in resetting $h_f(w_n, \delta_n)$ to the uncertainty of the incoming message. In the case of an asymmetric process X_n , this corresponds to a higher value when receiving the less likely state 1 or a lower value when

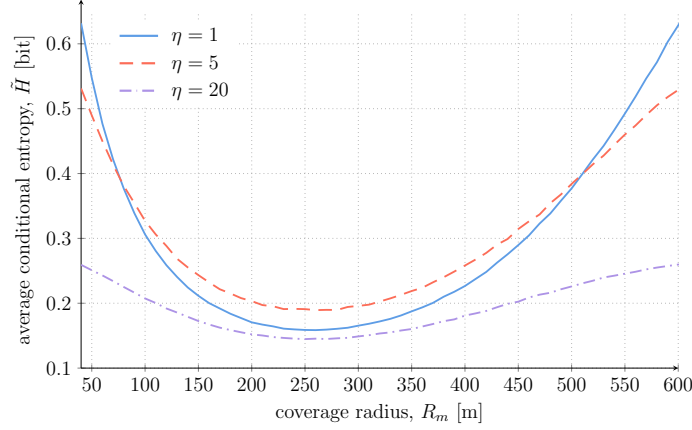


Figure 6. Average conditional uncertainty of the FS receiver, $\tilde{H}$, vs coverage range R_m . A binary source is monitored, considering different asymmetry factors. $\rho \simeq 0.05$, $R = 10$ m, $q = 0.005$, $\alpha = 0.02$, $\zeta = 10^{-4}$, $p_e = 0.1$.

collecting a 0.⁸ A comparison with Fig. 5b provides a visual intuition of the sub-optimality of the forgetful approach.

A. Numerical results and discussion

Unless otherwise stated, all the parameters for the results are set as specified in Sec. III-B. The uncertainty $\tilde{H}$ of the FS receiver was estimated via Monte Carlo simulations, taking the mean across multiple runs of $h(y^n)$ over sufficiently long intervals to estimate the limiting behavior in (7). As a first result, we report in Fig. 6 $\tilde{H}$ against the coverage range R_m . Three types of sources are considered: symmetric ($\eta = 1$, solid curve), slightly asymmetric ($\eta = 5$, dashed curve), and highly asymmetric ($\eta = 20$, dash-dotted curve). The fundamental trade-off driven by range selection that emerged in the forgetful case is confirmed even when all collected messages over time are taken into account. Large uncertainty is experienced both for a small coverage, due to sporadic updates, and for exceedingly large coverage due to an increase in collisions and to the reception of less reliable readings. An interesting trend emerges considering the behavior under different values of η . In intermediate coverage ranges, i.e., when the receiver is fed with a good mix of reliable and less reliable messages, a relatively low asymmetry in the source leads to higher values of $\tilde{H}$ in comparison to the tracking of a symmetric process. The result can be explained referring to the discussion carried out in Examples 2 and 3, as the retrieval of sequences of possibly erroneous messages

⁸Similar trends were noted in [31] for the UoI, which, as discussed, coincides with the studied metric for the forgetful receiver in the special case of perfectly reliable readings.

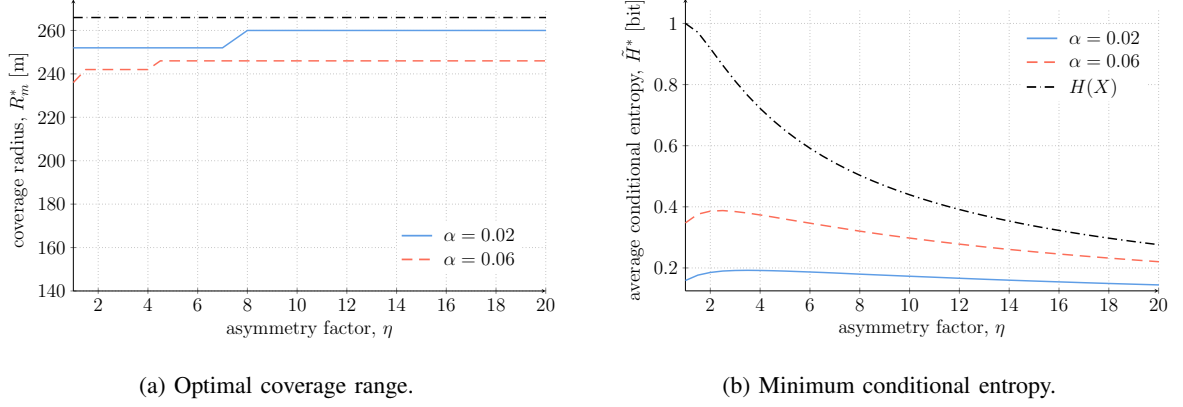


Figure 7. Optimal coverage range R_m^* and minimum average conditional entropy $\tilde{H}^*$ obtained with R_m^* vs asymmetry factor η for a FS receiver. The horizontal line corresponds to coverage for minimum AoI in (a) and to the source stationary entropy in (b).

reporting a visit to the less likely state hinders an accurate belief at the receiver. When asymmetry increases further (e.g., $\eta = 20$), the trend is more than counterbalanced by the higher predictability of the source behavior (see, e.g., Fig. 5c). Notably, asymmetry is always beneficial in terms of $\tilde{H}$ for very low and very large coverage radii. In the former case, updates are highly reliable, so that uncertainty is reset to low values with every incoming message, and even a slight reduction of the source stationary entropy at which $h(y^n)$ tends to converge without new readings is useful. In the latter, most messages are poorly informative and $\tilde{H}$ eventually converges to the source entropy $H(X)$.

These aspects are further explored in Fig. 7, showing the optimal coverage range and the corresponding minimum conditional entropy $\tilde{H}$ against the asymmetry factor, considering an FS receiver and two values of the exponent α of the reading reliability function. The role played by η is apparent, with asymmetry being detrimental up until $\eta \simeq 15$ for the set of parameters under study and $\alpha = 0.02$. As expected, a larger α leads to higher uncertainty, and the optimal system configuration is achieved for a smaller R_m^* , sacrificing the frequency of updates to favor their reliability. It is also interesting to compare the results with the trends reported in Fig. 4 for the forgetful case. The optimal coverage range for the FS approach is larger and converges to the value that minimizes AoI (i.e., ~ 265 m) more quickly as η increases. Overall, the possibility to leverage beliefs on the full history of received updates allows to better cope with potentially less reliable readings, and benefits from an increase in the frequency of update receptions. In all cases, extending the coverage radius beyond $(\pi\rho\zeta(1-p_e))^{-1/2}$ is detrimental, as it would trigger an overloaded channel (i.e., load larger than 1 pkt/slot) and result in a reduction in the number of received messages. In general, the

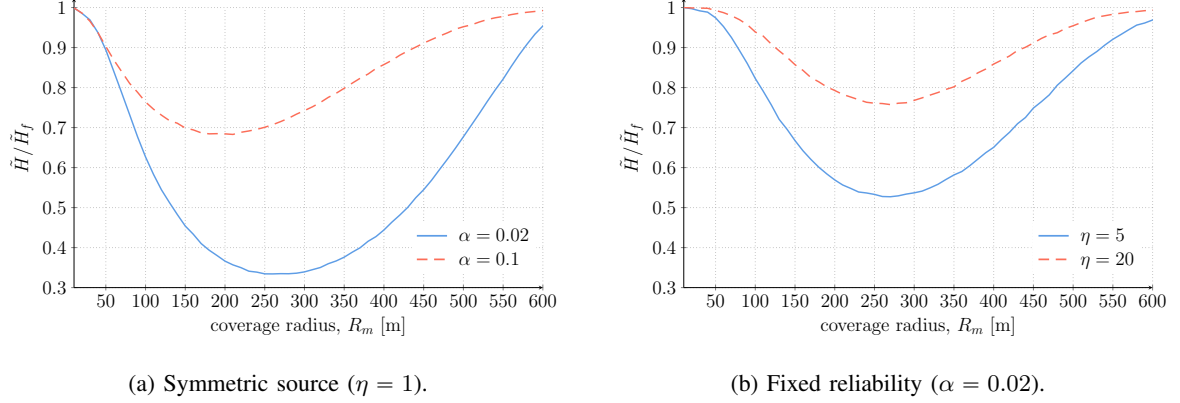


Figure 8. Ratio of the average conditional entropy of the FS and of the forgetful estimator, i.e., $\tilde{H}/\tilde{H}_f$, vs coverage range R_m . In (a), a symmetric source is tracked ($\eta = 1$) and different values of α are considered. In (b), $\alpha = 0.02$, and two values of the asymmetry factor η for the source are shown.

importance of properly tuning the coverage range is also evident for an FS receiver, although the optimal setting is closer to what would be obtained considering AoI alone. This can be seen in Fig. 6, noting how uncertainty deteriorates rather quickly when R_m moves away from its optimal value. We can also note that the FS receiver increases the gain over the stationary entropy even for higher values of α , as shown in Fig. 7b: this further confirms that the FS receiver is able to exploit even unreliable information, reducing the entropy more than the forgetful receiver even for low values of η .

