Risk-Aware Optimization of Age of Information in the Internet of Things

Bo Zhou*, Walid Saad*, Mehdi Bennis†, and Petar Popovski‡
*Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA,
†Center for Wireless Communications, University of Oulu, Finland,
‡Department of Electronic Systems, Aalborg University, Aalborg, Denmark,
Emails: *{ecebo, walids}@vt.edu, †mehdi.bennis@oulu.fi, ‡petarp@es.aau.dk.

***Abstract*—Minimization of the expected value of age of information (AoI) is a *risk-neutral* approach, and it thus cannot capture rare, yet critical, events with potentially large AoI. In order to capture the effect of these events, in this paper, the notion of conditional value-at-risk (CVaR) is proposed as an effective coherent risk measure that is suitable for minimization of AoI for real-time IoT status updates. In the considered monitoring system, an IoT device monitors a physical process and sends the status updates to a remote receiver with an updating cost. The optimal status update process is designed to jointly minimize the AoI at the receiver, the CVaR of the AoI at the receiver, and the energy cost. This stochastic optimization problem is formulated as an infinite horizon discounted risk-aware Markov decision process (MDP), which is computationally intractable due to the time inconsistency of the CVaR. By exploiting the special properties of coherent risk measures, the risk-aware MDP is reduced to a standard MDP with an augmented state space, for which we derive the optimal stationary policy using dynamic programming. In particular, the optimal history-dependent policy of the risk-aware MDP is shown to depend on the history only through the augmented system states and can be readily constructed using the optimal stationary policy of the augmented MDP. The proposed solution is shown to be computationally tractable and able to minimize the AoI in real-time IoT monitoring systems in a risk-aware manner.**

I. Introduction

Time-sensitive Internet of Things (IoT) applications [1], such as real-time surveillance and monitoring, drone navigation, and autonomous driving, must rely on a timely delivery of status information updates of the physical processes that are being monitored or operated by the IoT devices for control and monitoring purposes. In light of this, the concept of *age of information* (AoI) has been recently proposed to evaluate the freshness of the status updates at the information destination (e.g., an IoT control center or base station) [2], [3]. The AoI is a performance metric that quantifies the time elapsed since the latest received status update at the information destination was generated. Since the AoI captures the information freshness from the perspective of the remote destination and depends on both the generation and transmission of the status updates, it is fundamentally different from conventional performance metrics, such as throughput or delay.

Recently, there has been a growing body of research on minimizing the AoI in various communication systems [4]–[9]. In [4], the authors study the optimal status sampling and updating policy to minimize the average AoI for an IoT monitoring system under device energy constraints. The problem of AoI minimization for IoT monitoring systems with non-uniform status packet sizes is studied in [5] and [6]. The works in [7], [8] investigate the problem of AoI minimization for wireless status updating systems with noisy channels. The authors in [9] propose an online sampling policy to minimize the average AoI for energy harvesting systems.

These existing works, e.g., [4]–[9], adopt a *risk-neutral* approach, by focusing only on minimizing the expected value of the (random) AoI cost functions, e.g., the average AoI, the average peak AoI, and the average age penalty. Although the obtained algorithms through this approach can result in AoI performance that is minimized over a long run, they do not capture the risk of the uncertainty of the AoI cost function, e.g., the variability of the AoI distribution and the effects of rare but potentially detrimental AoI events. For example, for safety and state monitoring in industrial production scenarios, a certain status update with a very large AoI could result in a complete shutdown of the production. Thus, it is critical to focus on the AoI not only in the average sense, but also in a risk-related sense. Recently, the works in [10] and [11] considered the tail of the AoI distribution (with extremely large AoI) for vehicular networks and wireless industrial networks, respectively. Meanwhile, the work in [12] analyzed the violation probability of the peak AoI for a point-to-point communication system with short packets. However, these approaches in [10]–[12] focus on the probability that the peak AoI exceeds a certain threshold, and, thus, they can neither quantify nor minimize the expected losses that might be incurred in tail events in which the AoI is very large. Clearly, how to design the optimal status updating policy so as to jointly minimize the average AoI and the expected tail loss of the AoI, remains an open problem.

