Received 13 January 2024; accepted 29 January 2024. Date of publication 8 February 2024; date of current version 23 February 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3363731

Semantics-Aware Active Fault Detection in Status Updating Systems

GEORGE STAMATAKIS [1], NIKOLAOS PAPPAS [2] (Senior Member, IEEE), ALEXANDROS FRAGKIADAKIS [1], NIKOLAOS PETROULAKIS [1] (Member, IEEE), AND APOSTOLOS TRAGANITIS [1]

[1] Institute of Computer Science, Foundation for Research and Technology, 700 13 Heraklion, Greece

[2] Department of Computer and Information Science, Linköping University, Campus Valla, 581 83 Linköping, Sweden

CORRESPONDING AUTHOR: G. STAMATAKIS (e-mail: gstam@ics.forth.gr)

This work was supported in part by the European Union's Horizon Europe Research and Innovation Program under Grant 101057527 (NextGEM). The work of Nikolaos Pappas was supported in part by the Swedish Research Council (VR), ELLIIT, the European Union (ETHER) under Grant 101096526; in part by the European Union's Horizon Europe Research and Innovation Programme under the Marie Skłodowska-Curie Grant under Agreement 101131481 (SOVEREIGN); and in part by the Horizon Europe/JU SNS Project ROBUST-6G under Grant 101139068.

A shorter version of this work has been published in [1] [DOI: 10.23919/WiOpt56218.2022.9930549].

ABSTRACT With its growing number of deployed devices and applications, the Internet of Things (IoT) raises significant challenges for network maintenance procedures. In this work, we address a problem of active fault detection in an IoT scenario, whereby a monitor can probe a remote device to acquire fresh information and facilitate fault detection. However, probing could significantly impact the system's energy and communication resources. To this end, we utilize the Age of Information to measure the freshness of information at the monitor and adopt a semantics-aware communication approach between the monitor and the remote device. In semantics-aware communications, the processes of generating and transmitting information are treated jointly to consider the importance of information and the purpose of communication. We formulate the problem as a Partially Observable Markov Decision Process and show analytically that the optimal policy is of a threshold type. Finally, we use a computationally efficient stochastic approximation algorithm to approximate the optimal policy and present numerical results that exhibit the advantage of our approach compared to a conventional delay-based probing policy.

INDEX TERMS Semantics of information, age of information, fault detection, status updating systems.

I. INTRODUCTION

THE EMERGENCE of massive IoT ecosystems poses new challenges for their maintenance procedures. IoT networks are characterized by software, hardware, and communication protocols diversity. Furthermore, they are typically comprised of a large number of devices that are often deployed in remote and harsh environments. In this context, developing autonomous fault detection procedures is necessary to safely and efficiently operate an IoT network. Autonomous fault detection procedures are able to identify faulty, i.e., abnormal behavior, in IoT networks without requiring the direct inspection of the system by humans and can be utilized to autonomously perform maintenance of the system so as to prevent outages. Faults in IoT networks can be categorized into two distinct classes based on whether their mitigation requires the physical intervention by humans or not. Autonomous maintenance targets the latter class of faults which can usually be mitigated by remote maintenance of the IoT devices. Examples of such remote maintenance procedures include the rebooting the IoT device's operating system, the update of a device's firmware or an application's software, restoring configurations and network engineering. The majority of fault detection algorithms that have been proposed in the past [2], [3] fall under the category of passive fault detection mechanisms, i.e., they assume that the system is passively monitored and utilize statistical or machine learning techniques to infer the health status of its subsystems. However, a major drawback with passive fault detection is that faults can pass undetected if the faulty and the nominal operation overlap due to measurement and process uncertainties or in cases where control actions mask the influence of faults [4]. Furthermore, the data necessary to make informed maintenance decisions often arrive late at the passive fault detection mechanism so that the timely maintenance of the network is difficult.

To address this problem, we use an *active* fault detection scheme that utilizes probes to affect the system's response and thus increase the probability of detecting certain faults. With active fault detection, special care must be taken so that the extra network traffic due to probing is not detrimental to the system's performance. This requirement can become a significant challenge if active fault detection is to be deployed in IoT networks with a large number of remote devices. Blindly generating and transmitting probes could increase network congestion and prohibit other applications from satisfying their possibly strict real-time constraints. To this end, we adopt a *semantics*-aware [5], [6], [7] approach to active fault detection. Within the context of *semantics*-aware communications, the generation and transfer of information across a network are considered jointly to take into account the goal or purpose of the communication. What is more, the *importance/significance* of a communication event, i.e., the event of generation and transmission of information, constitutes the decisive criterion of whether it should take place or not. The definition of the importance of a communication event is application-specific; thus, in the context of active fault detection, we define it to be a function of the freshness of information that has been received from the remote device and of the operational status of the communication network and the remote device. Simply put, the importance of a probe increases when the information sent by the remote device has become stale, and the probing entity is not confident that the operational status of the communication network and the remote device is good. Numerical results indicate that the semantics-aware approach offers significant advantages in contrast to the classical communication paradigm, where information generation and its transmission are treated separately.

More specifically, in this work, we consider a basic active fault detection scenario for a discrete-time dynamic system comprising a sensor and a monitor. At the beginning of each time slot, the sensor probabilistically generates and transmits status updates to the monitor over an unreliable link. In contrast, the monitor decides whether or not to probe the sensor through a separate unreliable link. A successfully received probe results in a mandatory transmission of a fresh status update from the sensor. By the end of each time slot, the monitor may or may not receive a status update either because none was generated at the sensor or due to intermittent faults at the sensor and the wireless links. To detect intermittent faults, the monitor maintains a belief vector, i.e., a probability distribution, over the system's operational status (healthy or faulty) and a measure of its confidence in this belief vector that is expressed by the entropy of the belief vector. Probing, successfully or unsuccessfully, increases the confidence of the monitor in its belief state; however, it also induces a cost for the monitor that represents the negative impact of probing on the system's energy and communication resources. Our objective is to find a policy that decides at each time slot whether or not a probe should be sent to the sensor to optimally balance the probing cost with the need for fresh information at the monitor.

Our approach to solving this problem is to formulate it as a Partially Observable Markov Decision Process (POMDP) and derive the necessary conditions for probing to reduce the entropy of the belief state's vector. To the best of our knowledge, this work and its shorter version [1], are the only works on active fault detection to adopt this approach. Our analysis indicates that probing cost values exist such that the optimal policy is of a *threshold* type. In addition, we present a stochastic approximation algorithm that can compute such a policy and, subsequently, evaluate the derived policy numerically. The contributions of this work, especially in comparison to [1], are:

- We derive analytical expressions for the probability of the monitor to receive a fresh status update.
- We elaborate on the total cost function of the POMDP and the rational behind the belief state formulation.
- We give analytical proofs for the Lemmas and the Theorem presented in [1].
- We elaborate on the analysis of the results produced by the numerical experiments and provide intuitive explanations for the behavior of each tested policy.
- We introduce a new basic scenario that is different from the basic scenario we used in [1] and by comparing their numerical results we exhibit the effect that the health status entropy transition cost has on the total expected cost in cases where the monitor has a belief state vector with low entropy despite the fact that it doesn't receive status updates.

A. RELATED WORK

Fault detection methods can be categorized as passive, reactive, proactive, and active. Passive fault detection methods collect information from the data packets that the wireless sensors exchange as part of their normal operation. In contrast, in reactive and proactive fault detection methods, the wireless sensors collect information about their operational status and transmit it to the monitor. Finally, the monitor probes the wireless sensors for information specific to the fault detection process in active fault detection methods. In Wireless Sensor Networks (WSNs), the fault detection algorithms being cited in recent works and surveys [3], [8], [9], [10] fall in the passive, reactive, and proactive categories, with the majority of them being passive fault detection algorithms. Compared to the other three categories, active fault detection methods for WSNs have received limited attention [9]. In [11] and [12], the authors adopted an active approach to fault detection in WSNs. However, both tools were meant for pre-deployment testing of WSNs software rather than a health status monitoring mechanism.

Unlike these works, we propose an active fault detection method for assessing the health status of sensors. We believe that autonomous active fault detection methods can successfully complement the passive ones by addressing their limitations. More specifically, passive fault detection methods often fail to detect faults because the faulty and the nominal operation overlap due to measurement and process uncertainties. What is more, network control mechanisms specifically designed to increase the robustness of the IoT network, e.g., by delegating the job of a sensor to neighboring or redundant nodes, often compensate for the performance degradation due to intermittent faults and thus mask their influence, rendering them undetectable [4].

The main motivation for active fault detection is the timely and reliable detection of faults that can not be diagnosed through passive fault detection algorithms [13]. Active fault detection algorithms make timely fault detection possible by proactively investigating the system's fault status, i.e., they don't passively wait for sufficient data to be collected and processed by a machine learning algorithm in hope of detecting a fault at some point in time. Furthermore, active fault detection algorithms systematically account for uncertainty and operational constraints. Accounting for uncertainty is particularly important when dealing with faults that are small in magnitude or develop slowly. The effects of these faults may be indistinguishable from the variation arising from system uncertainty and disturbances; as a result they can become challenging to diagnose based on the nominal input–output data typically used in passive fault detection methods. By injecting probes or other types of test signals into the system one can enable the detection of small-magnitude faults before their severity increases. Another important feature of active fault detection is the ability to account for system constraints while injecting probes. Physical limits on capacity, energy and actuation are common reasons for specifying constraints. Other reasons include ensuring that the probes do not compromise operational safety or lead to unacceptable reduction in performance while investigating the fault status.

In this work, we acknowledge the fact that the network overhead due to active fault detection can be prohibitive, and to address this problem we adopted the semantics-aware communication paradigm [5], [6], [7], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25] which has exhibited its ability to eliminate the transmission of redundant and uninformative data and thus minimize the induced overhead. More specifically, in this work we adopted the semantic communications perspective of goal-oriented communications as described in [26]. However, a different line of research on semantic communications, which focuses on semantic interoperability [26], may also have a profound impact in the economy of communication and, thus, could be utilized in active fault detection methods. Along this line of research, the authors of [27] developed a unified deep learning-enabled semantic communication system where a unified end-to-end framework can serve different tasks with multiple data modalities, that dynamically adjusts the number of transmitted symbols for different tasks and can adaptively adjust the number of transmitted features under different channel conditions in order to optimize transmission efficiency. Furthermore, the authors in [28] propose a semantic aware communication scheme that uses a pre-trained language model to quantify the semantic importance of data frames and allocate transmission power accordingly.

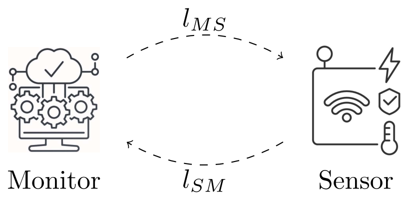


FIGURE 1. Basic IoT setup.

II. SYSTEM MODEL

We consider the system presented in Figure 1. It is comprised of a sensor that transmits status updates to a monitoring device over the wireless link labeled l_{SM}. The monitoring device, besides receiving the status updates from the sensor, is able to probe the sensor for a fresh status update over the wireless link labeled l_{MS}. Transmissions over the links l_{MS} and l_{SM} are subject to failure. Failures are independent between the two links. We assume that time is slotted and indexed by $t \in \mathbb{Z}^+$.

The state of the sensor is modeled as an independent two-state time-homogeneous Markov process. Let $F_t^S \in \{0, 1\}$ be the state of the sensor's Markov process at the beginning of the t-th time slot. When F_t^S has a value of 0/1, the sensor's operational status is healthy/faulty. We assume that the sensor will remain in the same state for the duration of a time slot and, afterward, it will make a probabilistic transition to another state as dictated by the state transition probability matrix P^S. Furthermore, at the beginning of each time slot, the sensor will generate a status update with probability P_g when in a healthy state, while it will not generate a status update when in a faulty state. In this work, we assume that $P_g < 1$; otherwise, the probing system is redundant. In the case of a status update generation, the sensor will transmit it over the l_{SM} link. At the end of the time slot, the status update is discarded independently from the outcome of the transmission.