A direct comparison of the performance achieved with the two types of receiver is shown in Fig. 8, reporting the ratio $\tilde{H}/\tilde{H}_f$ versus the coverage radius R_m , and considering different sets of parameters. In all cases, both receivers exhibit the same behavior for very small and very large coverage areas. The former condition is formally captured in Remark 5, as all readings are perfectly reliable for a single coverage area. As R_m diverges, the channel is dominated by collisions and the uncertainty converges to $H(X)$. On the other hand, significant improvements are obtained for intermediate coverage, with a reduction of the uncertainty up to a factor ~ 3 attained by leveraging the whole set of collected observations Y^n . The most notable gains are experienced when α increases, as the FS receiver manages to better handle incoming unreliable readings by relying on past messages. Smaller, yet still noticeable gaps can be observed when reliability degrades less sharply with distance or in the presence of highly asymmetric sources, as shown in Fig. 8b.

Further insight on the behavior of the two receivers is offered in Fig. 9, reporting the cumulative distribution function (CDF) of the receiver uncertainty at a generic point in time, i.e., of $h_f(w_n, \delta_n)$ (dashed lines) and $h(y^n)$ (solid lines). A symmetric source is considered

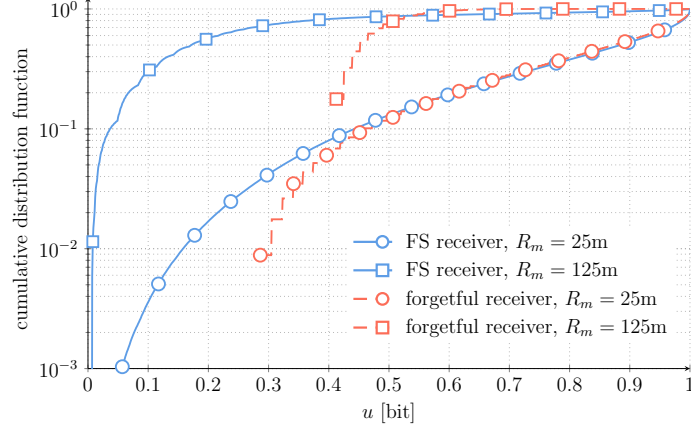


Figure 9. Cumulative distribution function of $h_f(w_n, \delta_n)$ for the forgetful receiver, and of $h(y^n)$ for the FS receiver. Different values of coverage range are considered, tracking a symmetric source.

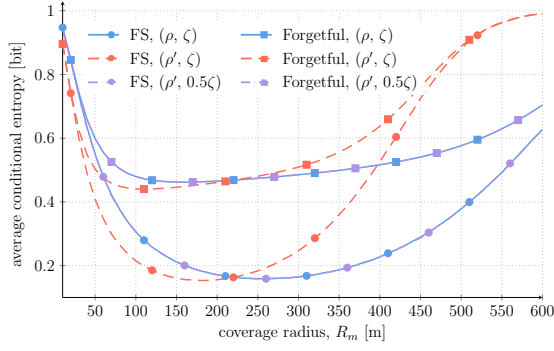
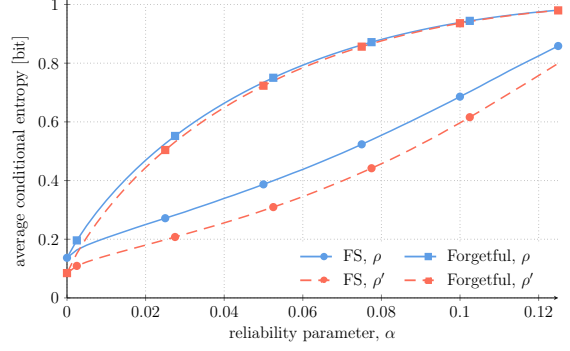
($\eta = 1$), and results are shown for two different coverage radii: a relatively small one (25 m, circle markers) and a larger one (125 m, square markers). For the forgetful receiver, the metric can be obtained from (17), (18)-(19) as

$$P[h(w_n, \delta_n) \leq u] = \sum_{\substack{(w_n, \delta_n) \text{ s.t.} \\ h(w_n, \delta_n) \leq u}} p(w_n, \delta_n).$$

In the FS case, the empirical CDF was derived via Monte Carlo simulations, tracking the values of $h(y^n)$ over multiple episodes.

The plot pinpoints how the uncertainty of the forgetful approach is bounded from below by the entropy of the state of the source conditioned on the value of an incoming message, determined by the distribution $\bar{p}(x_n | w_n)$ computed in (12). The more unreliable readings can be, i.e., the larger the radius, the higher the uncertainty of the receiver upon decoding (see Example 1). The situation changes when the whole sequence Y^n is accounted for. In this case, the receiver can experience a much lower uncertainty, e.g., by receiving a sequence of updates, reinforcing the belief of the source about being in a specific state (Example 2). The result is also interesting from a system design standpoint, providing guidelines on the choice of the receiver complexity and on the coverage radius when not only the average uncertainty but also the probability of not exceeding a possibly critical value for the application of interest is to be considered.

Finally, we analyze the interplay of the maximum coverage radius R_m , the accuracy decay parameter α , the transmission probability ζ , and the average number of nodes per unit area ρ . Fig. 10(a) shows the entropy of the two receivers as a function of the coverage radius

(a) Average entropy as a function of R_m , with $\alpha = 0.02$.(b) Average entropy as a function of the reliability parameter α with $m = 2454$ and $m' = 2m = 4908$, corresponding to ρ and ρ' nodes per unit area, respectively, with $R_m = 125\text{m}$.Figure 10. Average uncertainty at the FS ($\tilde{H}$) and forgetful ($\tilde{H}_f$) receiver, with $\rho \simeq 0.05$ and $\rho' = 2\rho \simeq 0.1$, $\zeta = 10^{-4}$, $R = 10\text{ m}$, $\eta = 1$, $q = 0.005$, $p_e = 0.1$.

under different settings: we note that increasing the number of nodes per unit area to $\rho' = 2\rho$ allows the receiver to obtain more accurate information if R_m is set to a low value, as there are more sensors providing accurate readings. However, the larger number of nodes increases the risk of collision: setting a relatively high R_m results in a much higher entropy due to the lower success probability. On the other hand, operating with 2ρ and $\zeta/2$ results in almost exactly the same performance as with the original values, because doubling the number of nodes in each area and halving their transmission probability leads to approximately the same collective behavior. This also highlights a system design trade-off: a denser sensor network may be more expensive to deploy, but the lifetime of individual nodes will be longer, as their transmission rates, and, thus, their energy expenditure, will decrease accordingly without any performance loss in terms of the receiver entropy.

The effect of the reliability parameter α is shown in Fig. 10(b), considering a fixed coverage radius and transmission probability. Following Remark 5, we see that the two receiver models are identical for $\alpha = 0$. The FS receiver is then better able to exploit a larger number of sensors, while the entropy of the forgetful receiver quickly grows with α independently of the number of nodes in the area. However, even relatively low values of α substantially degrade the entropy of both receiver types.

V. SYSTEM OPTIMIZATION

The results discussed so far depended on two simplifying assumptions: (1), the receiver cannot rely on any geographical information for received packets, only accounting for the

average reliability level of incoming readings; (2), all devices employ the same transmission probability ζ , irrespective of their location, and thus on the reliability of their updates. Both aspects trigger relevant questions on whether system performance could be optimized by relaxing the working hypotheses. In the first case, uncertainty might be reduced by properly weighting the reliability of the content of a decoded packet based on its source's accuracy. The problem is also relevant from a system design standpoint, as acquiring and reporting geographical information entails additional complexity at the device side (e.g., requiring GNSS capabilities) as well as overhead to notify the position of the transmitter in a sent message. Similarly, more complex protocols that adapt the transmission probability to the position of the source could be considered, accounting for the different accuracy as well as for the larger number of devices that populate farther regions.

A. The role of location awareness

Let us first consider the impact of position-aware nodes. To study this, we extend our model assuming that each device knows the region $\mathcal{A}_i$ it belongs to, and piggybacks this information whenever sending an update. Consequently, the observation Y_n has alphabet $(\mathcal{X} \times \mathcal{K}) \cup \{*\}$, and the analytical framework presented for the two types of receiver we consider can be readily extended to derive the conditional distribution on the current state of the source given the available knowledge. Specifically, we have:

- *Forgetful receiver*: in this case, the uncertainty at time n is driven not only by the latest received reading, W_n , and the time elapsed since its decoding, Δ_n , but also by the region from which it originated, D_n . Following the same approach discussed in Sec. III, upon reception of a packet at time n_0 , we get

$$\bar{p}(x_{n_0} | w_{n_0}, d_{n_0}) = \frac{\bar{p}(w_{n_0}, d_{n_0} | x_{n_0}) \pi_{x_{n_0}}}{\sum_{x'_{n_0} \in \mathcal{X}} \bar{p}(w_{n_0}, d_{n_0} | x'_{n_0}) \pi_{x'_{n_0}}}.$$