The main contribution of this paper is a novel design of a risk-aware status updating control policy that jointly minimizes the AoI at the receiver, the expected tail loss of the AoI at the receiver, and the energy cost, for a real-time IoT monitoring system. Specifically, we use a popular and

This research was supported by the U.S. National Science Foundation under Grant CNS-1836802, the Academy of Finland Project CARMA, the Academy of Finland Project MISSION, the Academy of Finland Project SMARTER, the INFOTECH Project NOOR, the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (ERC Consolidator Grant Nr. 648382 WILLOW), and the Danish Council for Independent Research (Grant Nr. 8022-00284B SEMIOTIC).

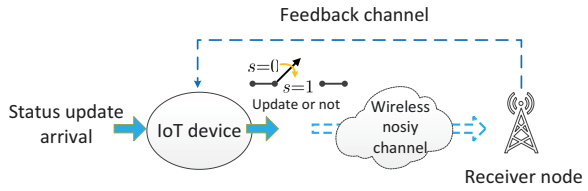

Fig. 1: Illustration of a real-time monitoring system.

effective coherent risk measure, called the conditional value-at-risk (CVaR) [13], to measure the *tail average* of the AoI distribution exceeding a given risk level. We formulate this stochastic control problem as an infinite horizon discounted risk-aware Markov decision process (MDP) and seek the optimal history-dependent updating control policy. This risk-aware MDP is challenging to solve due to two reasons: 1) Because of the time inconsistency of the CVaR, dynamic programming cannot be directly applied and 2) History-dependent policies are generally intractable due to the substantial requirements on the computation time and memory. By exploiting the dual representation and the temporal decomposition properties of the coherent risk measures, we reduce the risk-aware MDP to a standard MDP on the state space augmented by a two-dimensional risk level space and propose a dynamic programming based solution to derive the *optimal stationary policy* through a risk-aware Bellman operator. Thus, instead of working on the intractable space of history-dependent policies, it is sufficient to focus on the optimization over stationary policies of the augmented MDP. In particular, we show that the optimal history-dependent policy depends on the history only through the dynamics of the two risk levels, and can be constructed with the optimal stationary policy for the augmented MDP. The proposed solution can explicitly account for rare events with very large AoI in an IoT monitoring system and is computationally tractable to obtain the generally intractable history-dependent policies.

II. System Model

We consider a general real-time IoT monitoring system composed of an IoT device and a remote receiver node (see Fig. 1). The IoT device monitors an underlying time-varying physical process and sends the associated real-time status information to the receiver. We assume that the status information updates of the underlying process arrive at the IoT device stochastically and are queued at the device before being transmission to the receiver. We consider a discrete-time system with time slots indexed by $t = 0, 1, 2, \cdots$. At the beginning of each time slot, the status update (if any) of the underlying process arrives at the IoT device randomly. Similar to [5] and [6], the process of the status update arrivals is modeled by an independent and identically distributed (i.i.d.) Bernoulli process with mean rate $\lambda \in [0, 1]$. The device is equipped with a buffer to store the arriving status update and the newly arriving most up-to-date status update will replace the older one (if any) in the buffer, as the receiver will not benefit from obtaining an outdated status update. Hence, there is at most one status update at the device.

We consider a wireless packet erasure channel between the IoT device and the receiver, and, upon transmission, each status update will be successfully delivered to the receiver with probability p. As in [6]–[8], we further assume that the IoT device will be notified immediately upon a successful transmission, through a perfect feedback channel between the device and the receiver.

A. Monitoring Model

Due to the possible failure of each transmission, the status update currently in the buffer at the device may be outdated at the receiver. Thus, in each slot, the IoT device must decide whether to transmit the locally available status update or stay idle to wait for a possibly arriving fresher status update. Let $s_t \in \mathcal{S} \triangleq \{0, 1\}$ be the updating control action of the device at slot t, where $s_t = 1$ implies that the device transmits its locally available status update at slot t and $s_t = 0$ indicates the device stays idle. $\mathcal{S}$ denotes the control action space. Let C be the energy cost for transmitting a status update.