Similarly, we model the health status of the wireless links as two independent two-state time-homogeneous Markov processes. Let $F_t^{MS}, F_t^{SM} \in \{0, 1\}$ denote the state of the independent Markov processes for the l_{MS} and l_{SM} links, respectively, at the beginning of the t-th time slot. When F_t^{MS} and F_t^{SM} take a value of 0/1, the operation of the wireless links is healthy/faulty. We assume that the wireless links will remain in the same state for the duration of a single time slot and, subsequently, they will make a transition to another state as dictated by the transition probability matrices P^{MS} and P^{SM} respectively. When in a healthy state, the wireless link will successfully forward a status update to the monitor with probability 1. In contrast, a faulty wireless link will always fail to deliver the status update.

In this work, we consider the problem of a monitoring agent that must optimally decide whether to probe the sensor at the beginning of each time slot. As a result of its decision and the system's dynamics, a transition cost is induced on the agent by the end of each time slot. The agent aims to decide optimally so as to minimize the transition costs accumulated over a finite time horizon. Each transition cost is a function of how well-informed the agent is about the joint health status of the sensor and the l_{SM} link, the freshness of the status updates it has received up to that time slot, and a cost value c associated with the probing action.

More specifically, since the agent cannot observe the actual health status of the sensor and the l_{SM} link, it maintains a belief vector over their joint health status, i.e., a probability distribution over the following two events: (i) the sensor and the l_{SM} link are in a healthy state, and (ii) at least one of them is in a faulty state. This health status belief vector is updated at each time slot based on the agent's action; and the observation, or not, of a fresh status update arrival. Given the health status belief vector, the agent needs a measure of how well-informed it is about the health status of the system. Since the health status belief vector is a probability distribution with only two values we could use one of them to characterize how well-informed the agent is. However, this approach will not generalize to belief state vectors with multiple values. To this end, we introduce the entropy of the belief state vector, denoted with H_t, as a measure of how well-informed the agent is about the health status of the system. Details regarding the definition of the health status belief vector and its entropy H_t are presented in Section III-C.

To characterize the freshness of the status updates received at the monitor, we utilize the Age of Information (AoI) metric that has received significant attention in the research community [29], [30], [31], [32], [33], [34], [35], [36], [37], [38]. AoI was defined in [39] as the time that has elapsed since the generation of the last status update that has been successfully decoded by the destination, i.e., $\Delta(t) = t - U(t)$, where $U(t)$ is the time-stamp of the last packet received at the destination at time t. We use Δ_t, $t = 0, 1, \ldots, N$, to denote the AoI of the sensor at time t. However, as the time horizon of the optimal probing problem increases, Δ_t could assume values that would be disproportionately larger than H_t. To alleviate this problem we will use the normalized value of the AoI which we define as, $\bar{\Delta}_t = \frac{\Delta_t}{N}$, where N is the length of the finite horizon measured in time slots. Finally, the VoI is defined as,

$$V_t = \lambda_1 H_t + \lambda_2 \bar{\Delta}_t, \tag{1}$$

where λ_1 and λ_2 are weights that determine the relative importance of each component of the metric.

The VoI defined above is a semantic metric in the sense used in goal-oriented communications [26]. It captures the importance that a fresh status update has for the agent to achieve its goal, i.e., to perform autonomous maintenance, and motivates the agent to take action, i.e., to probe the sensor, in order to reduce future transition costs due to high VoI. As an example, consider the case where the agent is not well-informed about the health status of the system, then the health status belief vector involves two equiprobable events, and H_t will assume its maximum value, i.e., one. This will motivate the agent to probe the sensor for a mandatory status update transmission which will reduce the entropy of the health status belief vector and along with it future costs due to H_t. Similarly, if the agent hasn't received a status update over a long period the agent is motivated to probe the sensor due to the high value of normalized AoI. Based on the outcome of the probing action either the AoI will be minimized or the entropy of the health status belief vector will be further reduced. If the agent decides to transmit a probe then a cost c is incurred, which represents the ratio of resources spend on probing to those spent on transmitting information, i.e., a status update. As an example, consider the case where the system presented in Figure 1 is energy-constrained. Now, let $\bar{E}_P$ denote the average energy spent on the transmission of a probe and $\bar{E}_S$ the average energy spent on the transmission of a status update, then we define $c = \bar{E}_P/\bar{E}_S$. In IoT networks, the sensors typically transmit small-sized packets toward the monitor and we expect probes to be small-sized packets as well, so that $\bar{E}_P \approx \bar{E}_S$ and $c \approx 1$. Finally, after the successful reception of a probe through the l_{MS} link, the sensor will generate a fresh status update at the next time slot with a probability of 1 if it is in a healthy state and with a probability 0 if it is in a faulty state.

III. PROBLEM FORMULATION

In this section, we formulate the decision problem presented above as a Partially Observable Markov Decision Process (POMDP) denoted with $\mathcal{P}$. A POMDP model with a finite horizon N is a 7-tuple $(\mathcal{S}, \mathcal{A}, Z, P, r, g, g_N)$, where $\mathcal{S}$ is a set of states, $\mathcal{A}$ is a set of actions, Z is a set of possible observations, P is a probability matrix representing the conditional transition probabilities between states, r represents the observation probabilities, g is the transition reward function and g_N is the terminal cost incurred at the last decision stage. In the remaining part of this section, we present the individual elements of $\mathcal{P}$ and formulate the corresponding dynamic program based on a belief state formulation [40].

A. STATE SPACE ($\mathcal{S}$)

At the beginning of the t-th time slot the health state of the system is represented by the column vector

$$s_t = \left[F_t^{MS}, F_t^{S}, F_t^{SM}\right]^T, \tag{2}$$

where, as described in Section II, $F_t^i \in \{0, 1\}, i \in \{MS, S, SM\}$ indicate the health status of the wireless links

$(l_{MS},\ l_{SM})$ and of the sensor S, while T is the transpose operator. By the definition of the system's state we have eight discrete states which are indexed by $i = 0, 1, \ldots, 7$. The true state of the system is unknown to the agent at time t.

B. ACTIONS ($\mathcal{A}$)

The set of actions available to the agent is denoted with $\mathcal{A} = \{0, 1\}$, where 0 represents the no-probe action and 1 indicates the probe action. The result of the probe action, given that the probe is successfully received through the l_{MS} link is that the sensor will generate a fresh status update at the next time slot w.p. 1 if it is in a healthy state, and w.p. 0 if it is in a faulty state. Both actions are available in all system states. Finally, we denote the action taken by the agent at the beginning of the t-th time slot with $a_t \in \mathcal{A}$. The action taken by the agent does not directly affect the state of the system, nevertheless, it affects the observation made by the agent.

C. RANDOM VARIABLES

The state of the system presented in Fig. 1 will change stochastically at the beginning of each time slot. The transition to the new state is governed by the transition probability matrices P^{MS}, P^{S} and P^{SM}, presented in Section II, and the state of the system during the previous time slot.

As mentioned above, the agent has no knowledge of the system's actual state and is limited to observing the arrival of status updates. The observations are stochastic in nature and are determined by the action taken by the agent, the state of the system, and the following random variables. The random variable $W_t^g \in \{0, 1\}$ represents the random event of a status update generation at the t-th time slot. If a status update is generated by the sensor, then W_t^g takes the value 1. If the sensor does not generate a status update, then W_t^g takes the value 0. We have the following conditional distribution for W_t^g

$$P\Big[W^g = 0|F^S, a\Big] = \begin{cases} 1 - P_g, & \text{if } a = 0 \text{ and } F^S = 0, \\ 1, & \text{if } F^S = 1, \\ 0, & \text{if } a = 1 \text{ and } F^S = 0, \end{cases} \tag{3}$$

and $P[W^g = 1|F^S, a] = 1 - P[W^g = 0|F^S, a]$, where we omitted the time index since the distribution is assumed to remain constant over time.

The random variable $W_t^{MS} \in \{0, 1\}$ represents the random event of a successful transmission of a probe over the l_{MS} link during the t-th time slot. A value of 0 indicates an unsuccessful transmission over the link, and a value of 1 indicates a successful transmission. The conditional probability distribution for W_t^{MS} is given by,

$$P\Big[W^{MS} = 0|F^{MS}, a\Big] = \begin{cases} 1, \text{ if } a = 0 \text{ or} \\ \qquad \big(a = 1 \text{ and } F^{MS} = 1\big), \\ 0, \text{ if } a = 1 \text{ and } F^{MS} = 0, \end{cases} \tag{4}$$

and $P[W^{MS} = 1|F^{MS}, a] = 1 - P[W^{MS} = 0|F^{MS}, a]$, where again we omitted the time index t. Finally, the random variable $W_t^{SM} \in \{0, 1\}$ represents the random event of a successful transmission over the l_{SM} link during the t-th time slot. A value of 0 indicates an unsuccessful transmission over the link, and a value of 1 indicates a successful transmission. The conditional probability distribution for W^{SM} is given by,

$$P\Big[W^{SM} = 0|W^g, F^{SM}\Big] = \begin{cases} 1, \text{ if } W^g = 0 \text{ or} \\ \qquad \big(W^g = 1 \text{ and } F^{SM} = 1\big), \\ 0, \text{ if } W^g = 1 \text{ and } F^{SM} = 0, \end{cases} \tag{5}$$

and

$$P\Big[W^{SM} = 1|W^g, F^{SM}\Big] = 1 - P\Big[W^{SM} = 0|W^g, F^{SM}\Big]. \tag{6}$$

D. TRANSITION PROBABILITIES (P)

Let m be an index over the set of the three subsystems presented in Fig. 1, i.e., $m \in \{MS, S, SM\}$, then the transition probability matrices P^{MS}, P^{S}, and P^{SM} can be defined as follows, $P^m = \begin{bmatrix} p_{00}^m & p_{01}^m \\ p_{10}^m & p_{11}^m \end{bmatrix}$, where p_{00}^m represents the probability to make a transition from a healthy state (0) to a healthy state (0) for subsystem m. Transition probabilities p_{01}^m, p_{10}^m, and p_{11}^m are defined similarly. Furthermore, we introduce the shorthand notation $s = [s_0, s_1, s_2]$ and $s' = [s'_0, s'_1, s'_2]$ for states $s_t = [F_t^{MS}, F_t^S, F_t^{SM}]^T$ and s_{t+1} respectively so that the conditional probability distribution of state s' given the current state s can be expressed as, $P[s_{t+1} = s'|s_t = s] = p^{MS}_{s_0 s'_0} \cdot p^{S}_{s_1 s'_1} \cdot p^{SM}_{s_2 s'_2}$.

E. OBSERVATIONS (Z)

At the beginning of each time slot, the agent observes whether a status update arrived or not. Let $z_t \in \{0, 1\}$ denote the observation made at the t-th time slot, with 0 representing the event that no status update was received and 1 representing the event that a status update was received. We define $r_s(a, z)$ as the probability to make observation z at the t-th time slot, i.e., $z_t = z$, given that the system is in state s, i.e., $s_t = s$, and the preceding action was a, i.e., $a_{t-1} = a$. Thus we have, $r_s(a, z) = P[z_t = z|s_t = s, a_{t-1} = a]$. The expressions of $r_s(a, z)$ for all combinations of states and actions are presented in Expressions (7), (8), (9), and (10), show at the bottom of the next page. By utilizing the conditional probability distributions presented in Section III-C, we derived the observation probabilities for all possible combinations of states and controls and presented them in Table 1.

The evolution of the AoI value over time depends on the observation made by the agent and,

$$\Delta_t = \begin{cases} 1, & \text{if } z_t = 1, \\ \min\{N, \Delta_t + 1\}, & \text{if } z_t = 0, \end{cases} \tag{11}$$

where N is the finite time horizon of the optimization problem.

F. TRANSITION COST FUNCTION (G, G_N)

At the end of each time slot, the agent is charged with a cost that depends on the VoI and the action taken by the

TABLE 1. Observation probabilities $r_s(a, z)$ as a function of the health status of the sensor (F^S), of the l_{MS} link (F^{MS}), of the l_{SM} link (F^{SM}), the action (a_{t-1}), and the observation z_t.

i	F_t^{MS}	F_t^S	F_t^{SM}	$a_{t-1}=0$, $z_t=0$	$a_{t-1}=0$, $z_t=1$	$a_{t-1}=1$, $z_t=0$	$a_{t-1}=1$, $z_t=1$
0	0	0	0	$1-P_g$	P_g	0	1
1	0	0	1	1	0	1	0
2	0	1	0	1	0	1	0
3	0	1	1	1	0	1	0
4	1	0	0	$1-P_g$	P_g	$1-P_g$	P_g
5	1	0	1	1	0	1	0
6	1	1	0	1	0	1	0
7	1	1	1	1	0	1	0

agent as follows, $g_t = c \cdot \mathbb{1}_{\{a_t=1\}} + V_t$, where, $\mathbb{1}_{\{a_t=1\}}$ is the indicator function which takes a value of 1 when the probe action was taken by the agent and a value of 0 otherwise; V_t is computed using Equation (1). Parameter c is a cost value associated with probing and quantifies the consumption of system resources for the generation and transmission of a probe.