Note that, as the location information is available to the receiver, the condition is on both w_{n_0} and d_{n_0} , while the original forgetful receiver update in (12) only uses the former, computing the emission probability $\bar{p}(w_{n_0} | x_{n_0})$ by applying the law of total probability in (13). In this case, the emission probability is

$$\bar{p}(w_{n_0}, d_{n_0} | x_{n_0}) = \bar{p}(w_{n_0} | x_{n_0}, d_{n_0}) \bar{p}(d_{n_0}) \quad (21)$$

where $\bar{p}(w_{n_0} | x_{n_0}, d_{n_0})$ and $\bar{p}(d_{n_0})$ were derived in (15) and (14), respectively. The distribution for any $n = n_0 + \delta_n > 0$ between the current and next decoding follows by taking the δ_n -step evolution of the process via the transition matrix T , allowing us to

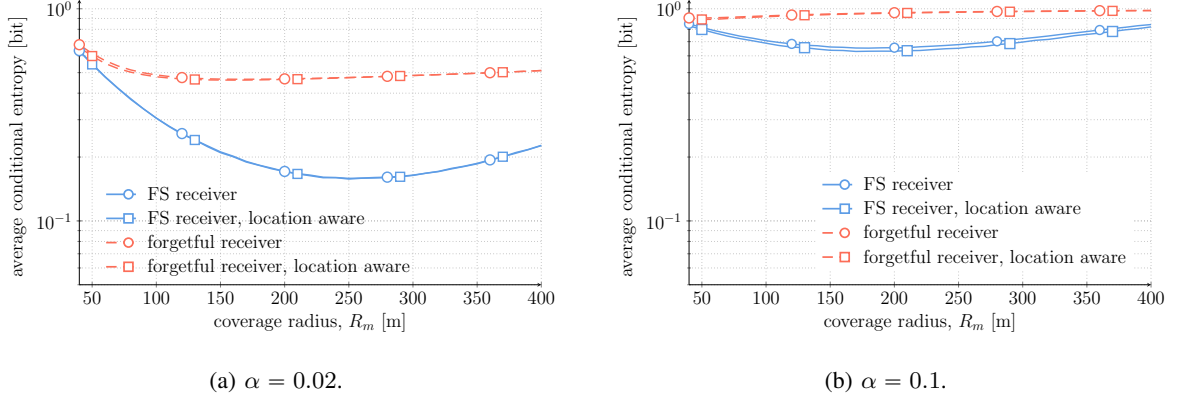


Figure 11. Average conditional entropy for a forgetful ($\tilde{H}_f$) and a FS ($\tilde{H}$) receiver, reported against the coverage radius R_m . Circle markers denote the performance achieved when the receiver has no knowledge about the position of the transmitters, whereas square markers refer to the case of location being piggybacked into sent messages. In (a), the reliability parameter α is set to 0.02, whereas (b) refers to the case $\alpha = 0.1$. A symmetric source is monitored.

compute $h_f(w_n, \delta_n, d_n)$ for a generic time slot n . The average uncertainty at the receiver, $\tilde{H}_f$, can in turn be evaluated by taking the expectation of this quantity over the joint distribution

$$p(w_n, \delta_n, d_n) = p(\delta_n) \sum_{x_{n_0} \in \mathcal{X}} \bar{p}(w_{n_0}, d_{n_0} | x_{n_0}) \pi_{x_{n_0}}$$

where we leveraged again the independence of the AoI from the content and origin of the last received message.

- *FS receiver*: the HMM characterizing the relationship between source state and channel output can again be applied, taking into account the fact that at every slot the receiver either observes $*$ (collision or idle slot), or a pair (w_n, d_n) in the event of successful decoding. The emission probability in the latter case is given in (21), allowing us to obtain the conditional probability via the standard HMM forward recursion.

The results of this analysis are shown in Fig. 11, where we report the average conditional entropy for the forgetful (dashed lines) and FS (solid lines) receivers, considering the case without (circle markers) and with (square markers) location awareness. A symmetric source was considered, and the groups of curves refer to two values for the reading reliability parameter α .

The first and key outcome that can be extracted is that piggybacking location information on the updates has a very limited impact on the overall uncertainty at the receiver, both in the forgetful and in the FS case. The result is non-obvious and provides a relevant design take-away, supporting the relevance of simpler IoT systems in which devices need not be

aware of their position. From this standpoint, the modeling assumptions introduced in Sec. II prove to be sound, leading to a broad applicability of the main setting considered in our framework. A detailed look at Fig. 11 also reveals that a slight improvement (up to 1.96%) can be noticed for the forgetful receiver already for $\alpha = 0.02$. In the FS case, instead, the role of weighing the reliability of incoming updates based on the distance of their sender only emerges when α is slightly larger, with an overall reduction in $\tilde{H}$ up to 3.73% for $\alpha = 0.1$. This relatively small effect for both receiver types is due to the relative frequency of low-quality updates, which is quite high due to the spatial distribution of nodes: when receiving sequences of low-quality updates, the two receivers still have a high entropy, and the knowledge of the location of the higher-quality updates has a small effect due to their low relative frequency.

B. Transmission probability optimization

To understand whether an adaptation of the transmission probability can be beneficial, consider an extension of the model in which all nodes within the same region $\mathcal{A}_i$ access the channel with probability ζ_i . We thus obtain a vector $\zeta = [\zeta_0, \dots, \zeta_{K-1}]$, and can find the optimal configuration ζ^* by solving the problem

$$\begin{aligned} \zeta^* = \underset{\zeta \in [0, \zeta_{\max}]^K}{\operatorname{argmin}} \quad & \tilde{H} \\ \text{s.t.} \quad & 0 \leq \zeta_i \leq \zeta_{\max} \quad \forall i \end{aligned} \quad (22)$$

where $\zeta_{\max}$ is the maximum transmission rate for a node, e.g., due to duty cycle or other operational constraints. To this aim, we define the set of threshold-based solutions $\mathcal{T}$. A solution ζ belongs to $\mathcal{T}$ if and only if it respects the following condition:

$$\zeta \in \mathcal{T} \iff \exists K^* \in \{0, \dots, K-1\} : \zeta_i = \zeta_{\max} \ \forall i < K^* \wedge \zeta_j = 0 \ \forall j > K^*.$$

We also define the accuracy $\bar{\lambda}(\zeta)$ of a solution, i.e., the probability of a packet received when using solution ζ reporting the correct value of the state, as

$$\bar{\lambda}(\zeta) = \bar{p}(w_{n_0} | x_{n_0}) \mathbb{1}(x_{n_0}, w_{n_0}).$$

We then state two Lemmas, which will be useful in determining the optimal solution.

Lemma 2. *Consider any solution ζ . If m_i is a non-decreasing function of i , i.e., there are more nodes in outer areas, there exists a threshold-based solution $\zeta' \in \mathcal{T}$ with an equal success probability, i.e.,*

$$m_j \geq m_i \ \forall j > i \implies \forall \zeta \ \exists \zeta' \in \mathcal{T} : p_s(\zeta') = p_s(\zeta).$$

Proof: The proof of the Lemma is provided as Appendix C. ■

Lemma 3. *If we consider the set $\mathcal{L}(p_s) = \{\zeta : p_s(\zeta) = p_s\}$, the solution*

$$\zeta^*(p_s) = \operatorname{argmax}_{\zeta \in \mathcal{L}(p_s)} \bar{\lambda}(\zeta),$$

belongs to the set $\mathcal{T}$ of threshold-based policies if R , α , and $\zeta_{\max}$ satisfy

$$\frac{1}{(1+R)^\alpha} \geq \frac{1-\zeta_{\max}}{(1+2R)^\alpha} + \zeta_{\max}. \quad (23)$$

Proof: The proof of the Lemma is provided as Appendix D. ■

We then use the Lemmas to prove the following result.

Theorem 1. *If the number of nodes in each zone m_i is a non-decreasing function of i , and $\frac{1}{(1+R)^\alpha} \geq \frac{1-\zeta_{\max}}{(1+2R)^\alpha} + \zeta_{\max}$, there exists an optimal activation probability vector ζ^* presenting a threshold structure, i.e., $\zeta^* \in \mathcal{T}$. In other words, the intersection between the set of optimal solutions to problem (22) and the set of threshold-based solutions $\mathcal{T}$ is not empty.*

Proof: We start with a trivial observation: any solution that generates a load higher than 1, i.e., with $\sum_{i=0}^{K-1} m_i \zeta_i > 1$, cannot be optimal, as we can always find a solution with the same ratio of packets from each zone and a higher success probability by reducing the load accordingly. We then concentrate only on solutions with a load lower than 1. Consider now any pair of solutions ζ and ζ' , with the same success probability, i.e., $p_s(\zeta) = p_s(\zeta')$. It is immediate to show that, if $\bar{\lambda}(\zeta') \geq \bar{\lambda}(\zeta)$, i.e., received packets have a higher probability of containing the correct state, the entropy of the receiver is never higher when the nodes transmit according to ζ' than when using ζ .

The theorem then straightforwardly follows from the two lemmas: as Lemma 2 states that we can always construct a threshold-based solution with the same throughput as any other solution, and Lemma 3 states that the highest-accuracy solution with a given success probability must be threshold-based, there is always a threshold-based solution which leads to an entropy at least as low as any other solution. The set of optimal solutions must then always include at least a threshold-based one. ■

Theorem 1 offers another relevant hint for system design, proving how optimization of the transmission probability based on the positions of the nodes does not play a key role. This is especially relevant, since implementing a per-zone channel access vector ζ^* would require once more some form of location awareness with respect to the tracked process for all nodes and entail further protocol overhead to distribute the transmission parameters. Both aspects can be critical in IoT systems, where the simplicity of devices and protocols is paramount.