B. Age of Information Model

We adopt the AoI as the key performance metric to quantify the freshness of the status information update at the receiver. The AoI is defined as time elapsed since the most recent status update delivered at the receiver. Let $A_{r,t}$ be the AoI at the receiver at the beginning of time slot t. By definition, we have $A_{r,t} = t - U_t^r$, where U_t^r is the time stamp of the freshest status update that was delivered to the receiver before t. Note that, the device can only transmit its currently available status update to the receiver and, thus, the AoI at the receiver depends on the age of the status update in the buffer at the device. We define $A_{d,t}$ as the AoI at the device at the beginning of slot t, to capture the freshness of the status information update at the device. Let $\hat{A}_d$ and $\hat{A}_r$ be the upper limits of the AoI at the device and the AoI at the receiver, respectively. Since a status update with an infinite age is not meaningful for real-time IoT monitoring systems, we assume that $\hat{A}_d$ and $\hat{A}_r$ are finite. Mathematically, $\hat{A}_d$ and $\hat{A}_r$ can be arbitrarily large. Let $\mathcal{A}_d \triangleq \{1, 2, \cdots, \hat{A}_d\}$ and $\mathcal{A}_r \triangleq \{1, 2, \cdots, \hat{A}_r\}$ be, respectively, the state space of the AoI at the device and the AoI at the receiver. We denote by $\boldsymbol{A}_t \triangleq (A_{d,t}, A_{r,t}) \in \mathcal{A} \triangleq \mathcal{A}_d \times \mathcal{A}_r$ the system AoI state at slot t, where $\mathcal{A}$ is the system AoI state space.

Now, we present how $\boldsymbol{A}_t$ evolves with the updating control action s_t. For the AoI at the device, if there is a status update arriving at the device during slot t, the AoI at the device will be reset to one, otherwise, the AoI will increase by one. Then, the dynamics of $A_{d,t}$ will be given by:

$$A_{d,t+1} = \begin{cases} 1, & \text{if an update arrives at } t, \\ \min\{A_{d,t} + 1, \hat{A}_d\}, & \text{otherwise.} \end{cases} \quad (1)$$

For the AoI at the receiver, if the device transmits the status update to the receiver at slot t and the transmission is successful, then the AoI at the receiver in the next slot will be the current AoI at the device plus one (due to the one slot transmission), otherwise, the AoI will increase by one. Note

that, the latter case includes the scenarios in which, the device attempts to send the status update while fails, or the device decides to stay idle. Thus, we have the dynamics of $A_{r,t}$:

$$A_{r,t+1} = \begin{cases} \min\{A_{d,t}+1, \hat{A}_r\}, & \text{if } s_t = 1 \text{ and the update} \\ & \text{transmission succeeds at } t, \\ \min\{A_{r,t}+1, \hat{A}_r\}, & \text{otherwise.} \end{cases} \quad (2)$$

By comparing (1) and (2), we observe that $A_{r,t} \geq A_{d,t}$ holds for all t, and, hence, we only need to focus on the system AoI state space $\mathcal{A}$ with $A_r \geq A_d$. Moreover, we also see that if $A_{d,t} = A_{r,t}$ for some t, then there is no need to choose the action $s_t = 1$, as the currently available status update at the device has already been delivered to the receiver before t.

III. Problem Formulation

The existing literature, e.g., [4]–[9], focuses only on minimizing the expected value of the (random) AoI cost function, and, thus, fails to capture the variability of the AoI distribution and accounts for rare events with potentially very large AoI. Hence, we consider a *risk-aware* approach, by taking into account the expected value of the AoI and the expected tail loss of the AoI based on the CVaR [13] – a popular and effective risk measure.

A. Preliminaries on CVaR and risk measures

For a bounded-mean random variable Z on a probability space $(\Omega, \mathcal{F}, P)$, the CVaR of Z at risk level $\alpha \in (0,1]$ is defined as the expectation of Z in its α-tail distribution [13]:

$$\text{CVaR}_\alpha(Z) \triangleq \min_{q\in\mathbb{R}} \left\{ q + \frac{1}{\alpha}\mathbb{E}[\max\{Z-q, 0\}] \right\}, \quad (3)$$

where the expectation is taken over the probability distribution P. Note that, $\text{CVaR}_\alpha(Z)$ decreases with α, $\text{CVaR}_1(Z) = \mathbb{E}[Z]$, and $\lim_{\alpha\to 0}\text{CVaR}_\alpha(Z) = \sup(Z)$. Thus, α can be seen as a kind of degree of risk aversion. It has been shown that the CVaR is a *coherent* risk measure [14]–[16].