The use VoI as a cost metric is justified by the fact that it quantifies how much the agent is in need of a fresh status update.

G. TOTAL COST FUNCTION

In a POMDP the agent doesn't have access to the current state of the system, thus, to optimally select actions it must utilize all previous observations and actions up to time t [40, Ch. 4]. Let $h_t = [z_0, z_1, \ldots, z_t, a_0, a_1, \ldots, a_{t-1}]$ be the *history* of all previous observations and actions, with $h_0 = \{z_0\}$. Furthermore, let $\mathcal{H}$ be the set of all possible histories for the system at hand. The agent must find a policy π^* that maps each history in $\mathcal{H}$ to a probability distribution over actions, i.e., $\pi : \mathcal{H} \to P(\mathcal{A})$, so that the expected value of the total cost accumulated over a horizon of N time slots is minimized. Let Π be the set of all feasible policies for the system at hand, then, assuming that the agent's policy is $\pi \in \Pi$ and has an initial history h_0 the expected value of the total cost accumulated over a horizon of N time slots is,

$$J_0(h_0) = \mathop{\mathbb{E}}_{W_0,\ldots,W_N}\left[\sum_{t=0}^{N} g_t \middle| h_0, \pi\right], \tag{12}$$

where expectation $\mathbb{E}\{\cdot\}$ is taken with respect to the joint distribution of the random variables in $W_t = [W_t^{MS}, W_t^g, W_t^{SM}]^T$ for $t = 0, 1, \ldots$ and the given policy π. Our objective is to find the optimal policy π^* which is defined as $\pi^* = \arg\min_{\pi\in\Pi} J_{\pi,N}(h_0)$.

For finite N, the optimal policy π^* can be obtained using the dynamic programming algorithm. However, the difficulty with this approach is that the dynamic programming algorithm is carried out over a state space of expanding dimension. As new observations are made and new actions are taken, the dimension of h_t increases accordingly. To overcome this difficulty, h_t can be replaced by a sufficient statistic, i.e., a quantity that summarizes all the essential content of h_t necessary for control purposes. In the POMDP literature, a sufficient statistic that is often used is the belief state, which is presented in the following section.

H. BELIEF STATE

At each time slot t the agent maintains a belief state P_t, i.e., a probability distribution over all possible system states, $P_t = [p_t^0, \ldots, p_t^7]^T$. Starting from an arbitrarily initialized belief state P_0, the agent updates its belief about the actual state of the system at the beginning of each time slot as follows,

$$p_{t+1}^j = \frac{\sum_{i=0}^{7} p_t^i \cdot p_{ij} \cdot r_j(a, z)}{\sum_{s=0}^{7}\sum_{i=0}^{7} p_t^i \cdot p_{is} \cdot r_s(a, z)}, \tag{13}$$

where $p_{ij} = P[s_{t+1} = j | s_t = i]$. In the literature p_{ij} is usually a function of the action selected at time t, i.e., $p_{ij}(a_t)$, however, in our case the actions taken by the agent do not affect the system's state. In any case, the action taken by the agent affects the observation z_{t+1} made by the agent and thus directly affects the evolution of the belief state over time. As mentioned in Section I, based on P_t the agent forms the *health status belief vector* P_t^h that represents our belief regarding the health status of the sub-system comprised of the sensor and the l_{SM} link. We have, $P_t^h = [p_t^h, p_t^f]$, where p_t^h and p_t^f represent, respectively, the probabilities for the sub-system to be in a healthy or faulty state. We

$$r_s(0,0) = P\left[W_t^g = 0 | F_t^S, 0\right] + P\left[W_t^g = 1 | F_t^S, 0\right] P\left[W_t^{SM} = 0 | F_t^{SM}, 0\right] \tag{7}$$

$$r_s(0,1) = P\left[W_t^g = 1 | F_t^s, 0\right] P\left[W_t^{SM} = 1 | F_t^{SM}, 0\right] \tag{8}$$

$$\begin{aligned} r_s(1,0) = {} & P\left[W_t^{MS} = 0 | F_t^{MS}, 1\right] + P\left[W_t^{MS} = 0 | F_t^{MS}, 1\right] P\left[W_t^g = 1 | F_t^s, 1\right] P\left[W_t^{SM} = 0 | F_t^{SM}, 1\right] \\ & + P\left[W_t^{MS} = 1 | F_t^{MS}, 1\right] P\left[W_t^g = 0 | F_t^S, 1\right] + P\left[W_t^{MS} = 1 | F_t^{MS}, 1\right] P\left[W_t^g = 1 | F_t^s, 1\right] P\left[W_t^{SM} = 0 | F_t^{SM}, 1\right] \end{aligned} \tag{9}$$

$$\begin{aligned} r_s(1,1) = {} & P\left[W_t^{MS} = 1 | F_t^{MS}, 1\right] P\left[W_t^g = 1 | F_t^S, 1\right] P\left[W_t^{SM} = 1 | F_t^{SM}, 1\right] + \\ & P\left[W_t^{MS} = 0 | F_t^{MS}, 1\right] P\left[W_t^g = 1 | F_t^S, 1\right] P\left[W_t^{SM} = 1 | F_t^{SM}, 1\right] \end{aligned} \tag{10}$$

define $p_t^h = p_t^0 + p_t^4$, since states with index 0 and 4 in Table 1 are the only states where both the sensor and the l_{SM} link are in a healthy state. Correspondingly, we define $p_t^f = \sum_{i\neq 0,4} p_t^i$, $i = 0, \ldots, 7$. It holds that P_t^h is a probability distribution since p_t^h and p_t^f are computed over complementary subsets of the system's state space and P_t is a probability distribution. Finally, the health status entropy is computed as $H_t = -[p_t^h \cdot log_2(p_t^h) + p_t^f \cdot log_2(p_t^f)]$.

For the agent to have all the information necessary to proceed with the decision process it must also keep the value of the AoI as part of its state, thus we augment the belief state with the value of AoI and define the following representation of the current state,

$$x_t = \left[P_t, \bar{\Delta}_t\right], \tag{14}$$

and define X to be the set of all states.

I. DYNAMIC PROGRAM OF $\mathcal{P}$

By utilizing the belief state formulation and assuming a finite horizon N the optimal policy π^* can be obtained by solving the following dynamic program,

$$J_t(x_t) = \min_{a_t\in\{0,1\}}\left[g_t + \sum_z \sum_s \sum_i p_t^i p_{is}\, r_s(a_t, z)\, J_{t+1}(x_{t+1})\right], \tag{15}$$

for all $x_t \in X$ and $t = 0, 1, \ldots, N$, where $x_{t+1} = [P_{t+1}^{a,z}, \Delta_{t+1}^z]$, $z \in \{0, 1\}$, $s, i \in \{0, 1, \ldots, 7\}$ and the terminal cost is given by $J_N(x_N) = g_N$.

The formulation of (15) differs from the typical dynamic program for the general case of a POMDP [40], [41], due to the fact that the transition cost depends only on the observed values.

It is known that for (15) there do exist optimal stationary policies [40], [41], i.e., $\pi^* = \{\pi_0^*, \pi_1^*, \ldots, \pi_{N-1}^*\}$. However, since the state space X is uncountable the recursion in (15) does not translate into a practical algorithm. Nevertheless, based on (15) we can prove that the optimal policy has certain structural properties that can be utilized for its efficient computation.

IV. ANALYSIS

In this section we present structural results for the optimal policy of the POMDP $\mathcal{P}$ defined in the previous section. In order to represent the belief state at the $(t+1)$-th time slot one has to consider the action that was taken at time t, i.e., a_t, and the observation made at $(t+1)$, i.e., z_{t+1}, thus, we use $P_{t+1}^{a,z}$ to represent the belief state at the $(t+1)$-th time slot, when $a_t = a$ and $z_{t+1} = z$. In this work we assume that POMDP $\mathcal{P}$ satisfies the following two assumptions.

Assumption 1: Let $x_t = [P_t, \bar{\Delta}_t]$ and $x_t^+ = [P_t^+, \bar{\Delta}_t]$ be states such that $H(P_t^{h,+}) \geq H(P_t^h)$ then $H(P_{t+1}^{h,+,a,z}) \geq H(P_{t+1}^{h,a,z})$, $a, z \in \{0, 1\}$.

Assumption 1 states that, given the action at t and the observation at $t+1$, if the system starts in a belief state with higher health status entropy, i.e., $H(P_t^{h,+}) \geq H(P_t^h)$, then it will make a transition to a state with higher health status entropy, i.e., $H(P_{t+1}^{h,a,z,+}) \geq H(P_{t+1}^{h,a,z})$.

Assumption 2: Let $I_S = \{0, 1, \ldots, 7\}$ and $i \in I_S$ be the index of the system's state $s_t = [i_0, i_1, i_2]^T$ at time $t = 0, 1, \ldots, N$, where $i_0 = F^{MS}$, $i_1 = F^S$, and $i_2 = F^{SM}$ (see Table 1). Furthermore, let $p_{i_10}^S$, $p_{i_20}^{SM}$ be the probabilities for the sensor S and the link l_{SM} to make a transition from health status i_1 and i_2, respectively, to a healthy status (indicated by 0) at $t+1$. We assume that for the POMDP $\mathcal{P}$ the following inequality is true,

$$\sum_{i\in I_S} p_t^i\Big[p_{i_10}^S p_{i_20}^{SM}(2 - P_g) - 1\Big] \leq 0, \quad t = 0, 1, \ldots, N. \tag{16}$$

Assumption 2 expresses the necessary conditions and system's parametrization for the probing action to always result in a lower health status entropy compared to the no probe action. It may seem intuitive that probing reduces entropy since it makes the generation of a status update from the sensor mandatory, i.e., it reduces the uncertainty induced in the system due to the probabilistic generation of status updates from the sensor. However, one should also consider that probing introduces a new type of uncertainty in the system due to the transmission failures occurring in the l_{MS} link. For example, consider the case where a probe was sent to the sensor, yet the monitor received no status update. It is uncertain whether this happened because the probe didn't reach the sensor due to a faulty l_{MS} link, or because the sensor, or the l_{SM} link, or both were in a faulty state. Assumption 2 expresses the effect of faults in the l_{MS} link along with that of parameters $p_{i_10}^S$, $p_{i_20}^{SM}$ and P_g on the health status entropy (for details see Appendices A and B) and it is used in the following lemma to prove that the probe action will always result in the same or reduced health status entropy compared to the no-probe action for a given observation z at time $t+1$.

Lemma 1: Let $P_{t+1}^{0,z}$ and $P_{t+1}^{1,z}$ be the belief states of $\mathcal{P}$ at the $(t+1)$-th time slot when $a_t = 0$ and 1, respectively, and let $P_{t+1}^{h,0,z}$, $P_{t+1}^{h,1,z}$ be their corresponding health status belief vectors, then, if Assumption 2 is satisfied, it holds that, $H(P_{t+1}^{h,0,z}) \geq H(P_{t+1}^{h,1,z})$, $\quad z \in \{0, 1\}$.

The proof is given in Appendix A. Next, in Lemma 2, we show that the expected cost-to-go from decision stage t up to N is an increasing function of the health status entropy.

Lemma 2: Let $x_t^+ = [P_t^+, \bar{\Delta}_t]$ and $x_t^- = [P_t^-, \bar{\Delta}_t]$ be states such that $H(P_t^{h,+}) \geq H(P_t^{h,-})$ and $J_t(\cdot)$ be the dynamic program of $\mathcal{P}$ then for $t = 1, \ldots, N$, it holds that $J_t(P_t^+, \bar{\Delta}_t) \geq J_t(P_t^-, \bar{\Delta}_t)$.