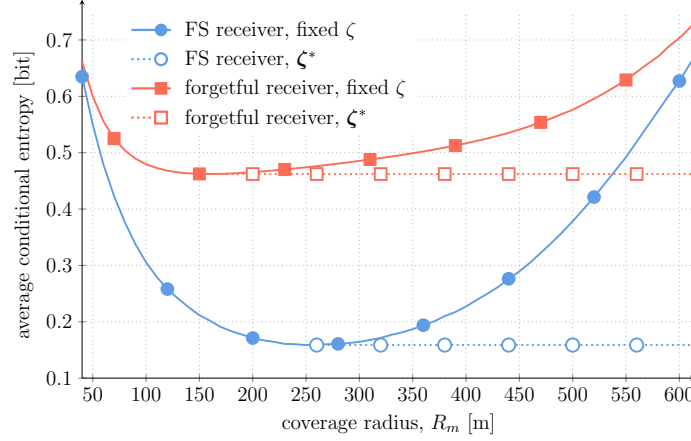


Figure 12. Average conditional entropy vs coverage range R_m , considering a symmetric source. Solid lines report the behavior of an FS and a forgetful receiver when all nodes within coverage transmit with the same probability $\zeta = 10^{-4}$. Dashed lines denote the performance of the two receivers when the transmission probabilities for each coverage zone are obtained by solving the optimization problem (22).

Moreover, the result confirms that the modeling assumptions introduced in Sec. II are valid: by simply setting an appropriate coverage range and allowing nodes in the area to transmit as dictated by their duty cycles, the system achieves a nearly optimal solution, as the only difference from ζ^* is given by (at most) the most external active region. This is clearly shown in Fig. 12, reporting $\tilde{H}$ against R_m when tracking a symmetric source, and considering the case in which nodes access the channel with probability $\zeta = \zeta_{\max} = 10^{-4}$ (solid lines) or with ζ^* (dashed lines). The latter has been obtained by solving the optimization problem in (22) via standard numerical tools. The performance of both FS (circle markers) and forgetful (square markers) receivers is reported. Up until $R_m = R_m^*$, the solutions coincide, as the optimal approach foresees all nodes to transmit with the same probability. Conversely, for larger radii, the uncertainty obtained for a constant ζ increases due to the reception of unreliable updates, and the optimal solution is simply to switch off any transmission from outside R_m^* . In this perspective, the simpler modeling approach tackled in this paper proves to be especially valuable in determining the optimal coverage and the minimum attainable uncertainty. Also note how the effect of keeping the nodes in region K^* active, with $\zeta_{K^*} < \zeta_{\max}$ in this case, as dictated by the optimal structure of Theorem 1, does not play a significant role.

To conclude, we also observe that, while the theorem sets some conditions on the values of α , R , and $\zeta_{\max}$, practical system parameters are well within these bounds. This is illustrated in Fig. 13, where, for any value of the granularity $R = R_m/K$, i.e., the width of each annular region within the coverage area, the region below the corresponding curve gives all α values

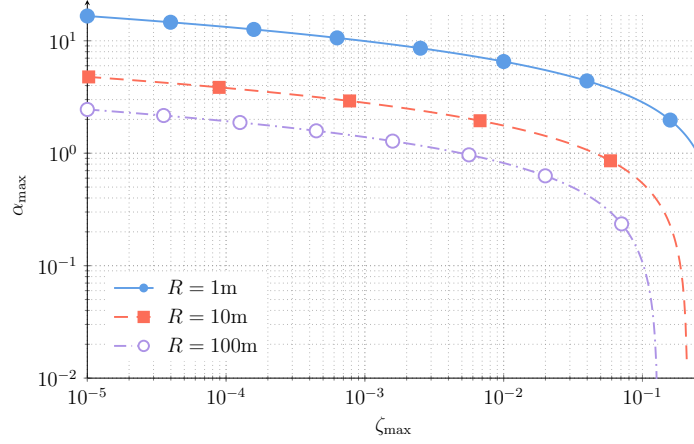


Figure 13. Maximum value of α under which Theorem 1 holds as a function of $\zeta_{\max}$ for different values of R .

for which the threshold-based solution is proven to be optimal. Decreasing the duty cycle or the granularity of the zone division allows for larger values of α . All results in the simulations we show in this work fall within the condition in the theorem.

VI. MONITORING OF AN M -STATE PROCESS

To conclude our study, we tackle the more general case of tracking a Markov source with an arbitrary number of states M and alphabet $\mathcal{X} = \{0, \dots, M-1\}$. We assume once more that the nodes are not location aware and that all use a common transmission probability ζ .

The one-step transition probabilities of the considered source model are illustrated in Fig. 14a, providing a direct extension of the symmetric source case studied so far and leading to a stationary distribution $\pi_i = 1/M$, $\forall i \in \mathcal{X}$. In the remainder, we will focus solely on the FS receiver case, whose uncertainty $\tilde{H}$ can be captured extending the HMM that describes the relationship between X_n and the channel outputs Y_n . Specifically, for any distribution $p(z_n | x_n, i)$, $i \in \mathcal{K}$, describing the probability that a device will produce reading Z_n given the current state of the tracked process and its distance from it, the emission probability of the HMM can be obtained as

$$p(y_n | x_n) = \begin{cases} p_s \sum_{d_n \in \mathcal{K}} p(z_n | x_n, d_n) \bar{p}(d_n) & y_n \in \mathcal{X} \\ 1 - p_s & y_n = * \end{cases}$$

where $\bar{p}(d_n)$ is as derived in (14).

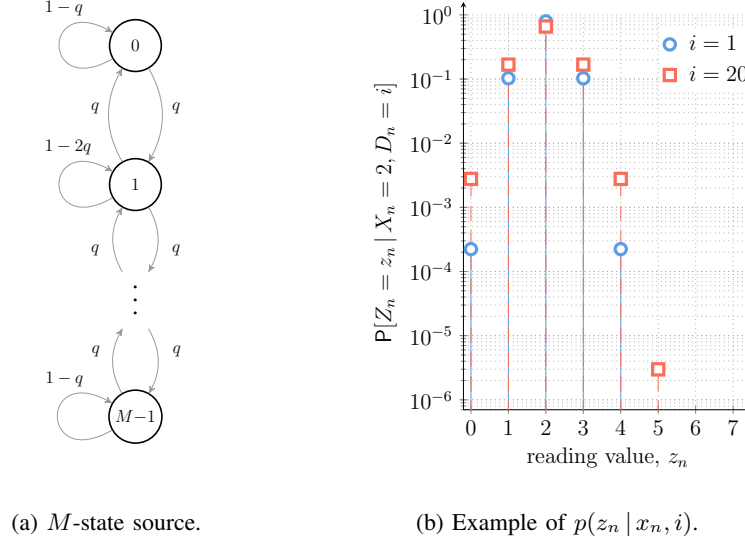


Figure 14. A diagram of the M -state Markov source, with transition probabilities, and an example of the PMF of the observation produced by a sensor in area $\mathcal{A}_i$ given the state of the source, i.e., $p(z_n | x_n, i)$. The result is obtained assuming that $\mathcal{X} = \{0, \dots, 7\}$ and the true source state is $X_n = 2$, and considering the model in (24)-(25). To determine σ_i^2 , the reliability exponent of the 2-state model was set to $\alpha = 0.02$.

In the remainder, we focus on a specific example for the readings reliability, considering a discrete Gaussian distribution

$$p(z_n | x_n, i) = \frac{e^{-(z_n - x_n)^2 / (2\sigma_i^2)}}{\sum_{z'_n \in \mathcal{X}} e^{-(z'_n - x_n)^2 / (2\sigma_i^2)}} \quad (24)$$

where the parameter σ_i^2 controls how likely it is to obtain readings farther away from the true value. An example of the distribution is reported in Fig. 14b, for $M = 8$, showing the PMF of Z_n conditioned on having the source in state $X_n = 2$ when the reading is produced by a node very close to the center ($i = 2$, solid line) or farther away ($i = 10$, dashed line). In order to have a relevant comparison with the results presented earlier, we set the parameter σ_i^2 so that, for $M = 2$, the probability that a device in area $\mathcal{A}_i$ produces a correct reading corresponds to the one considered in Sec. II. This is obtained if $p(z_n | x_n, i) = \lambda_i$ when $z_n = x_n$. Imposing this condition within (24) leads, after straightforward manipulations, to

$$\sigma_i^2 = \left(2 \ln \left(\frac{\lambda_i}{1 - \lambda_i} \right) \right)^{-1}. \quad (25)$$

Using this setting, we report in Fig. 15 the average uncertainty $\tilde{H}$ against the coverage range for a source with $M = 2$ (solid line), $M = 4$ (dashed line), and $M = 8$ (dash-dotted line) states. In all cases, the transition probability q is set to $5 \cdot 10^{-3}$ and $\alpha = 0.02$. As expected, a larger number of source states leads to a higher uncertainty, with the effect being especially visible for very low or large coverage, where $\tilde{H} \rightarrow H(X) = -\sum_{x \in \mathcal{X}} \pi_x \log_2 \pi_x$.