Definition 1: A *coherent risk measure* $\rho(Z)$ is a mapping from the space $\mathcal{Z}$ of the random variable Z to $\mathbb{R}$ that obeys the following four axioms [14]–[16]. For any $Z, Z' \in \mathcal{Z}$:

1) Monotonicity: if $Z \leq Z'$, then $\rho(Z) \leq \rho(Z')$;
2) Subadditivity: $\rho(Z+Z) \leq \rho(Z) + \rho(Z')$;
3) Translation invariance: $\rho(Z+a) = \rho(Z) + a, \forall a \in \mathbb{R}$;
4) Positive homogeneity: if $b > 0$, then $\rho(bZ) = b\rho(Z)$.

Note that, based on the translation invariance and positive homogeneity axioms of a coherent risk measure ρ, we can easily obtain $\rho(c) = c$ for all constants c. One important result in risk measure theory is that each coherent risk measure has its dual representation as the maximum of certain expected value over a risk envelope [15, Theorem 6.4], i.e.,

$$\rho(Z) = \max_{\xi\in\Xi} \mathbb{E}_\xi[Z], \quad (4)$$

where $\mathbb{E}_\xi[Z] \triangleq \sum_{\omega\in\Omega}\xi(\omega)P(\omega)Z(\omega)$ denotes the ξ-weighed expectation of Z and Ξ is a specific set of probability density functions, referred to as the risk envelop. For example, the risk envelop of the CVaR is $\Xi = \{\xi : \xi(\omega) \in [0, 1/\alpha], \forall\omega \in \Omega, \text{ and } \sum_{\omega\in\Omega}\xi(\omega)P(\omega) = 1\}$.

B. Risk-Aware MDP Formulation

We consider history-dependent updating control policies, and the updating action at each time slot depends on the past history of the system, as represented by the sequence of the previous system AoI states and updating actions. For each time slot $t = 0, 1, \cdots$, let $\boldsymbol{h}_t \triangleq (\boldsymbol{A}_0, s_0, \boldsymbol{A}_1, s_1, \cdots, \boldsymbol{A}_{t-1}, s_{t-1}, \boldsymbol{A}_t) \in \mathcal{H}_t$ be the history up to slot t, which satisfies the recursion $\boldsymbol{h}_t = (h_{t-1}, s_{t-1}, \boldsymbol{A}_t)$ for all $t \geq 1$. Here, $\mathcal{H}_t$ is the space of all histories up to slot t, where $\mathcal{H}_0 \triangleq \mathcal{A}$ and $\mathcal{H}_t \triangleq \mathcal{H}_{t-1}\times\mathcal{S}\times\mathcal{A}$ for all $t \geq 1$.

Definition 2: A *history-dependent updating control policy* π is a sequence of decision rules for each time slot, i.e., $\pi \triangleq (\mu_0, \mu_1, \cdots)$, where μ_t is a mapping from the set of histories $\mathcal{H}_t$ at slot t to the control action space $\mathcal{S}$, i.e., $s_t = \mu_t(\boldsymbol{h}_t)$. Let Π_H be the set of all history-dependent policies π.

By the dynamics in (1) and (2), and the i.i.d. assumptions on the status updates arrival process, the induced random process $\{\boldsymbol{A}_t\}_{t=0,1,\cdots}$ under a history-dependent policy π is a controlled Markov chain, with the transition probability:

$$\begin{aligned} &\Pr[\boldsymbol{A}'|\boldsymbol{A}, s] \quad (5)\\ &= \Pr[\boldsymbol{A}_{t+1} = \boldsymbol{A}'|\boldsymbol{A}_t = \boldsymbol{A}, s_t = s] \\ &= \begin{cases} 1-\lambda, & \text{if } \boldsymbol{A}' = (A_d^0, A_r^0) \text{ and } s = 0, \\ \lambda, & \text{if } \boldsymbol{A}' = (A_d^1, A_r^0) \text{ and } s = 0, \\ (1-\lambda)p, & \text{if } \boldsymbol{A}' = (A_d^0, A_r^1) \text{ and } s = 1, \\ (1-\lambda)(1-p), & \text{if } \boldsymbol{A}' = (A_d^0, A_r^0) \text{ and } s = 1, \\ \lambda p, & \text{if } \boldsymbol{A}' = (A_d^1, A_r^1) \text{ and } s = 1, \\ \lambda(1-p), & \text{if } \boldsymbol{A}' = (A_d^1, A_r^0) \text{ and } s = 1, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

A_d^0 and A_d^1 are for the cases that a new status update arrives and no status update arrives, respectively. A_r^0 is for the case that either the device stays idle or the transmission fails, and A_r^1 is for the case that the transmission succeeds. From (1) and (2), we have $A_d^0 = \min\{A_d+1, \hat{A}_d\}$, $A_d^1 = 1$, $A_r^0 = \min\{A_r+1, \hat{A}_r\}$, and $A_r^1 = \min\{A_d+1, \hat{A}_r\}$.