Proof: The proof of Lemma 2 is given in Appendix C. ■

In Lemma 3 we state a similar property for the expected cost-to-go when the value of AoI increases. We omit the proof of Lemma 3 since it is intuitive and its proof follows a similar line of arguments as in Lemma 2.

Lemma 3: Let $\bar{\Delta}_t^+$ and $\bar{\Delta}_t^-$ be normalized AoI values such that $\bar{\Delta}_t^+ \geq \bar{\Delta}_t^-$ and $J_t(\cdot)$ be the cost-to-go function in

Algorithm 1 Policy Gradient Algorithm for Probing Control

1: Initialize threshold $\theta_0 = [\theta_0^H, \theta_0^{\bar{\Delta}}]$ and γ, A, η, β, ζ
2: **for** $k = 1$ to K **do**
3: $\gamma_k = \frac{\gamma}{(k+A)^\beta}$ and $\eta_k = \frac{c}{k^\zeta}$
4: Randomly set ω_k^H, $\omega_k^{\bar{\Delta}}$ to the equiprobable values $\{-1, 1\}$ and define $\omega_k = [\omega_k^H, \omega_k^{\bar{\Delta}}]^T$
5: $\theta_k^+ = \min\{\mathbf{1}, \max\{\mathbf{0}, \theta_{k-1} + \eta_k \cdot \omega_k\}\}$ and $\theta_k^- = \min\{\mathbf{1}, \max\{\mathbf{0}, \theta_{k-1} - \eta_k \cdot \omega_k\}\}$
6: $y_k^+ = \hat{J}(\theta_k^+)$, $y_k^- = \hat{J}(\theta_k^-)$
7: $\hat{e_k} = (y_k^+ - y_k^-) \oslash (2 \cdot c_k \cdot \omega_k)$
8: $\theta_k = \theta_{k-1} - \gamma_k \hat{e}$
9: **end for**

the dynamic program (15) then for $t = 0, 1, \ldots, N-1$, it holds that $J_t(P_t, \bar{\Delta}_t^+) \geq J_t(P_t, \bar{\Delta}_t^-)$.

In Lemma 4 we prove properties of the cost-to-go function $J_t(\cdot)$ that are necessary to establish the structural properties of the optimal policy in Theorem 1.

Lemma 4: Let $J_t(x_t)$ be the value of the dynamic program of $\mathcal{P}$ at $x_t = [P_t, \bar{\Delta}_t]$ then $J_t(x_t)$ is piece-wise linear, increasing, and concave with respect to $H(P_t^h)$ and $\bar{\Delta}$ for $t = 1, \ldots, N$.

Proof: The proof of Lemma 4 is given in Appendix D. ■

Finally, in Theorem 1 we show that there exist configurations of POMDP $\mathcal{P}$ such that the optimal policy is threshold based.

Theorem 1: At each decision stage $t = 0, 1, \ldots, N-1$ there exists a positive probing cost c such that the probing action is optimal for state $x_t^T = [P_t, \bar{\Delta}_t]$ and for all states $x_t^+ = [P_t^+, \bar{\Delta}_t^+]$ with $H(P_t^{h,+}) \geq H(P_t^h)$ and $\bar{\Delta}_t^+ \geq \bar{\Delta}_t$.

Proof: The proof of Theorem 1 is given in Appendix E. ■

V. OPTIMAL POLICY APPROXIMATION

According to Theorem 1, given a proper probing cost c, the optimal policy $\pi^* = \{\pi_0^*, \pi_1^*, \ldots, \pi_{N-1}^*\}$ for the finite horizon POMDP $\mathcal{P}$ is of a threshold type. This means that π^* is comprised of different threshold values at *each* decision stage $t = 0, 1, \ldots, N$. More specifically, let $\theta_t^{H,*}$ and $\theta_t^{\bar{\Delta},*}$ be the optimal threshold values for the health status entropy and the normalized AoI at stage t, then the optimal policy can be expressed as $\pi^* = \{[\theta_0^{H,*}, \theta_0^{\bar{\Delta},*}], [\theta_1^{H,*}, \theta_1^{\bar{\Delta},*}], \ldots, [\theta_{N-1}^{H,*}, \theta_{N-1}^{\bar{\Delta},*}]\}$. Computing $\theta_k^* = [\theta_k^{H,*}, \theta_k^{\bar{\Delta},*}]^T$ for $t = 0, 1, \ldots, N$ can be a computationally demanding task, especially if one considers large time horizons. To address this problem we approximate the optimal policy π^* with a *single* threshold and utilize a Policy Gradient algorithm, namely, the Simultaneous Perturbation Stochastic Approximation (SPSA) Algorithm [42] in order to find it.

The SPSA algorithm appears in Algorithm 1 and operates by generating a sequence of threshold estimates, $\theta_k = [\theta_k^H, \theta_k^{\bar{\Delta}}]^T$, $k = 1, 2, \ldots, K$ that converges to a local minimum, i.e., an approximation of the best single threshold policy for POMDP $\mathcal{P}$. The SPSA algorithm picks a single random direction ω_k along which the derivative is evaluated at each step k, i.e., ω_k^H and $\omega_k^{\bar{\Delta}}$ are independently generated according to a Bernoulli distribution as presented in line 4 of Algorithm 1. Subsequently, in line 5 the algorithm generates threshold vectors θ_k^+ and θ_k^-, which are bounded element-wise in the interval [0, 1], i.e., **0** and **1** in line 5 are column vectors whose elements are all zeros and ones respectively. $\theta_k^{\bar{\Delta}}$ is also bounded in [0, 1] since we assumed a normalized value for the AoI, and, this is also true for θ_k^H since the maximum health status entropy occurs for $P_t^h = [0.5, 0.5]$ which evaluates to 1. In line 6 the estimates $\hat{J}(\theta^+)$ and $\hat{J}(\theta^-)$ are computed by simulating M_s times the POMDP $\mathcal{P}$ under the corresponding single threshold policy. Finally, the gradient is estimated in line 7, where $\oslash$ represents an element-by-element division, and θ_k is updated in line 8. Since the SPSA algorithm converges to local optima it is necessary to try several initial conditions θ_0.

VI. NUMERICAL RESULTS

In this section, we evaluate numerically the cost efficiency of the threshold probing policy we derived previously. Furthermore, we provide comparative results with an alternative probing policy that is often used in practice. The latter policy will probe the sensor whenever the time that has elapsed since the last arrival of a status update at the monitor exceeds a certain threshold. We will refer to this policy as the *delay* based policy. In contrast, we will refer to the single threshold policy, that approximates the optimal policy, as the *threshold* policy. We also note here, that the delay metric, as described above, and the AoI metric coincide, i.e., delay based policies are also AoI based policies. This holds because the sensor does not buffer status updates and the status update generation scheme is fixed. The results we present in this section exhibit the advantage of using VoI instead of AoI when deciding whether to probe or not. This advantage of VoI can be attributed to the fact that AoI captures only the timeliness of information and misses out semantical information related to the health status of the system.

Furthermore, to gain insight into how the various system parameters affect the performance of the probing policies, we formulated a *basic* scenario and subsequently varied its parameters. For the basic scenario, the system was configured as follows, $c = 1$, $\lambda_1 = 1$, $\lambda_2 = 1$, $P_g = 0.1$ and the transition probability kernels were set as, $P^{MS} = \begin{bmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{bmatrix}$, $P^S = \begin{bmatrix} 0.9 & 0.1 \\ 0.9 & 0.1 \end{bmatrix}$, and $P^{SM} = \begin{bmatrix} 0.9 & 0.1 \\ 1 - p_{SM}^{11} & p_{SM}^{11} \end{bmatrix}$, where $p_{SM}^{11} = 0.1, 0.2, \ldots, 0.9$. Furthermore, we set the parameters of the SPSA algorithm as follows, $\eta = 1$, $\gamma = 10^{-3}$, $A = 1$, $\beta = 1$, and $\zeta = 1$. We derived the threshold policy by executing $K = 20$ iterations of the SPSA algorithm. At each iteration $k = 1, 2, \ldots, K$ we calculated each of y_k^+ and y_k^- as the sample average of 100 Monte-Carlo simulations. Each Monte-Carlo simulation had a time horizon of N time slots and during that period the system was controlled by the

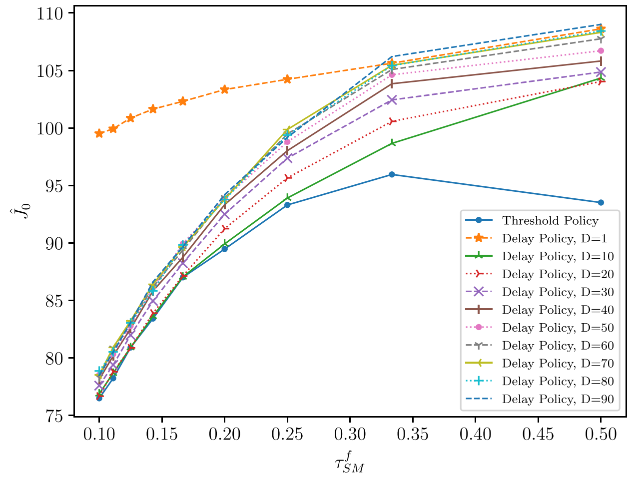


FIGURE 2. $\hat{J}_0(\cdot)$ vs τ^f_{SM} for a horizon of 100 time slots.

single threshold policy defined by θ_k^+, in the case of y_k^+, and θ_k^-, in the case of y_k^-, as presented in Algorithm 1. Subsequently, we used the threshold θ_K to evaluate the efficiency of the derived threshold policy. More specifically, for all policies appearing in Figure 2, was calculated the average cost $\hat{J}_0$ as the sample average over $M = 2000$ Monte-Carlo simulations of the system while it was being controlled by the corresponding policy over a period of N time slots, i.e., $\hat{J}_0 = \frac{1}{M}\sum_{m=1}^{M}\sum_{t=0}^{N} g_t$. Finally, for each Monte-Carlo simulation we set randomly the initial health status for the sensor and the l_{MS} and l_{SM} links.

In Figure 2 we present the evolution of $\hat{J}_0$ with respect to the *steady state probability of link l_{SM} to be in a faulty state*,

$$\tau^f_{SM} = \frac{p^{01}_{SM}}{1 - p^{11}_{SM} + p^{01}_{SM}}, \tag{17}$$

where $p^{11}_{SM} = 0.1, 0.2, \ldots, 0.9$. We utilize the steady state probability of link l_{SM} to be in a faulty state instead of the transition probability of link l_{SM} from a faulty state to a faulty state, p^{11}_{SM}, because it assumes a more intuitive interpretation, i.e., it expresses the expected time that link l_{SM} will spend in the faulty state over a large time horizon. For each different value of p^{11}_{SM} we derived the approximately optimal threshold policy by executing K iterations of the SPSA algorithm as described in the previous paragraph.

In Figure 2 we consider delay policies that will probe the sensor whenever a new status update arrival has been delayed for more than $D = 1, 10, 20, \ldots, 90$ time slots. The results presented in Figure 2 indicate that the threshold based policy achieves a lower cost $\hat{J}_0$ compared to the delay based policies for all values of D. In order to provide insights into these results we have to elaborate on the behavior of the delay policy for the two extreme values of D, i.e., for D equal to 1 and 90. For $D = 1$ the delay policy probes the sensor almost at every single time slot. This is because the policy requires a fresh status at every time slot while the sensor generated and transmitted status updates at any time slot with a small probability, $P_g = 0.1$. On the other hand, the delay policy with $D = 90$ practically never probed the sensor within the time horizon of 100 time slots. This is because at least one status update arrival at the monitor, within the first 90 time slots, would be adequate for the delay between status update arrivals to never exceed the threshold of 90 time slots.