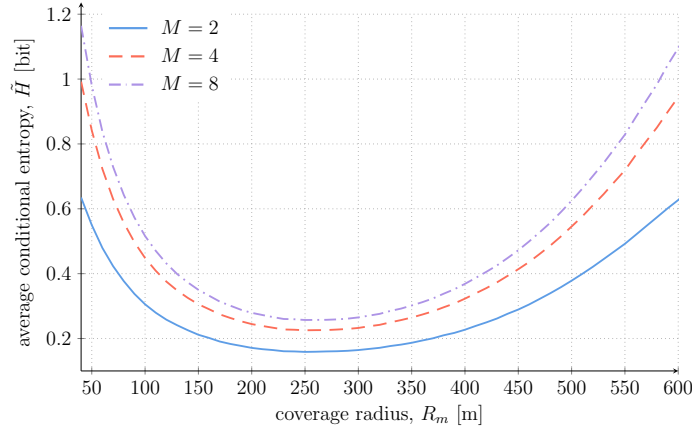


Figure 15. Average conditional entropy of the FS receiver when tracking a multi-state source with $M = 8$. The source transitions with probability $q = 5 \cdot 10^{-3}$, and the reliability parameters for the produced readings are set following (24) and (25). To determine σ_t^2 , the reliability exponent of the 2-state model was set to $\alpha = 0.02$.

More interestingly, the fundamental trends and trade-offs discussed for the two-state case are confirmed, validating the approach of considering a simpler system model and allowing us to broaden the messages provided in the paper.

VII. CONCLUDING REMARKS

The freshness of information is a key metric for the development of 6G massive access, and its study is moving beyond the simple definition of AoI and toward more expressive metrics that directly target application performance. At the same time, the temporal aspect of distributed IoT observations, which has been extensively studied, needs to be complemented by a spatial component, integrating readings from sensors that may be farther from the event of interest and thus less reliable, but still useful.

In this work, we characterized the uncertainty of the receiver by studying its entropy in such a spatio-temporal observation system, showing its behavior under state asymmetry and multi-state Markov chains, and provide some general insights for optimization. Further, we showed that system optimization can be carried out with simple policies, which can be implemented in devices with limited hardware, with a negligible performance loss.

Future avenues of research include the characterization of a wider set of spatio-temporal systems, considering multiple points of interest and structural distance metrics based on graph theory instead of simple Euclidean distance, in order to fully represent modern IoT systems while still providing a solid theoretical characterization that can be used for further optimization. This can be extended to a wider consideration of the system: we can imagine

a vehicular traffic estimation problem measuring flow at crossings, i.e., nodes on a graph, in which entropy can capture the structure and system dynamics holistically. Furthermore, our research aims at triggering novel contributions on the topic of spatio-temporal freshness, e.g., exploring it under different channel models and taking into account the specific physical layer and protocol aspects of practical IoT systems.

APPENDIX

A. On the definition of $\tilde{H}$

We prove that $\tilde{H}$ is well defined by showing that $\lim_{n \rightarrow \infty} H(X_n | Y^n)$ exists. To this aim, consider the joint process (X_n, Y_n) . We have $p(x_n, y_n | x^{n-1}, y^{n-1}) = p(x_n, y_n | x_{n-1})$, since X_n is Markovian, and Y_n is only driven by the current value of the source. Accordingly, (X_n, Y_n) is a finite-state Markov chain, which is time-homogeneous (as X_n is homogeneous and the observation probabilities over the channel do not change over time). The chain is also irreducible. To prove this, it is sufficient to show that the one-step transition probability $p(x_{n+1}, y_{n+1} | x_n, y_n)$ between any two states is strictly positive [52]. We then observe that $p(x_{n+1}, y_{n+1} | x_n, y_n)$ can be factorized as $p(y_{n+1} | x_{n+1})p(x_{n+1} | x_n)$, and that both factors are positive. For the transition probabilities of the tracked process, $p(x_{n+1} | x_n) > 0$ by definition, as we assume $q > 0$ (see Sec. II) and do not consider pathological cases in which the monitored Markov chain stays indefinitely in one state. In turn, $p(y_{n+1} | x_{n+1}) > 0$ follows from (20), recalling that $p_s > 0$ (as clarified in (5) since the erasure probability is $p_e < 1$ and $\zeta \in (0, 1)$) and that $\bar{p}_{W_n | X_n}$ is also strictly positive from (16). Incidentally, we also remark that, in the case $K = 1$, i.e., coverage only consists of the innermost region, all updates are reliable, and the process only visits states (x_n, y_n) with $y_n \in \{x_n, *\}$. In this special case, one would thus consider a reduced chain including only those states, which is irreducible following the same reasoning. Finally, (X_n, Y_n) is also aperiodic, since, e.g., state $(0, 0)$ has a strictly positive self-transition probability from the arguments above. These conditions ensure that the process is ergodic and a stationary distribution exists. In the remainder, we will assume that the initial distribution is the stationary one, so that the process is stationary.

We now note that

$$H(X_n | Y^n) = H(X_n | Y_n, \dots, Y_0) \leq H(X_n | Y_n, \dots, Y_1) \stackrel{(a)}{=} H(X_{n-1} | Y_{n-1}, \dots, Y_0)$$

where (a) follows from the stationarity of (X_n, Y_n) . The sequence $H(X_n | Y^n)$ is thus non-increasing and non-negative, proving its convergence to a limit.

B. Proof of Lemma 1

Consider first Δ_n . The process evolves from Δ_{n-1} as $\Delta_n = \Delta_{n-1} + 1$ with probability $1 - p_s$ (no packet received) or $\Delta_n = 0$ otherwise. We thus have a countable state Markov chain, which can easily be seen to be irreducible. The chain is also aperiodic (as the self-transition probability for state 0 is $p_s > 0$) and positive recurrent (since the return time to state 0 is a geometric r.v. of parameter p_s , with mean value $1/p_s > 0$). The process is then ergodic, and we will assume its initial distribution to be equal to the stationary distribution. Ergodicity can be inferred also for (X_n, W_n) , following the same arguments provided in Appendix 1 for (X_n, Y_n) . Observing that Δ_n is independent of (X_n, W_n) , the joint process (X_n, W_n, Δ_n) is also a Markov chain, irreducible and aperiodic. Its stationary distribution is the product of the stationary distributions of (X_n, W_n) and Δ_n , and is thus proper, proving the ergodicity and stationarity of the chain. Accordingly, $H(X_n | W_n, \Delta_n)$ is the same for all n . The limit thus exists and can be computed using the stationary distribution of (W_n, Δ_n) as in the rightmost equality of the Lemma statement.

C. Proof of Lemma 2

The success probability of a setting in which an arbitrary number n of nodes transmit with probability $\zeta_{\max}$ is

$$p_s(n) = n\zeta_{\max}(1 - p_e)(1 - \zeta_{\max}(1 - p_e))^{n-1}.$$

In the following, we define symbol $\zeta_{\text{eff}} = (1 - p_e)\zeta_{\max}$ for the sake of brevity.

We can then find the value n^* that solves $p_s(n) = p_s(\zeta)$. The equation does not have a closed-form solution, but it can be solved numerically, and n^* exists if the number of nodes m_i is non-decreasing, as the maximum throughput for n transmitting nodes is limited to $(1 - p_e)/n$. If we take the second derivative of $p_s(n)$ with respect to n , we get

$$\frac{\partial^2 p_s(n)}{\partial n^2} = \zeta_{\text{eff}}(1 - \zeta_{\text{eff}})^{n-1} \log(1 - \zeta_{\text{eff}})(2 + n \log(1 - \zeta_{\text{eff}})).$$

We know that $\zeta_{\text{eff}}(1 - \zeta_{\text{eff}})^{n-1}$ is always positive. On the other hand, $\log(1 - \zeta_{\text{eff}})$ is always negative, as $\zeta_{\text{eff}} < 1$. In order for the second derivative to be negative, we then need to set $n \log(1 - \zeta_{\text{eff}}) > -2$.

Observe that, to maintain a load lower than 1, we must have $n\zeta_{\text{eff}} \leq 1$. Any solution with a load higher than 1 has a lower success probability than solutions with a lower load, so this does not reduce the validity of the Lemma. We can thus set a more restrictive condition:

$$n \log(1 - \zeta_{\text{eff}}) > -2n\zeta_{\text{eff}} \implies n \log(1 - \zeta_{\text{eff}}) > -2 \iff \frac{\partial^2 p_s(n)}{\partial n^2} < 0.$$

Removing the factor n on each side, we have $\log(1 - \zeta_{\text{eff}}) > -2\zeta_{\text{eff}}$, which is true for $\zeta_{\text{eff}} < 0.796$. As $n > 1$, this is a less restrictive condition than $n\zeta_{\text{eff}} \leq 1$, and the second derivative is always negative in the interval of interest.

The function is thus concave, and we can find the value of n^* with the bisection method. We then determine K^* as:

$$K^* = 1 + \sup \left\{ k \in \{0, \dots, K-2\} : \sum_{i=0}^k m_i < n^* \right\}.$$

We denote the number of nodes in the zones before K^* as $m_b = \sum_{i=0}^{K^*-1} m_i$. If the inequality is strict, we then need to find the correct value ζ_{K^*} , which is the solution of

$$p_s(\zeta) = (1 - \zeta_{\text{eff}})^{m_b} (1 - (1 - p_e)\zeta_{K^*})^{m_{K^*}} \left[\frac{(1 - p_e)\zeta_{K^*}m_{K^*}}{1 - (1 - p_e)\zeta_{K^*}} + \frac{\zeta_{\text{eff}}m_b}{1 - \zeta_{\text{eff}}} \right].$$

We can find this solution numerically using the bisection method, as the success probability monotonically increases with ζ_{K^*} as long as $(m_{-K^*} + m_{K^*})\zeta_{\text{eff}} \leq 1$.

D. Proof of Lemma 3

Consider a generic solution $\zeta \notin \mathcal{T}$. If ζ is not threshold-based, we need to prove that a threshold-based solution ζ' exists, with $p_s(\zeta') = p_s(\zeta)$, and ζ' has a higher accuracy. We can do so through a water-filling approach, by proving the existence of a swap in ζ towards the inner regions that leaves $p_s(\zeta)$ unchanged but improves the accuracy. Thus, we can reach the desired threshold-based ζ' at the end of a necessarily finite-step procedure where, as long as two values ζ_j and ζ_k are both inside $(0, \zeta_{\text{max}})$, with $j < k$, a transfer is made from ζ_k to ζ_j to either saturate ζ_j to ζ_{max} or decrease ζ_k to 0.

Formally, we define a series of intermediate solutions $\zeta^{(n)}$, with $\zeta^{(0)} = \zeta$. We consider zones $j = \inf \{i : \zeta_i^{(n)} < \zeta_{\text{max}}\}$, and $k = \sup \{i : \zeta_i^{(n)} > 0\}$. If $j \geq k$, the n -th solution is threshold-based, and the procedure is over. Otherwise, we construct a solution $\zeta^{(n+1)}$, with $\zeta_i^{(n+1)} = \zeta_i^{(n)} \forall i \notin \{j, k\}$ and $p_s(\zeta^{(n+1)}) = p_s(\zeta^{(n)})$, such that $\zeta_j^{(n+1)} = \zeta_{\text{max}} \vee \zeta_k^{(n+1)} = 0$. This is always possible thanks to Lemma 2, and, due to the finite dimension of vector ζ , the sequence must converge to a threshold-based policy in a finite number of steps. Lemma 3 must then hold if every intermediate step improves the accuracy, i.e., if $\bar{\lambda}(\zeta^{(n+1)}) \geq \bar{\lambda}(\zeta^{(n)}) \forall n$. This can be proven to hold under the relatively mild condition of the hypothesis and after a long sequence of algebraic steps. In the following, we consider the case in which $p_e = 0$ for the sake of brevity, without loss of generality. The Lemma holds under the same conditions for other values of p_e , after a change of variable from ζ_i to $z_i = (1 - p_e)\zeta_i$. As this is a linear

transformation that preserves the sign, and $(1 - p_e)\zeta_{\max} \leq \zeta_{\max}$, all the following inequalities are true for $p_e > 0$.

We start by modeling all zones other than j and k as a combined zone c , and denote the probability of all nodes in combined zone c being silent as s_c , the probability of a single one transmitting as a_c , and the expected accuracy of packets received from the combined zone c is $\bar{\lambda}_c$, all of which stay constant in the water-filling transfer from k to j . The success probability $p_s(\zeta^{(n)})$ can then be expressed as

$$p_s(\zeta^{(n)}) = \left(1 - \zeta_j^{(n)}\right)^{m_j} \left(1 - \zeta_k^{(n)}\right)^{m_k} \left[a_c + s_c \left(\frac{\zeta_j^{(n)}}{1 - \zeta_j^{(n)}} + \frac{\zeta_k^{(n)}}{1 - \zeta_k^{(n)}} \right) \right]$$

and condition $p_s(\zeta^{(n+1)}) = p_s(\zeta^{(n)})$ becomes

$$\frac{\left(1 - \zeta_j^{(n)}\right)^{m_j} \left(1 - \zeta_k^{(n)}\right)^{m_k}}{\left(1 - \zeta_j^{(n+1)}\right)^{m_j} \left(1 - \zeta_k^{(n+1)}\right)^{m_k}} = \frac{a_c + s_c \left(\frac{m_j \zeta_j^{(n+1)}}{1 - \zeta_j^{(n+1)}} + \frac{m_k \zeta_k^{(n+1)}}{1 - \zeta_k^{(n+1)}} \right)}{a_c + s_c \left(\frac{m_j \zeta_j^{(n)}}{1 - \zeta_j^{(n)}} + \frac{m_k \zeta_k^{(n)}}{1 - \zeta_k^{(n)}} \right)}. \quad (26)$$

Similarly, the the overall expected accuracy $\bar{\lambda}(\zeta^{(n)})$ is

$$\bar{\lambda}(\zeta^{(n)}) = \frac{\left(1 - \zeta_j^{(n)}\right)^{m_j} \left(1 - \zeta_k^{(n)}\right)^{m_k}}{p_s(\zeta^{(n)})} \left[\bar{\lambda}_c a_c + s_c \left(\frac{m_j \zeta_j^{(n)} \lambda_j}{1 - \zeta_j^{(n)}} + \frac{m_k \zeta_k^{(n)} \lambda_k}{1 - \zeta_k^{(n)}} \right) \right]$$

and we have $\bar{\lambda}(\zeta^{(n+1)}) \geq \bar{\lambda}(\zeta^{(n)})$ if and only if

$$\frac{\left[\bar{\lambda}_c a_c + s_c \left(\frac{m_j \zeta_j^{(n+1)} \lambda_j}{1 - \zeta_j^{(n+1)}} + \frac{m_k \zeta_k^{(n+1)} \lambda_k}{1 - \zeta_k^{(n+1)}} \right) \right]}{\left(1 - \zeta_j^{(n)}\right)^{m_j} \left(1 - \zeta_k^{(n)}\right)^{m_k} p_s(\zeta^{(n+1)})} \geq \frac{\left[\bar{\lambda}_c a_c + s_c \left(\frac{m_j \zeta_j^{(n)} \lambda_j}{1 - \zeta_j^{(n)}} + \frac{m_k \zeta_k^{(n)} \lambda_k}{1 - \zeta_k^{(n)}} \right) \right]}{\left(1 - \zeta_j^{(n+1)}\right)^{m_j} \left(1 - \zeta_k^{(n+1)}\right)^{m_k} p_s(\zeta^{(n)})}. \quad (27)$$

For the sake of compactness, we define auxiliary symbols $\xi_j = m_j \left(\frac{\zeta_j^{(n+1)}}{1 - \zeta_j^{(n+1)}} - \frac{\zeta_j^{(n)}}{1 - \zeta_j^{(n)}} \right)$ and $\xi_k = m_k \left(\frac{\zeta_k^{(n)}}{1 - \zeta_k^{(n)}} - \frac{\zeta_k^{(n+1)}}{1 - \zeta_k^{(n+1)}} \right)$. Due to (26), (27) is equivalent to

$$\frac{a_c}{s_c} ((\bar{\lambda}_c - \lambda_k)\xi_k - (\bar{\lambda}_c - \lambda_j)\xi_j) + (\lambda_j - \lambda_k)\xi_j\xi_k \geq 0.$$

We then introduce an additional Lemma:

Lemma 4. *Under the condition in (23), we have*

$$\xi_j(\lambda_j - \lambda_k) \geq (\xi_j - \xi_k)(\bar{\lambda}_c - \lambda_k). \quad (28)$$

Proof: The proof of the Lemma is provided in Appendix E. ■

We then note that (28) represents a more restrictive condition, so that (27) always holds if (28) is verified. Lemma 4 thus proves that each water-filling transfer improves the expected accuracy of received packets. We thus proved that any individual step in the sequence has

$p_s(\zeta^{(n+1)}) = p_s(\zeta^{(n)})$ and $\bar{\lambda}(\zeta^{(n+1)}) \geq \bar{\lambda}(\zeta^{(n)})$. As the sequence must reach a threshold-based policy, the Lemma is proven.

E. Proof of Lemma 4

We first note that, if we have $j = 0$, i.e., the innermost zone does not have a transmission probability $\zeta_{\max}$, while external zones have a non-zero transmission probability, the condition is trivially true, as $\lambda_j \geq \bar{\lambda}_c$. We must then handle the two cases with $j = 1$ and $j > 1$ separately, but first, we introduce another Lemma to simplify the condition in (28).

Lemma 5. *Considering the n -th step in the sequence, the ratio between the two values $\xi_k = m_k \left(\frac{\zeta_k^{(n)}}{1-\zeta_k^{(n)}} - \frac{\zeta_k^{(n+1)}}{1-\zeta_k^{(n+1)}} \right)$ and $\xi_j = m_j \left(\frac{\zeta_j^{(n+1)}}{1-\zeta_j^{(n+1)}} - \frac{\zeta_j^{(n)}}{1-\zeta_j^{(n)}} \right)$ must respect the lower bound*

$$\frac{\xi_k}{\xi_j} \geq 1 - \zeta_{\max}. \quad (29)$$

Proof: The proof of the Lemma is provided in Appendix F. ■

In the case in which $j = 1$, we consider the loosest possible bound, i.e., $\bar{\lambda}_c \leq 1$. We then substitute the bound in (29) into (28) to get

$$\lambda_j \geq (1 - \zeta_{\max})\lambda_k + \zeta_{\max}.$$

Naturally, the worst case is $k = 2$, as it represents the highest possible value of λ_k . We substitute the definition from (4) to get

$$\frac{1}{(1+R)^\alpha} \geq \frac{1 - \zeta_{\max}}{(1+2R)^\alpha} + \zeta_{\max},$$

which is identical to the condition in (23).