Figure 2 shows that when τ^f_{SM} is less than 0.20 the threshold policy and the delay policy with $D = 90$ have similar expected cost values $\hat{J}_0$. Given the behavior of the delay policy with $D = 90$, as presented above, one can conclude that the arrival of status updates at the monitor occurs with enough frequency to render probing unnecessary. When τ^f_{SM} is within the range of 0.20 and 0.30 the cost induced by the delay policy with $D = 90$ increases with a higher rate compared to all other policies. This indicates that probing effectively contributes to the reduction of the expected cost $\hat{J}_0$ and since the delay policy with $D = 90$ doesn't engage in probing it suffers a high cost. This is evident also by the fact that the delay based policy with $D = 10$ performs closer to the threshold policy now compared to all other delay based policies. Finally, when τ^f_{SM} becomes larger than 0.30 the expected cost $\hat{J}_0$ increases for all delay policies and decreases for the threshold based policy. For this range of τ^f_{SM} values the periods that the l_{SM} link remains in a faulty state increase in duration due to the higher probability value for link l_{SM} to remain in a faulty state, p^{11}_{SM}. As a consequence of this, the delay between status update arrivals exceeds the threshold value D for all delay policies more often than before and, consequently, they probe the sensor with higher frequency. However, whenever the l_{SM} link is in a faulty state, the dynamics of the system make it impossible for a status update to reach the monitor despite the transmission of a probe. As a result, neither the health status entropy nor the normalized AoI can be reduced effectively by probing. This behavior is particularly evident in the abrupt increase of the expected cost for the delay policies with $D = 1$ and $D = 10$ which probed the sensor with higher frequency due to their low values of D. On the other hand, the threshold based policy succeeded in reducing the expected cost by considering, for its probing decisions, its confidence on the health status of the system, the normalized AoI and the probing const c.

In Figure 3 we present cost $\hat{J}_0$ for a wider range of values for τ^f_{SM}. More specifically, we modified the basic scenario by increasing the probability of link l_{SM} to enter a faulty state, given that it is in a healthy state, p^{01}_{SM}, from 0.1 to 0.2. As a result, the l_{SM} link enters more often its faulty state compared to the basic scenario and this provides for a wider range of τ^f_{SM} values. All policies exhibit the same behavior as in the basic scenario for values of τ^f_{SM} up to 0.5. However, when τ^f_{SM} becomes larger than 0.5 we observe a reduction in the expected cost $\hat{J}_0$ for all policies except for the delay based policies with $D = 1$ and $D = 10$. The observed reduction in $\hat{J}_0$ was mainly due to the reduction in the cost induced by the health status entropy. More specifically, for large values

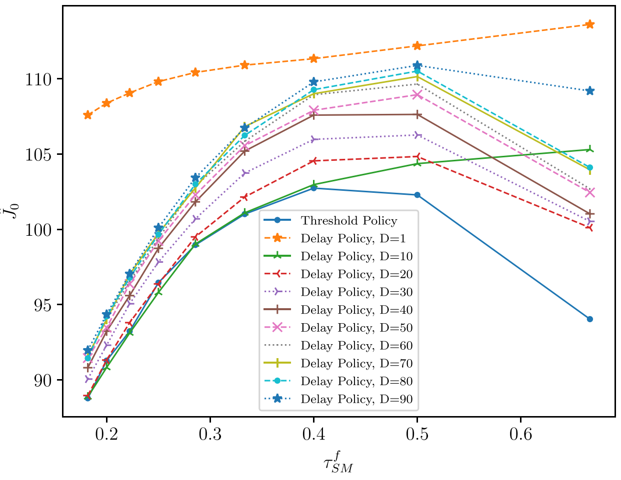


FIGURE 3. $J_0(\cdot)$ vs. τ^f_{SM} when we increase the probability for the sensor to enter a faulty state.

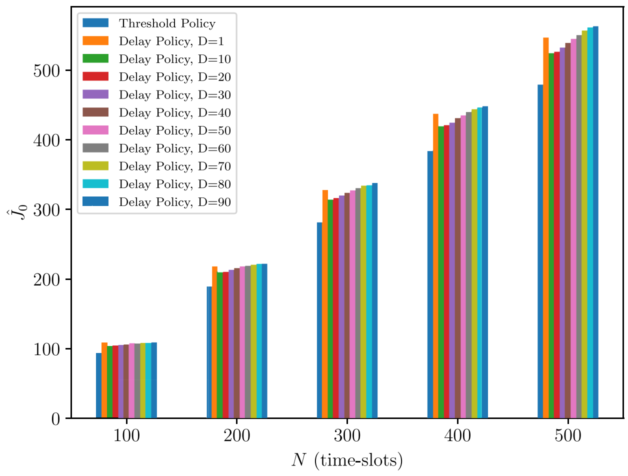


FIGURE 5. Sample average of cost $J_0(\cdot)$ vs the optimization horizon.

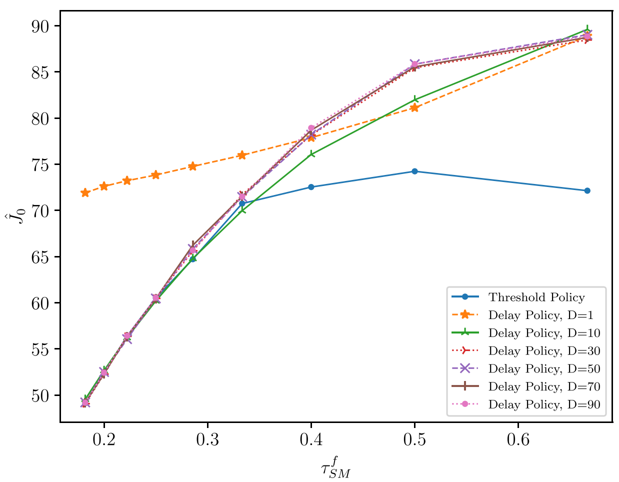


FIGURE 4. $J_0(\cdot)$ vs. τ^f_{SM} when only the l_{SM} link is subject to failures.

τ^f_{SM}, i.e., for large values of p^{11}_{SM} and p^{01}_{SM}, the monitor can be confident that the l_{SM} link is in a faulty state, and this resulted in an overall reduction of the health status entropy cost, H_t. The delay policies with $D = 1$ and $D = 10$ exhibit an increased cost $\hat{J}_0$ which is mainly due to the persistent probing of the sensor while the l_{SM} link was in a faulty state and thus no status update could be transmitted successfully to the monitor. As expected, for large values of τ^f_{SM}, this behavior resulted in a large number of unnecessary probes. Policies with a larger value of D were also engaged into this type of probing albeit with a lower frequency.

In Figure 4 we present the effect of the health status entropy on the expected cost $\hat{J}_0$. We modified the basic scenario by setting, $P^{MS} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, $P^S = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, and by increasing p^{01}_{SM} from 0.1 to 0.2 in order to get the same range of τ^f_{SM} values as in Figure 3. By setting the matrices P^{MS} and P^S to the values presented above both the link l_{MS} and the sensor S would never enter a faulty state and, even if they were randomly initialized to a faulty state they would return to the healthy state deterministically in the next time slot. Thus, the system can be in one of two possible states. In the first state both links and the sensor are healthy while in the second state the l_{MS} link and the sensor are healthy whereas the l_{SM} is faulty. This comes in contrast to the eight possible states of the basic scenario and results in diminishing the uncertainty in the system's health status, i.e., the health status entropy diminishes for all policies and across the whole range of τ^f_{SM} values. As a consequence of this, we do not observe in Figure 4 the significant reduction in the expected cost $\hat{J}_0$ we observed in Figure 3 for τ^f_{SM} values greater than 0.5. To elaborate more on this result, the results in Figure 4 point out the significance of that the health status entropy transition cost can have on the expected cost and by contrasting the results in Figure 4 with those of Figure 3 we conclude that in Figure 3 the threshold based policy was able to successfully reduce the expected cost by correctly identifying the health status of the system and thus reducing the health status entropy transition costs.

Finally, in Figure 5 we present the effect of an increasing time horizon N on $\hat{J}_0(\cdot)$. We modified the basic system setup by setting $p^{11}_{SM} = 0.9$ and $\lambda_2 = \frac{N}{100}$. By setting $\lambda_2 = \frac{N}{100}$ we had, a normalized AoI cost of $\bar{\Delta} = \frac{\Delta}{100}$, which was analogous to that of the basic scenario irrespective to the time horizon N. Figure 5 depicts that an increment of N results in an increased $\hat{J}_0(\cdot)$ for all policies. Furthermore, by calculating the relative difference of the expected cost of the threshold policy with respect to that of the delay policy with $D = 90$ we observed that the threshold policy achieved a constant reduction of 16% in the expected cost across all experiments.

VII. CONCLUSION

In this work, we address the problem of deriving an efficient policy for sensor probing in IoT networks with intermittent faults. We adopted a semantics-aware communications paradigm for the transmission of probes whereby the importance (semantics) of the probe is considered before its generation and transmission. We formulated the problem as a POMDP and proved that the optimal policy is of a threshold type. We used a computationally efficient stochastic approximation algorithm to derive the probing policy. Finally, the numerical results presented in this work exhibit a significant cost reduction when the derived probing policy is followed instead of a conventional delay-based one.

APPENDIX A
PROOF OF LEMMA 1

First we present the part of the proof for the case where the monitor observes a fresh status update at $t+1$, i.e., $z_{t+1}=1$. From Equation (13) and the column of Table 1 that corresponds to $a_t=0$ and $z_{t+1}=1$ we have that, $P^{0,1}_{t+1}$ has only the following two non-zero elements, $p^0_{t+1}=\frac{\sum_{i=0}^7 p^i_t p_{i0}P_g}{\sum_{i=0}^7 p^i_t(p_{i0}+p_{i4})P_g}$ and $p^4_{t+1}=\frac{\sum_{i=0}^7 p^i_t p_{i4}P_g}{\sum_{i=0}^7 p^i_t(p_{i0}+p_{i4})P_g}$. We represent the numerators of p^0_{t+1} and p^4_{t+1} with $\phi_1=\sum_{i=0}^7 p^i_t p_{i0}P_g$ and $\phi_2=\sum_{i=0}^7 p^i_t p_{i4}P_g$ and we get, $p^0_{t+1}=\frac{\phi_1}{\phi_1+\phi_2}$ and $p^4_{t+1}=\frac{\phi_2}{\phi_1+\phi_2}$. Furthermore, we can express the belief state vector at the next time slot as, $P^{0,1}_{t+1}=[\frac{\phi_1}{\phi_1+\phi_2},0,0,0,\frac{\phi_2}{\phi_1+\phi_2},0,0,0]^T$, and the corresponding health status belief vector as, $P^{h,0,1}_{t+1}=[p^h_{t+1},p^f_{t+1}]=[p^0_{t+1}+p^4_{t+1},0]=[1,0]$.

Similarly, from Equation (13) and the column of Table 1 that corresponds to $a_t=1$ and $z_{t+1}=1$ we have that $P^{1,1}_{t+1}$ has again only two non-zero elements, $p^0_{t+1}=\frac{\sum_{i=0}^7 p^i_t p_{i0}}{\sum_{i=0}^7 p^i_t(p_{i0}+p_{i4}P_g)}$ and $p^4_{t+1}=\frac{\sum_{i=0}^7 p^i_t p_{i4}P_g}{\sum_{i=0}^7 p^i_t(p_{i0}+p_{i4}P_g)}$. We represent the numerators of p^0_{t+1} and p^4_{t+1} with $\xi_1=\sum_{i=0}^7 p^i_t p_{i0}$ and $\xi_2=\sum_{i=0}^7 p^i_t p_{i4}P_g$ and we get, $p^0_{t+1}=\frac{\xi_1}{\xi_1+\xi_2}$ and $p^4_{t+1}=\frac{\xi_2}{\xi_1+\xi_2}$. The belief state vectors at the next time slot, will be, $P^{1,1}_{t+1}=[\frac{\xi_1}{\xi_1+\xi_2},0,0,0,\frac{\xi_2}{\xi_1+\xi_2},0,0,0]^T$ and the corresponding health status belief vector will be, $P^{h,1,1}_{t+1}=[p^0_{t+1}+p^4_{t+1},0]=[1,0]$ From the expressions for $P^{h,0,1}_{t+1}$ and $P^{h,1,1}_{t+1}$ given above we get that $H(P^{h,0,1}_{t+1})=H(P^{h,1,1}_{t+1})$.