We now consider the case in which $j > 1$, where we can exploit the structure of λ_i that we defined in (4) to obtain a tighter upper bound on $\bar{\lambda}_c$. We first note that c contains zones with indices $\{0, \dots, j-1\} \cup \{j+1, \dots, k-1\}$. The former set has $\lambda_i > \lambda_j$, while the latter has $\lambda_i < \lambda_j$, as λ is a strictly decreasing function. We then note that considering only zones $\{0, \dots, j-1\}$ yields an upper bound on $\bar{\lambda}_c$. As all these zones have $\zeta_i^{(n)} = \zeta_{\max}$ by the definition of j , we know that

$$\bar{\lambda}_c \leq \sum_{i=0}^{j-1} \frac{m_i \lambda_i}{\sum_{\ell=0}^{j-1} m_\ell} = \sum_{i=0}^{j-1} \frac{2i+1}{(1+iR)^\alpha j^2}.$$

where we consider the expected number of nodes $\mathbb{E}[m_i] = \frac{(2i+1)m}{K^2}$ and $\lambda_i = (1+iR)^{-\alpha}$ by applying the definitions in (3) and (4). Under our model, the actual values of m_j and m_k are

random, as nodes are uniformly distributed across the area. However, we use the expected value as a proxy for the sake of readability. This does not affect the rest of the proof, as the linearity of the expected value allows us to extend the result directly. We note that $\frac{\partial^2 \lambda_i}{\partial i^2}$ is always positive for $\alpha \geq 0$, and λ_i is thus convex in i . By the definition of convexity, we get

$$\lambda_i \leq \frac{(j-i)\lambda_0 + i\lambda_j}{j},$$

and $\lambda_0 = 1$ by the definition in (4). We then get a looser bound on $\bar{\lambda}_c$, as all values are positive:

$$\bar{\lambda}_c \leq \sum_{i=0}^{j-1} \frac{(2i+1)(j-i(1-\lambda_j))}{j^2} = \frac{2j^2+3j+1}{6j^2} + \frac{(2j^2-3j+1)\lambda_j}{6j^2}.$$

We then substitute the bound into (28):

$$\lambda_j [\xi_j(4j^2+3j-1) + \xi_k(2j^2-3j+1)] \geq [\xi_j - \xi_k(1-\lambda_k)](2j^2+3j+1).$$

As $\lambda_k \leq \lambda_j$, we can further restrict ourselves to

$$\lambda_j [\xi_j(4j^2+3j-1) - 6\xi_k j] \geq (\xi_j - \xi_k)(2j^2+3j+1) = \left(1 - \frac{\xi_k}{\xi_j}\right) \frac{(2j^2+3j+1)}{4j^2+3j-1 - \frac{6\xi_k j}{\xi_j}}. \quad (30)$$

If we take the derivative of the right side of the inequality with respect to $\frac{\xi_k}{\xi_j}$, we get

$$\frac{\partial \left(1 - \frac{\xi_k}{\xi_j}\right) \frac{(2j^2+3j+1)}{4j^2+3j-1 - \frac{6\xi_k j}{\xi_j}}}{\partial \frac{\xi_k}{\xi_j}} = - \frac{8j^4 + 6j^3 - 7j^2 - 6j - 1}{\left(4j^2 + 3j - 1 - \frac{6\xi_k j}{\xi_j}\right)^2}$$

which is always negative for $j > 1$. We then note that setting $\xi_k/\xi_j \geq 1 - \zeta_{\max}$ makes the condition in (30) always true for $\zeta_{\max} > \frac{1}{2}$.

F. Proof of Lemma 5

We first introduce two parameters, $\varphi = \zeta_j^{(n+1)} - \zeta_j^{(n)}$ and $\psi = \left(\zeta_k^{(n)} - \zeta_k^{(n+1)}\right) \varphi^{-1}$. By the definition of the swap operation between zones j and k , we must have $\varphi \in (0, \zeta_{\max} - \zeta_j^{(n)})$ and $\psi \in \left(0, \zeta_k^{(n)} \varphi^{-1}\right]$. Substituting the two values into (26), and using the relations

$$\begin{aligned} \frac{m_j \left(\zeta_j^{(n)} + \varphi\right)}{1 - \zeta_j^{(n)} - \varphi} &= \frac{m_j \zeta_j^{(n)}}{1 - \zeta_j^{(n)}} + \frac{m_j \varphi}{\left(1 - \zeta_j^{(n)}\right) \left(1 - \zeta_j^{(n)} - \varphi\right)}; \\ \frac{m_k \left(\zeta_k^{(n)} - \psi \varphi\right)}{1 - \zeta_k^{(n)} + \psi \varphi} &= \frac{m_k \zeta_k^{(n)}}{1 - \zeta_k^{(n)}} - \frac{m_k \psi \varphi}{\left(1 - \zeta_k^{(n)}\right) \left(1 - \zeta_k^{(n)} + \psi \varphi\right)} \end{aligned}$$

the condition can be transformed after some manipulations into

$$A(\varphi) \left[1 + \left(\frac{B\varphi m_j}{(1 - \zeta_j^{(n)}) (1 - \zeta_j^{(n)} - \varphi)} - \frac{B\psi\varphi m_k}{(1 - \zeta_k^{(n)}) (1 - \zeta_k^{(n)} + \psi\varphi)} \right) \right] = A(0) \quad (31)$$

where we introduced utility symbols $A(\varphi) = (1 - \zeta_j^{(n)} - \varphi)^{m_j} (1 - \zeta_k^{(n)} + \psi\varphi)^{m_k}$ and $B = s_c \left(a_c + s_c \left(\frac{m_j \zeta_j^{(n)}}{1 - \zeta_j^{(n)}} + \frac{m_k \zeta_k^{(n)}}{1 - \zeta_k^{(n)}} \right) \right)^{-1}$.

We first note that the condition in (31) is trivially true if $\varphi = 0$, as the two solutions are identical. We then compute the condition for the derivative of the left side of (31) with respect to φ to be positive:

$$\begin{aligned} A(\varphi)B & \left(\frac{m_j}{(1 - \zeta_j^{(n)} - \varphi)^2} - \frac{m_k\psi}{(1 - \zeta_k^{(n)} + \psi\varphi)^2} \right) + \left(\frac{A(\varphi)m_k\psi}{1 - \zeta_k^{(n)} + \psi\varphi} - \frac{A(\varphi)m_j}{1 - \zeta_j^{(n)} - \varphi} \right) \\ & \times \left[1 + \left(\frac{B\varphi m_j}{(1 - \zeta_j^{(n)}) (1 - \zeta_j^{(n)} - \varphi)} - \frac{B\psi\varphi m_k}{(1 - \zeta_k^{(n)}) (1 - \zeta_k^{(n)} + \psi\varphi)} \right) \right] \geq 0. \end{aligned}$$

We then define value ψ^* as follows:

$$\psi^* = \frac{m_j (1 - \zeta_k^{(n)})}{m_k (1 - \zeta_j^{(n)} - \varphi) - m_j \varphi}.$$

If we set $\psi = \psi^*$, $p_s(\zeta^{(n+1)}) = p_s(\zeta^{(n)})$ for $\varphi = 0$ and $p_s(\zeta^{(n+1)}) \geq p_s(\zeta^{(n)})$ for $\varphi > 0$. As solution $\zeta^{(n)}$ is not threshold-based, we must have $\varphi > 0$. We then remark that the system clearly has to operate at load lower than 1, and thus the success probability must be a strictly decreasing function of ψ , which controls the reduction of the transmission probability from zone k . This means that, in order to have $p_s(\zeta^{(n+1)}) = p_s(\zeta^{(n)})$, we must have $\psi \geq \psi^*$. We then compute the value of ξ_k/ξ_j :

$$\frac{\xi_k}{\xi_j} = \frac{m_k \psi \varphi (1 - \zeta_j^{(n)}) (1 - \zeta_j^{(n)} - \varphi)}{m_j \varphi (1 - \zeta_k^{(n)}) (1 - \zeta_k^{(n)} + \psi\varphi)}.$$

As the value is a strictly increasing function of ψ , we can substitute ψ^* to obtain a lower bound. After some algebraic manipulations, we get:

$$\frac{\xi_k}{\xi_j} \geq \frac{1 - \zeta_j^{(n)}}{1 - \zeta_k^{(n)}} \geq 1 - \zeta_{\max}.$$