Next we present the case where the monitor does not observe a fresh status update at $t+1$, i.e., $z_{t+1}=0$. We substitute, for each state index j in (13), the observation probabilities presented in the column of Table 1 that corresponds to $a_t=0$ and $z_{t+1}=0$ and we derive for $j=0$, $p^0_{t+1}=\frac{\sum_{i=0}^7 p^i_t p_{i0}(1-P_g)}{\sum_{i=0}^7 p^i_t[p_{i0}(1-P_g)+p_{i4}(1-P_g)+\sum_{j\neq 0,4}p_{ij}]}$, for $j=4$, $p^4_{t+1}=\frac{\sum_{i=0}^7 p^i_t p_{i4}(1-P_g)}{\sum_{i=0}^7 p^i_t[p_{i0}(1-P_g)+p_{i4}(1-P_g)+\sum_{j\neq 0,4}p_{ij}]}$, and for $j=\{0,1,\ldots,7\}\setminus\{0,4\}$, $p^j_{t+1}=\frac{\sum_{i=0}^7 p^i_t p_{ij}}{\sum_{i=0}^7 p^i_t[p_{i0}(1-P_g)+p_{i4}(1-P_g)+\sum_{j\neq 0,4}p_{ij}]}$. By setting $\xi_1=\sum_{i=0}^7 p^i_t p_{i0}(1-P_g)$, $\xi_2=\sum_{i=0}^7 p^i_t p_{i4}(1-P_g)$, $\phi_j=\sum_{i=0}^7 p^i_t p_{ij}$ for $j=\{0,1,\ldots,7\}\setminus\{0,4\}$ and $\phi_s=\sum_{j\neq 0,4}\phi_j$ we get that the belief state at $t+1$ is, $P^{0,0}_{t+1}=\frac{1}{\xi_1+\xi_2+\phi_s}\cdot[\xi_1,\phi_1,\phi_2,\phi_3,\xi_2,\phi_5,\phi_6,\phi_7]^T$. The resulting health status belief vector will be, $P^{h,0,0}_{t+1}=[\frac{\xi_1+\xi_2}{\xi_1+\xi_2+\phi_s},\frac{\phi_s}{\xi_1+\xi_2+\phi_s}]^T$.

Similarly, by substituting, for each state index j in Equation (13), the observation probabilities presented in the column of Table 1 that corresponds to $a_t=1$ and $z_{t+1}=0$, we derive for $j=0$, $p^0_{t+1}=0$ for $j=4$, $p^4_{t+1}=\frac{\sum_{i=0}^7 p^i_t p_{i4}(1-P_g)}{\sum_{i=0}^7 p^i_t[p_{i4}(1-P_g)+\sum_{j\neq 0,4}p_{ij}]}$, and for $j=\{0,1,\ldots,7\}\setminus\{0,4\}$, $p^j_{t+1}=\frac{\sum_{i=0}^7 p^i_t p_{ij}}{\sum_{i=0}^7 p^i_t[p_{i0}(1-P_g)+p_{i4}(1-P_g)+\sum_{j\neq 0,4}p_{ij}]}$. The belief state at $t+1$ can be expressed as, $P^{1,0}_{t+1}=\frac{1}{\xi_1+\xi_2+\phi_s}\cdot[0,\phi_1,\phi_2,\phi_3,\xi_2,\phi_5,\phi_6,\phi_7]^T$. Then, the resulting health status belief vector will be, $P^{h,1,0}_{t+1}=[\frac{\xi_2}{\xi_2+\phi_s},\frac{\phi_s}{\xi_2+\phi_s}]^T$.

In order to prove that $H(P^{h,0,0}_{t+1})\geq H(P^{h,1,0}_{t+1})$ it is adequate to show that the probability distribution $P^{h,0,0}_{t+1}$ is closer to a uniform distribution compared to $P^{h,1,0}_{t+1}$, i.e., the following inequality is true

$$\left|\frac{\xi_1+\xi_2}{\xi_1+\xi_2+\phi_s}-\frac{\phi_s}{\xi_1+\xi_2+\phi_s}\right|\leq\left|\frac{\xi_2}{\xi_2+\phi_s}-\frac{\phi_s}{\xi_2+\phi_s}\right|. \tag{18}$$

In Appendix B we show that under Assumption 2 it holds that, $\xi_1+\xi_2\leq\phi_s$ and, consequently, that $\xi_2\leq\phi_s$, thus we have, $\frac{\phi_s-\xi_1+\xi_2}{\xi_1+\xi_2+\phi_s}\leq\frac{\phi_s-\xi_2}{\xi_2+\phi_s}$. With simple algebraic manipulations the equation presented above can be expressed as, $-2\xi_1\phi_s\leq 0$ which is true since both ξ_1, ϕ_s are probabilities. This concludes the proof and shows that the probing action results in the same or reduced health status entropy.

APPENDIX B

In this appendix we show that Assumption 2 is equivalent to $\xi_1+\xi_2\leq\phi_s$. For convenience we repeat here the definitions of ξ_1, ξ_2 and ϕ_s, i) $\xi_1=\sum_{i=0}^7 p^i_t p_{i0}(1-P_g)$ ii) $\xi_2=\sum_{i=0}^7 p^i_t p_{i4}(1-P_g)$ iii) $\phi_s=\sum_{i=0}^7 p^i_t\sum_j p_{ij}$, $j=I_S\setminus\{0,4\}$, where $I_S=\{0,1,\ldots,7\}$ and in ϕ_s we changed the order of summation. Now, by substituting to $\xi_1+\xi_2\leq\phi_s$ we get,

$$\sum_{i=0}^{7}p^i_t p_{i0}(1-P_g)+\sum_{i=0}^{7}p^i_t p_{i4}(1-P_g)\leq\sum_{i=0}^{7}p^i_t\sum_{j\in I_s\setminus\{0,4\}}p_{ij}, \tag{19}$$

which can be written as $\sum_{i=0}^7 p^i_t[p_{i0}(1-P_g)+p_{i4}(1-P_g)-\sum_{j\in I_s\setminus\{0,4\}}p_{ij}]\leq 0$. Subsequently we express the transition probability from state with index i to state with index j as $P[s_{t+1}=j|s_t=i]\ =p_{ij}=p^{MS}_{i_0j_0}\cdot p^S_{i_1j_1}\cdot p^{SM}_{i_2j_2}$, where i_0, i_1, and i_2 represent respectively the states of the l_{MS} link, the sensor and the l_{SM} link at time t, while j_0, j_1, and j_2 represent respectively the states of the l_{MS} link, the sensor and the l_{SM} link at time $t+1$. Thus expression, $[p_{i0}(1-P_g)+p_{i4}(1-P_g)-\sum_{j\in I_s\setminus\{0,4\}}p_{ij}]$ can be written as, $[p^{MS}_{i_00}p^S_{i_10}p^{SM}_{i_20}(1-P_g)+p^{MS}_{i_01}p^S_{i_10}p^{SM}_{i_20}(1-P_g)-p^{MS}_{i_00}p^S_{i_10}p^{SM}_{i_21}-p^{MS}_{i_00}p^S_{i_11}p^{SM}_{i_20}-p^{SM}_{i_20}p^S_{i_11}p^{SM}_{i_21}-p^{MS}_{i_01}p^S_{i_10}p^{SM}_{i_21}-p^{MS}_{i_01}p^S_{i_11}p^{SM}_{i_20}-p^{MS}_{i_01}p^S_{i_11}p^{SM}_{i_21}]$.

Through simple algebraic manipulations the latter expression can be shown to be equal to $[p^S_{i_10}p^{SM}_{i_20}(2-P_g)-1]$ and, based on this result, (19) can be expressed as $\sum_{i=0}^7 p^i_t[p^S_{i_10}p^{SM}_{i_20}(2-P_g)-1]\leq 0$.

APPENDIX C
PROOF OF LEMMA 2

To prove Lemma 2 we will use induction. For the N-th decision stage Lemma 2 holds trivially, i.e., for $x_N^+ = [P_N^+, \bar{\Delta}_N]$ and $x_N^- = [P_N^-, \bar{\Delta}_N]$ with $H(P_N^{h,+}) \geq H(P_N^{h,-})$ we have,

$$H\left(P_N^{h,+}\right) + \bar{\Delta}_N \geq H\left(P_N^{h,-}\right) + \bar{\Delta}_N \Rightarrow$$
$$J_N\left(P_N^+, \Delta_N\right) \geq J_N\left(P_N^-, \Delta_N\right).$$

Let it be true that for $x_{t+1}^+ = [P_{t+1}^+, \bar{\Delta}_{t+1}]$ and $x_{t+1}^- = [P_{t+1}^-, \bar{\Delta}_{t+1}]$ with $H(P_{t+1}^{h,+}) \geq H(P_{t+1}^{h,-})$ it holds that, $J_{t+1}(P_{t+1}^+, \Delta_{t+1}) \geq J_{t+1}(P_{t+1}^-, \Delta_{t+1})$ then we will prove that, for $x_t^+ = [P_t^+, \bar{\Delta}_t]$ and $x_t^- = [P_t^-, \bar{\Delta}_t]$ with $H(P_t^{h,+}) \geq H(P_t^{h,-})$ it is also true that,

$$J_t\left(P_t^+, \Delta_t\right) \geq J_t\left(P_t^-, \Delta_t\right), \tag{20}$$

We have,

$$J_t\left(P_t^+, \Delta_t\right) = \min\left[A_0^+, A_1^+\right], \tag{21}$$

where,

$$A_0^+ = H\left(P_t^{h,+}\right) + \bar{\Delta}_t + \sum_z \sum_s \sum_i p_t^{i,+} p_{is} r_s(0,z) J_{t+1}\left(P_{t+1}^{0,z,+}, \bar{\Delta}_{t+1}^z\right), \tag{22}$$

$$A_1^+ = c + H\left(P_t^{h,+}\right) + \bar{\Delta}_t + \sum_z \sum_s \sum_i p_t^{i,+} p_{is} r_s(1,z) J_{t+1}\left(P_{t+1}^{1,z,+}, \bar{\Delta}_{t+1}^z\right), \tag{23}$$

where $P_{t+1}^{0,z,+}$ ($P_{t+1}^{1,z,+}$) is the belief state calculated at the $(t+1)$-th time slot from P_t^+ using Equation (13), when $a_t = 0$ ($a_t = 1$) and $z_{t+1} = z$. Similarly, we have,

$$J_t\left(P_t^-, \Delta_t\right) = \min\left[A_0^-, A_1^-\right] \tag{24}$$

where A_0^-, and A_1^- are defined by replacing $+$ with $-$ in (22) and (23) respectively. In order to prove Equation (20) it suffices to show that the following two inequalities hold,

$$A_1^+ \geq A_1^-, \tag{25}$$
$$A_0^+ \geq A_0^-. \tag{26}$$

This can be verified by considering all possible combinations for the values of $J_t(x_t^+)$ and $J_t(x_t^-)$ from (21) and (24): i) Let $J_t(x_t^+) = A_0^+$ and $J_t(x_t^+) = A_0^-$ then by (26) Equation (20) holds. ii) Let $J_t(x_t^+) = A_0^+$ and $J_t(x_t^+) = A_1^-$ then $A_0^+ \geq A_0^-$ by (26) and $A_0^- \geq A_1^-$ due to A_1^- being optimal, i.e., due to the $\min[\cdot]$ operator in (24), thus we have $A_0^+ \geq A_1^-$ and as a result Equation (20) holds. iii) Let $J_t(x_t^+) = A_1^+$ and $J_t(x_t^+) = A_0^-$ then $A_1^+ \geq A_1^-$ by (25) and $A_1^- \geq A_0^-$ due to A_0^- being optimal in (24), thus we have $A_1^+ \geq A_0^-$ and as a result Equation (20) holds. iv) Finally, let $J_t(x_t^+) = A_1^+$ and $J_t(x_t^+) = A_1^-$ then by (25) Equation (20) holds.

Next we show that inequality (25) holds. Proof that Equation (26) holds can be derived in a similar way. Firstly, from the basic assumption of Lemma 2 we have, $H(P_t^{h,+}) \geq H(P_t^{h,-}) \Rightarrow H(P_t^{h,+}) + \bar{\Delta}_t \geq H(P_t^{h,-}) + \bar{\Delta}_t$. Secondly, from Assumption 1, the fact that the value of AoI at $t+1$, i.e., $\bar{\Delta}_{t+1}^z$, will have the same value for a given observation $z_{t+1} = z$ independently of the starting belief state and the induction hypothesis we have that, $J_{t+1}(P_{t+1}^{1,z,+}, \bar{\Delta}_{t+1}^z) \geq J_{t+1}(P_{t+1}^{1,z,-}, \bar{\Delta}_{t+1}^z)$, $z \in \{0, 1\}$. Consequently, for each observation z, $J_{t+1}(P_{t+1}^{1,z,+}, \bar{\Delta}_{t+1}^z)$ is uniformly larger than $J_{t+1}(P_{t+1}^{1,z,-}, \bar{\Delta}_{t+1}^z)$ and thus its expected value at $t+1$ will also be larger [43, Ch. 7, p. 299], i.e.,

$$\sum_s \sum_i p_t^{i,+} p_{is} r_s(a_t, z) J_{t+1}\left(P_{t+1}^{1,z,+}, \bar{\Delta}_{t+1}^z\right) \geq \sum_s \sum_i p_t^{i,-} p_{is} r_s(a_t, z) J_{t+1}\left(P_{t+1}^{1,z,-}, \bar{\Delta}_{t+1}^z\right),$$

and by summing over all possible observations we get,

$$\sum_z \sum_s \sum_i p_t^{i,+} p_{is} r_s(a_t, z) J_{t+1}\left(P_{t+1}^{1,z,+}, \bar{\Delta}_{t+1}^z\right) \geq \sum_z \sum_s \sum_i p_t^{i,-} p_{is} r_s(a_t, z) J_{t+1}\left(P_{t+1}^{1,z,-}, \bar{\Delta}_{t+1}^z\right),$$

which concludes the proof.

APPENDIX D
PROOF OF LEMMA 4

We prove Lemma 4 using induction. At the final stage, $t = N$, we have that $J_N(P, \bar{\Delta}) = \lambda_1 H(P^h) + \lambda_2 \bar{\Delta}$ which is linear in $H(P^h)$ and $\bar{\Delta}$. For stage $t = N-1$, we have that,

$$J_{N-1}(x_{N-1}) = \min\left[A_{0,N-1}, A_{1,N-1}\right], \tag{27}$$

and

$$A_{0,N-1} = \lambda_1 H\left(P_{N-1}^h\right) + \lambda_2 \bar{\Delta}_{N-1} + \sum_z \sum_s \sum_i p_{N-1}^i p_{is} r_s(0,z) J_N\left(P_N^{0,z}, \bar{\Delta}_N^z\right), \tag{28}$$

$$A_{1,N-1} = c + \lambda_1 H\left(P_{N-1}^h\right) + \lambda_2 \bar{\Delta}_{N-1} + \sum_z \sum_s \sum_i p_{N-1}^i p_{is} r_s(1,z) J_N\left(P_N^{1,z}, \bar{\Delta}_N^z\right). \tag{29}$$

By the minimization operator in Equation (27), the fact that $A_{0,N-1}$ and $A_{1,N-1}$ are linear and increasing in $H(P)$ and $\bar{\Delta}$ we have $J_{N-1}(x)$ is piece-wise linear, increasing and concave with respect to $H(P)$ and $\bar{\Delta}$. Assuming that $J_{t+1}(P_{t+1}, \bar{\Delta}_{t+1})$ is piece-wise linear, increasing and concave with respect to $H(P_{t+1})$ and $\bar{\Delta}_{t+1}$ then we will show that $J_t(P_t, \bar{\Delta}_t)$ will also be piece-wise linear, increasing and concave with respect to $H(P_{t+1})$ and $\bar{\Delta}_{t+1}$. For $J_t(P_t, \bar{\Delta}_t)$ we have,

$$J_t(x_t) = \min[A_0, A_1], \tag{30}$$

and

$$A_{0,t} = \lambda_1 H\left(P_t^h\right) + \lambda_2 \bar{\Delta}_t + \sum_z \sum_s \sum_i p_t^i p_{is} r_s(0,z) J_{t+1}\left(P_{t+1}^{0,z}, \bar{\Delta}_{t+1}^z\right), \tag{31}$$

$$A_{1,t} = c + \lambda_1 H\Big(P_t^h\Big) + \lambda_2 \bar{\Delta}_t + \sum_z \sum_s \sum_i p_t^i p_{is}\, r_s(1,z)\, J_{t+1}\Big(P_{t+1}^{1,z}, \bar{\Delta}_{t+1}^z\Big). \quad (32)$$

From (31) and (32) and the induction hypothesis we have that $A_{0,t}$ and $A_{1,t}$ are piece-wise linear, increasing and concave with respect to $H(P_{t+1}^h)$ and $\bar{\Delta}_{t+1}$. Finally, considering (30) we have that $J_t(x)$ is piece-wise linear, increasing and concave with respect to $H(P^h)$ and $\bar{\Delta}$ for $t = 1, \dots, N$.

APPENDIX E PROOF OF THEOREM 1

In order to prove Theorem 1 we will show that there exist positive probing cost values c such that the optimal action at decision stage $t = N-1, \dots, 1, 0$ is the probe action. Subsequently we will show that, if the optimal action is the probe action for a given state $x_t = [P_t, \bar{\Delta}_t]$ and cost value c then the optimal action will be the probe action for all states x_t with higher health status entropy. With similar arguments it can be shown that the probe action will also be optimal for states with higher normalized AoI than x_t. As a result, at each decision stage t, the optimal policy will be threshold based with respect to V_t.

The Bellman Equation (15), indicates that the probe action will be optimal if the following inequality is true,

$$c + V_t + \sum_{z,s,i} p_t^i p_{is}\, r_s(1,z)\, J_{t+1}\Big(P_{t+1}^{1,z}, \bar{\Delta}_{t+1}^z\Big) \leq V_t + \sum_{z,s,i} p_t^i p_{is}\, r_s(0,z)\, J_{t+1}\Big(P_{t+1}^{0,z}, \bar{\Delta}_{t+1}^z\Big), \quad (33)$$

where we have substituted x_{t+1} with its constituent elements P_{t+1} and $\bar{\Delta}_{t+1}$. The superscripts a, z in $P_{t+1}^{a,z}$ present action a_t and observation z_{t+1}, respectively, which determine the evolution of the belief state from time slot t to time slot $t+1$. Similarly, the superscript z in $\bar{\Delta}_{t+1}^z$ presents the observation z_{t+1} that determines the evolution of the normalized AoI from time slot t to time slot $t+1$. With simple algebraic manipulations Equation (33) can be written as,

$$\begin{aligned} c \leq \sum_s \sum_i p_t^i p_{is} \Big[r_s(0,0)\, J_{t+1}\Big(P_{t+1}^{0,0}, \bar{\Delta}_{t+1}^0\Big) &- r_s(1,0)\, J_{t+1}\Big(P_{t+1}^{1,0}, \bar{\Delta}_{t+1}^0\Big)\Big] \\ + \sum_s \sum_i p_t^i p_{is} \Big[r_s(0,1)\, J_{t+1}\Big(P_{t+1}^{0,1}, \bar{\Delta}_{t+1}^1\Big) &- r_s(1,1)\, J_{t+1}\Big(P_{t+1}^{1,1}, \bar{\Delta}_{t+1}^1\Big)\Big]. \end{aligned}$$

Substituting $r_s(a,z)$ for each combination of s, a and z as presented in Table 1 we get,

$$\begin{aligned} c \leq \sum_i p_t^i \Big[& p_{i0}\Big[(1-P_g) J_{t+1}\Big(P_{t+1}^{0,0}, \bar{\Delta}_{t+1}^0\Big)\Big] \\ & + p_{i4}(1-P_g)\Big[J_{t+1}\Big(P_{t+1}^{0,0}, \bar{\Delta}_{t+1}^0\Big) - J_{t+1}\Big(P_{t+1}^{1,0}, \bar{\Delta}_{t+1}^0\Big)\Big] \\ & + p_{i0}\Big[P_g J_{t+1}\Big(P_{t+1}^{0,1}, \bar{\Delta}_{t+1}^1\Big) - J_{t+1}\Big(P_{t+1}^{1,1}, \bar{\Delta}_{t+1}^1\Big)\Big] \\ & + p_{i4} P_g \Big[J_{t+1}\Big(P_{t+1}^{0,1}, \bar{\Delta}_{t+1}^1\Big) - J_{t+1}\Big(P_{t+1}^{1,1}, \bar{\Delta}_{t+1}^1\Big)\Big] \\ & + \sum_{s \in S \setminus \{0,4\}} p_{is}\Big[J_{t+1}\Big(P_{t+1}^{0,0}, \bar{\Delta}_{t+1}^0\Big) - J_{t+1}\Big(P_{t+1}^{1,0}, \bar{\Delta}_{t+1}^0\Big)\Big]\Big], \end{aligned} \quad (34)$$

where $S = \{0, 1, \dots, 7\}$. If the right hand side of (34) is greater than zero then there exists a probing cost c such that the optimal action is the probe action. Given that the transition cost is positive for all decision stages, i.e., $g_t > 0$ for $t = N-1, \dots, 0, 1$, we have that the minimum expected cost from decision stage $t+1$ up to the last decision stage $N-1$ will be greater than 0, i.e., $J_{t+1}(P_{t+1}^{0,0}, \bar{\Delta}_{t+1}^0) > 0$ and consequently the first term in (34) is strictly positive. It remains to show that, $J_{t+1}(P_{t+1}^{0,0}, \bar{\Delta}_{t+1}^0) - J_{t+1}(P_{t+1}^{1,0}, \bar{\Delta}_{t+1}^0) \geq 0$ and $J_{t+1}(P_{t+1}^{0,1}, \bar{\Delta}_{t+1}^1) - J_{t+1}(P_{t+1}^{1,1}, \bar{\Delta}_{t+1}^1) \geq 0$. From Lemma 1 we have that $H(P_{t+1}^{h,0,0}) \geq H(P_{t+1}^{h,1,0})$ and $H(P_{t+1}^{h,0,1}) \geq H(P_{t+1}^{h,1,1})$ and according to Lemma 2 we have that $J_k(P, \bar{\Delta})$ is an increasing function of the health status entropy for all decision stages, thus the above inequalities are true.

Now assume that the probe action is optimal for the state $x_t = [P_t, \bar{\Delta}_t]$ and let $x_t^+ = [P_t^+, \bar{\Delta}_t]$ be such that $H(P_t^{h,+}) \geq H(P_t^h)$. Furthermore, let $P_{t+1}^{a,z,+}$ be the belief state at $t+1$ given that the belief state at t was P_t^+, action a was taken at time t and observation z was made at $t+1$. From Lemma 1 we have that probing results in a reduction of entropy, i.e., $H(P_{t+1}^{h,0,0,+}) \geq H(P_{t+1}^{h,1,0,+})$ and $H(P_{t+1}^{h,0,1,+}) \geq H(P_{t+1}^{h,1,1,+})$, from Lemma 2 and Lemma 4 we have that $J_t(\cdot)$, $t = 0, 1, \dots, N$ is piece-wise linear, increasing and concave with respect entropy, thus the following inequalities, i) $J_{t+1}(P_{t+1}^{0,0,+}, \bar{\Delta}_{t+1}^0) \geq 0$ ii) $J_{t+1}(P_{t+1}^{0,0,+}, \bar{\Delta}_{t+1}^0) - J_{t+1}(P_{t+1}^{1,0,+}, \bar{\Delta}_{t+1}^0) \geq 0$ iii) $J_{t+1}(P_{t+1}^{0,1,+}, \bar{\Delta}_{t+1}^1) - J_{t+1}(P_{t+1}^{1,1,+}, \bar{\Delta}_{t+1}^1) \geq 0$ hold. As a result there exists a probing cost $c > 0$ such that the probing action is optimal for x_t^+ as well. What is more, according to Lemmas 3 and 4, $J_t(\cdot)$ is also increasing in $\bar{\Delta}_t$ and, as a consequence, it will also be increasing when both $\bar{\Delta}_t$ and P_t increase. Thus, the same proposition as above will be true for all states $x_t^+ = [P_t^{h,+}, \bar{\Delta}_t^+]$ such that $H(P_t^{h,+}) \geq H(P_t^h)$ and $\bar{\Delta}_t^+ \geq \bar{\Delta}_t$.

ACKNOWLEDGMENT

Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

[1] G. J. Stamatakis, N. Pappas, A. Fragkiadakis, and A. Traganitis, "Semantics-aware active fault detection in IoT," in *Proc. 20th Int. Symp. Model. Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, 2022, pp. 161–168.

[2] T. Muhammed and R. A. Shaikh, "An analysis of fault detection strategies in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 78, pp. 267–287, Jan. 2017.