REFERENCES

- [1] A. Munari, F. Chiariotti, L. Badia, and P. Popovski, “Spatio-temporal information freshness for remote source monitoring in IoT systems,” in *Proc. IEEE SPAWC*, July 2025.
- [2] Ericsson, “Mobility report – November 2024,” Ericsson, Tech. Rep., 2024.
- [3] A. E. Kalør, G. Durisi, S. Coleri, S. Parkvall, W. Yu, A. Mueller, and P. Popovski, “Wireless 6G connectivity for massive number of devices and critical services,” *Proc. IEEE*, 2024.
- [4] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. Leung, “Enabling massive IoT toward 6G: A comprehensive survey,” *IEEE Internet Things J.*, vol. 8, no. 15, pp. 11 891–11 915, 2021.
- [5] D. Gündüz, F. Chiariotti, K. Huang, A. E. Kalør, S. Kobus, and P. Popovski, “Timely and massive communication in 6G: Pragmatics, learning, and inference,” *IEEE BITS Inf. Theory Mag.*, vol. 3, no. 1, pp. 27–40, 2023.
- [6] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, “Minimizing age of information in vehicular networks,” in *Proc. IEEE SECON*, June 2011.
- [7] R. D. Yates and S. K. Kaul, “The age of information: Real-time status updating by multiple sources,” *IEEE Trans. Inf. Theory*, vol. 65, no. 3, 2019.
- [8] R. Talak, S. Karaman, and E. Modiano, “Distributed scheduling algorithms for optimizing information freshness in wireless networks,” in *Proc. IEEE SPAWC*, June 2018.
- [9] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, “The age of incorrect information: a new performance metric for status updates,” *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, 2020.
- [10] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, “Age-of-information vs. value-of-information scheduling for cellular networked control systems,” in *Proc. ACM/IEEE ICCPS*, April 2019, pp. 109–117.
- [11] H. H. Yang, A. Arafa, T. Q. Quek, and H. V. Poor, “Spatiotemporal analysis for age of information in random access networks under last-come first-serve with replacement protocol,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2813–2829, 2021.
- [12] J. Tong, L. Fu, and Z. Han, “Age-of-information oriented scheduling for multichannel IoT systems with correlated sources,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9775–9790, 2022.
- [13] A. Zancanaro, G. Cusotto, and L. Badia, “Modeling value of information in remote sensing from correlated sources,” *Comput. Commun.*, vol. 203, pp. 289–297, 2023.
- [14] M. Fidler, F. Gallistl, J. P. Champati, and J. Widmer, “2D-AoI: Age-of-information of distributed sensors for spatio-temporal processes,” *IEEE Trans. Commun.*, vol. 74, pp. 645–661, 2026.
- [15] J. Luo and D. Guo, “On the entropy rate of hidden Markov processes observed through arbitrary memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1460–1467, 2009.
- [16] G. Cocco, A. Munari, and G. Liva, “Remote monitoring of two-state Markov sources via random access channels: An information freshness vs. state estimation entropy perspective,” *IEEE J. Sel. Areas Inf. Theory*, vol. 4, pp. 651–666, 2023.
- [17] G. N. Saridis, “Entropy formulation of optimal and adaptive control,” *IEEE Trans. Autom. Control*, vol. 33, no. 8, pp. 713–721, 2002.
- [18] B. Chen, L. Dang, Y. Gu, N. Zheng, and J. C. Principe, “Minimum error entropy Kalman filter,” *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 9, pp. 5819–5829, 2019.
- [19] Y. Chen and A. Ephremides, “Preempting to minimize age of incorrect information under transmission delay,” *IEEE Trans. Netw.*, vol. 34, pp. 668–680, 2026.
- [20] I. Cosandal, N. Akar, and S. Ulukus, “Multi-threshold aoii-optimum sampling policies for continuous-time markov chain information sources,” *IEEE Trans. Inf. Theory*, vol. 71, no. 9, pp. 6968–6988, 2025.
- [21] I. Cosandal, S. Ulukus, and N. Akar, “Joint age-state belief is all you need: Minimizing AoII via pull-based remote estimation,” in *Proc. IEEE ICC Workshops*, June 2025, pp. 1098–1103.

- [22] J. Luo and N. Pappas, “Minimizing the age of missed and false alarms in remote estimation of Markov sources,” in *Proc. 25th ACM MOBIHOC*, October 2024, pp. 381–386.
- [23] M. Salimnejad, M. Kountouris, and N. Pappas, “State-aware real-time tracking and remote reconstruction of a Markov source,” *J. Commun. Netw.*, vol. 25, no. 5, pp. 657–669, 2023.
- [24] S. Liyanaarachchi, S. Ulukus, and N. Akar, “Structured estimators: A new perspective on information freshness,” in *Proc. IEEE ITW*, Sep. 2025.
- [25] P. Talli, E. D. Santi, F. Chiariotti, T. Soleymani, F. Mason, A. Zanella, and D. Gündüz, “Pragmatic communication for remote control of finite-state Markov processes,” *IEEE J. Sel. Areas Commun.*, vol. 43, no. 7, pp. 2589–2603, 2025.
- [26] J. Holm, F. Chiariotti, A. E. Kalør, B. Soret, T. B. Pedersen, and P. Popovski, “Goal-oriented scheduling in sensor networks with application timing awareness,” *IEEE Trans. Commun.*, vol. 71, no. 8, pp. 4513–4527, 2023.
- [27] A. Zakeri, M. Moltafet, and M. Codreanu, “Semantic-aware sampling and transmission in real-time tracking systems: A POMDP approach,” *IEEE Trans. Commun.*, vol. 73, no. 7, pp. 4898–4913, 2025.
- [28] A. Munari, “On the uncertainty of a simple estimator for remote source monitoring over ALOHA channels,” in *Proc. IEEE ISIT*, June 2025.
- [29] M. Rezaeian, “Sensor scheduling for optimal observability using estimation entropy,” in *Proc. IEEE PERCOM Workshops*, March 2007.
- [30] —, “Hidden Markov process: A new representation, entropy rate and estimation entropy,” *arXiv preprint cs/0606114*, 2006.
- [31] G. Chen, S.-C. Liew, and Y. Shao, “Uncertainty-of-information scheduling: A restless multiarmed bandit framework,” *IEEE Trans. Inf. Theory*, vol. 68, no. 9, pp. 6151–6173, 2022.
- [32] G. Chen and S.-C. Liew, “An index policy for minimizing the uncertainty-of-information of Markov sources,” *IEEE Trans. Inf. Theory*, vol. 70, no. 1, pp. 698–721, 2024.
- [33] N. L. Johnson and S. Kotz, *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*. New York, NY, USA: John Wiley & Sons, 1977.
- [34] J. Hribar, M. Costa, N. Kaminski, and L. A. DaSilva, “Using correlated information to extend device lifetime,” *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2439–2448, 2018.
- [35] N. Abramson, “The throughput of packet broadcasting channels,” *IEEE Trans. Commun.*, vol. COM-25, no. 1, pp. 117–128, 1977.
- [36] E. Perron, M. Rezaeian, and A. Grant, “The On-Off Fading Channel,” in *Proc. IEEE ISIT*, June 2003.
- [37] Z. Sun, Y. Xie, J. Yuan, and T. Yang, “Coded Slotted ALOHA Schemes for Erasure Channels,” in *Proc. IEEE ICC Workshops*, May 2016.
- [38] M. Ivanov, F. Brännström, A. Graell i Amat, and P. Popovski, “Broadcast coded slotted ALOHA: A finite frame length analysis,” *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 651–662, 2017.
- [39] J.-H. Lee, C. Park, S. Park, and A. F. Molisch, “Handover protocol learning for LEO satellite networks: Access delay and collision minimization,” *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7624–7637, 2024.
- [40] G. Callebaut and L. Van der Perre, “Characterization of LoRa point-to-point path loss: Measurement campaigns and modeling considering censored data,” *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1910–1918, 2020.
- [41] L. Beltramelli, A. Mahmood, P. Österberg, and M. Gidlund, “LoRa beyond ALOHA: An investigation of alternative random access protocols,” *IEEE Trans. Industrial Informatics*, vol. 17, no. 5, pp. 3544–3554, 2020.
- [42] O. T. Yavascan and E. Uysal, “Analysis of slotted ALOHA with an age threshold,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1456–1470, 2021.
- [43] E. Testi and E. Paolini, “Packet collision probability of direct-to-satellite iot systems,” *IEEE Internet Things J.*, vol. 12, no. 2, pp. 1843–1855, 2025.
- [44] H. Asgari, A. Munari, G. Liva, and G. Cocco, “Remote monitoring of two-state Markov sources in random access channels: Joint model and state estimation,” in *Proc. ITG SCC*, March 2025.

- [45] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New York: Wiley, 2006.
- [46] V. A. Kovalevsky, Читающие автоматы и распознавание образов [*Character readers and pattern recognition*]. Kyiv, Ukraine: Naukova Dumka, 1968.
- [47] M. Feder and N. Merhav, “Relations between entropy and error probability,” *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 259–266, 1994.
- [48] R. D. Yates and S. K. Kaul, “Status updates over unreliable multiaccess channels,” in *Proc. IEEE ISIT*, June 2017, pp. 331–335.
- [49] R. Yates and S. Kaul, “Status updates over unreliable multiaccess channels,” in *Proc. IEEE ISIT*, 2017.
- [50] A. Munari, “Modern random access: an age of information perspective on irregular repetition slotted ALOHA,” *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3572–3585, Jun. 2021.
- [51] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [52] H. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, 3rd ed. San Diego, CA, USA: Academic Press, 1998.