[3] Z. Zhang, A. Mehmood, L. Shu, Z. Huo, Y. Zhang, and M. Mukherjee, "A survey on fault diagnosis in wireless sensor networks," *IEEE Access*, vol. 6, pp. 11349–11364, 2018.

[4] S. L. Campbell and R. Nikoukhah, *Auxiliary Signal Design for Failure Detection*, vol. 11. Princeton, NJ, USA: Princeton Univ. Press, 2015.

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

[6] P. Popovski et al., "A perspective on time toward wireless 6G," *Proc. IEEE*, vol. 110, no. 8, pp. 1116–1146, Aug. 2022.

[7] P. Kutsevol, O. Ayan, N. Pappas, and W. Kellerer, "Experimental study of transport layer protocols for wireless networked control systems," in *Proc. 20th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, 2023, pp. 438–446.

[8] E. Effah and O. Thiare, "Survey: Faults, fault detection and fault tolerance techniques in wireless sensor networks," *Int. J. Comput. Sci. Inf. Secur.*, vol. 16, no. 10, pp. 1–14, 2018.

[9] A. Mehmood, N. Alrajeh, M. Mukherjee, S. Abdullah, and H. Song, "A survey on proactive, active and passive fault diagnosis protocols for WSNS: Network operation perspective," *Sensors*, vol. 18, no. 6, p. 1787, 2018.

[10] G. Stamatakis, N. Pappas, A. Fragkiadakis, and A. Traganitis, "Autonomous maintenance in IoT networks via AoI-driven deep reinforcement learning," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–7.

[11] M. M. H. Khan, H. K. Le, H. Ahmadi, T. F. Abdelzaher, and J. Han, "Dustminer: Troubleshooting interactive complexity bugs in sensor networks," in *Proc. 6th ACM Conf. Embedded Netw. Sensor Syst.*, 2008, pp. 99–112. [Online]. Available: https://doi.org/10.1145/1460412.1460423

[12] J. Yang, M. L. Soffa, L. Selavo, and K. Whitehouse, "Clairvoyant: A comprehensive source-level debugger for wireless sensor networks," in *Proc. 5th Int. Conf. Embedded Netw. Sens. Syst.*, 2007, pp. 189–203. [Online]. Available: https://doi.org/10.1145/1322263.1322282

[13] T. A. N. Heirung and A. Mesbah, "Input design for active fault diagnosis," *Annu. Rev. Control*, vol. 47, pp. 35–50, Jun. 2019.

[14] R. A. Howard, "Information value theory," *IEEE Trans. Syst. Sci. Cybern.*, vol. 2, no. 1, pp. 22–26, Aug. 1966.

[15] B. Juba and M. Sudan, "Universal semantic communication i," in *Proc. 40th Annu. ACM Symp. Theory Comput.*, 2008, pp. 123–132. [Online]. Available: https://doi.org/10.1145/1374376.1374397

[16] J. Bao et al., "Towards a theory of semantic communication," in *Proc. IEEE Netw. Sci. Workshop*, 2011, pp. 110–117.

[17] B. Güler, A. Yener, and A. Swami, "The semantic communication game," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 4, pp. 787–802, Dec. 2018.

[18] O. Goldreich, B. Juba, and M. Sudan, "A theory of goal-oriented communication," *J. ACM*, vol. 59, no. 2, p. 8, May 2012. [Online]. Available: https://doi.org/10.1145/2160158.2160161

[19] P. Popovski, O. Simeone, F. Boccardi, D. Gündüz, and O. Sahin, "Semantic-effectiveness filtering and control for post-5G wireless connectivity," *J. Indian Inst. Sci.*, vol. 100, pp. 435–443, May 2020.

[20] E. C. Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Comput. Netw.*, vol. 190, May 2021, Art. no. 107930. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128621000773

[21] G. Stamatakis, N. Pappas, and A. Traganitis, "Control of status updates for energy harvesting devices that monitor processes with alarms," in *Proc. IEEE Globecom Workshops*, 2019, pp. 1–6.

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

[23] Q. Lan et al., "What is semantic communication? A view on conveying meaning in the era of machine intelligence," *J. Commun. Inf. Netw.*, vol. 6, no. 4, pp. 336–371, 2021.

[24] J. Dommel, D. Wieruch, Z. Utkovski, and S. Stańczak, "A semantics-aware communication scheme to estimate the empirical measure of a quantity of interest via multiple access fading channels," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, 2021, pp. 521–525.

[25] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 210–219, Feb. 2022.

[26] Z. Utkovski et al., "Semantic communication for edge intelligence: Theoretical foundations and implications on protocols," *IEEE Internet Things Mag.*, vol. 6, no. 4, pp. 48–53, Dec. 2023.

[27] G. Zhang, Q. Hu, Z. Qin, Y. Cai, G. Yu, and X. Tao, "A unified multi-task semantic communication system for multimodal data," 2022, *arXiv:2209.07689*.

[28] S. Guo, Y. Wang, S. Li, and N. Saeed, "Semantic importance-aware communications using pre-trained language models," *IEEE Commun. Lett.*, vol. 27, no. 9, pp. 2328–2332, Sep. 2023.

[29] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Found. Trends Netw.*, vol. 12, no. 3, pp. 162–259, 2017.

[30] Y. Sun, I. Kadota, R. Talak, and E. Modiano, "Age of information: A new metric for information freshness," *Synth. Lectures Commun. Netw.*, vol. 12, no. 2, pp. 1–224, 2019.

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

[32] M. Moltafet, M. Leinonen, M. Codreanu, and N. Pappas, "Power minimization for age of information constrained dynamic control in wireless sensor networks," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 419–432, Jan. 2022.

[33] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The cost of delay in status updates and their value: Non-linear ageing," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4905–4918, Aug. 2020.

[34] X. Wang, W. Lin, C. Xu, X. Sun, and X. Chen, "Age of changed information: Content-aware status updating in the Internet of Things," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 578–591, Jan. 2022.

[35] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the Internet of Things," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7468–7482, Nov. 2019.

[36] F. Wu, H. Zhang, J. Wu, Z. Han, H. V. Poor, and L. Song, "UAV-to-device underlay communications: Age of information minimization by multi-agent deep reinforcement learning," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4461–4475, Jul. 2021.

[37] M. A. Abd-Elmagid, H. S. Dhillon, and N. Pappas, "A reinforcement learning framework for optimizing age of information in RF-powered communication systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4747–4760, Aug. 2020.

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

[39] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, 2012, pp. 2731–2735.

[40] D. Bertsekas, *Dynamic Programming and Optimal Control*, 4th ed., vol. 1. Belmont, MA, USA: Athena Sci., 2017.

[41] V. Krishnamurthy, *Partially Observed Markov Decision Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[42] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, vol. 65. Hoboken, NJ, USA: Wiley, 2005.

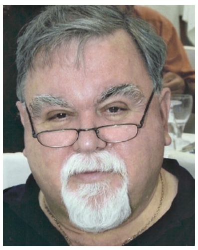
[43] S. Ross, *A First Course in Probability*. Hoboken, NJ, USA: Prentice Hall, 2010.

GEORGE STAMATAKIS received the bachelor's degree in electronics and computer engineering from the Technical University of Crete in 2001, and the M.Sc. degree in computer science from the University of Crete in 2005. Since 2003, he has been a member of the Telecommunications and Networks Laboratory, ICS-FORTH and has participated in numerous National and European projects. His research interests are in the area of optimal control and resource allocation in communication systems and networks with an emphasis on the development of intelligent, autonomous networks by utilizing dynamic programming, and reinforcement learning techniques.

NIKOLAOS PAPPAS (Senior Member, IEEE) received the first B.Sc. degree in computer science, the second B.Sc. degree in mathematics, the M.Sc. degree in computer science, and the Ph.D. degree in computer science from the University of Crete, Greece, in 2005, 2012, 2007, and 2012, respectively. During his Ph.D. studies, he was a Visiting Scholar with the Institute of Systems Research, University of Maryland at College Park, College Park, MD, USA. From 2012 to 2014, he was a Postdoctoral Researcher with the Department of Telecommunications, CentraleSupélec, Gif-sur-Yvette, France. He is currently an Associate Professor with the Department of Computer and Information Science, Linköping University, Linköping, Sweden. His main research interests include the field of wireless communication networks with an emphasis on semantics-aware communications, energy harvesting networks, network-level cooperation, age of information, and stochastic geometry. He has served as the Symposium Co-Chair of the IEEE International Conference on Communications in 2022 and the IEEE Wireless Communications and Networking Conference in 2022. He is an Area Editor of the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY and an Expert Editor for invited papers of the IEEE COMMUNICATIONS LETTERS. He is an Editor of the IEEE TRANSACTIONS ON MACHINE LEARNING IN COMMUNICATIONS AND NETWORKING and the IEEE/KICS JOURNAL OF COMMUNICATIONS AND NETWORKS. He is a Guest Editor of the IEEE NETWORK and the IEEE INTERNET OF THINGS MAGAZINE. He has served as an Editor of the IEEE COMMUNICATIONS LETTERS and the IEEE TRANSACTIONS ON COMMUNICATIONS. He was a Guest Editor of the IEEE INTERNET OF THINGS JOURNAL on "Age of Information and Data Semantics for Sensing, Communication and Control Co-Design in IoT."

ALEXANDROS FRAGKIADAKIS received the bachelor's degree in electronic and electrical engineering from the University of West Attica, the master's degree in digital communication systems (with Distinction) from Loughborough University, U.K., and the Ph.D. degree in computer networks from the Department of Electronic and Electrical Engineering, Loughborough University. He is a Researcher with the Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH), Greece. He has worked as a Research Associate with the High Speed Networks Group, Department of Electronic and Electrical Engineering, Loughborough University, and as a Laboratory Associate with the Department of Electrical and Computer Engineering, Hellenic Mediterranean University, Greece. He has also taught at the Hellenic Open University in one of its M.Sc. programs. At FORTH, he has participated in many research projects in the field of wireless communications and networking. His research interests include wireless networks, intrusion detection and security in wireless networks, reprogrammable devices, open source architectures, RF cognitive networks, wireless sensor networks, cloud infrastructure, and Internet of Things. He is actively involved in research projects related to the Internet of Things, in several of them as a Scientific Leader. He has several publications in book chapters, journals, and international conferences. He has served as a member of technical committees in a number of scientific conferences, and has served as an evaluator of research proposals at the European level.

NIKOLAOS PETROULAKIS (Member, IEEE) received the Diploma degree in mathematics from the Mathematics Department, National and Kapodistrian University of Athens, the M.Sc. degree in digital communications from the Department of Engineering and Design, University of Sussex, Brighton, U.K., and the Ph.D. degree in network security from the School of Mathematics, Computer Science and Engineering, City, University of London, U.K. He has been a Research Scientist with the Institute of Computer Science of Foundation for Research and Technology—Hellas (FORTH) since 2007. He has coordinated and participated in a number of European and national projects and proposals especially from the FP7, H2020, and Horizon Europe frameworks, being also the Project Coordinator of the Horizon Europe NextGEM Project. He is also interested to contribute the results of his research in different conferences and journals, having authored more than 50 scientific publications. His research interests resolve around wireless/wired networks, resource management, 5G networking, IoT, electromagnetic field, cybersecurity, semantic interoperability, and artificial intelligence. He is an IEEE ComSoc Member, having served as a reviewer and a member of the technical and organizing committees in various academic activities, including journal, conferences, and summer schools in his area of expertise.



APOSTOLOS TRAGANITIS received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1970, and the M.S. and Ph.D. degrees from Princeton University, Princeton, NJ, USA, in 1972 and 1974, respectively. He has been a Senior Researcher with the Hellenic Navy Research Laboratory, where he was in charge of the Electronic Warfare Unit. In 1993 and 1994, he was a Visiting Research Fellow with the Center of Satellite and Hybrid Communications Networks, Institute of Systems Research, University of Maryland at College Park, College Park, MD, USA. He is currently a Professor Emeritus with the Department of Computer Science, University of Crete, having served as the Chair from 2000 to 2002. He is also the Head of the Telecommunications and Networks Laboratory, Institute of Computer Science, Foundation for Research and Technology - Hellas, Heraklion, Greece. His research interests include digital communications, wireless networks, communications security, hardware design, and biomedical engineering.