For a given history-dependent policy π, an initial system AoI state $\boldsymbol{A}$, and a discount factor $\gamma \in (0,1)$, the infinite horizon expected total discounted AoI at the receiver and the infinite horizon expected total discounted energy cost are, respectively, given by:

$$\bar{A}_{r,\pi}^\gamma(\boldsymbol{A}) \triangleq \mathbb{E}\left[\limsup_{T\to\infty}\sum_{t=0}^{T}\gamma^t A_{r,t}|\boldsymbol{A}_0 = \boldsymbol{A}, \pi\right], \quad (6)$$

$$\bar{C}_\pi^\gamma(\boldsymbol{A}) \triangleq \mathbb{E}\left[\limsup_{T\to\infty}\sum_{t=0}^{T}\gamma^t s_t C|\boldsymbol{A}_0 = \boldsymbol{A}, \pi\right], \quad (7)$$

where the expectation is taken under the measure induced by policy π. By using the discounted AoI and energy cost, we weight the immediate cost more heavily than expected future costs. We use CVaR to capture the expected tail loss of the infinite horizon total discounted AoI at the receiver, given by:

$$\rho_\pi^\gamma(\boldsymbol{A}) \triangleq \text{CVaR}_\alpha\left(\limsup_{T\to\infty}\sum_{t=0}^{T}\gamma^t A_{r,t}|\boldsymbol{A}_0 = \boldsymbol{A}, \pi\right), \quad (8)$$

where $\alpha \in (0,1]$ is the risk level. Note that, $\gamma \in (0,1)$ ensures that $\bar{A}^{\gamma}_{r,\pi}(\boldsymbol{A})$, $\bar{C}^{\gamma}_{\pi}(\boldsymbol{A})$, and $\rho^{\gamma}_{\pi}(\boldsymbol{A})$ are upper-bounded.

Our goal is to find the optimal history-dependent policy that jointly minimizes the infinite horizon expected total discounted AoI at the receiver, the infinite horizon expected total discounted energy cost, and the CVaR of the infinite horizon total discounted AoI at the receiver. By adopting the weighted-sum method, which a widely used method for multi-objective optimization problem [17], we formulate the following problem:

$$\min_{\pi \in \Pi_H} \bar{A}^{\gamma}_{r,\pi}(\boldsymbol{A}) + \eta \rho^{\gamma}_{\pi}(\boldsymbol{A}) + \nu \bar{C}^{\gamma}_{\pi}(\boldsymbol{A}), \tag{9}$$

where $\boldsymbol{A}$ is a given initial system AoI state, and $\eta, \nu \geq 0$ are the weighing factors on the CVaR of the AoI and the energy cost. η and ν can be regarded as the penalty factors, mimicking the soft constraints on the CVaR of the AoI and the energy cost. Thus, we can think η and ν as the corresponding Lagrange multipliers.

We refer to problem (9) as an infinite horizon discounted risk-aware MDP. Note that, for standard MDPs with expected cost objectives (e.g., [18]), it is generally sufficient to focus on the optimization over deterministic stationary Markovian policies without loss of optimality. However, for the considered risk-aware MDP in (9), the more general class of history-dependent (non-stationary) policies could be required. This is because the CVaR measure is time-inconsistent, which intuitively implies that a policy that is optimal at the current stage is not necessarily optimal in subsequent stages [16]. Such time-inconsistency could further couple risk preferences over time, and, thus prevents us from directly applying dynamic programming to decompose the problem in stages [19].

IV. Optimal Risk-Aware AoI Solution

In general, computing the optimal history-dependent updating policy $\pi \in \Pi_H$ for the risk-aware MDP in (9) is practically intractable due to the substantial requirements in terms of memory and computation time. Inspired from [16], [20], we show that the risk-aware MDP in (9) can be reduced to a standard MDP with an augmented system state space, by exploiting the properties of dual representation and temporal decomposition of coherent risk measures. In particular, the optimal history-dependent policy for (9) depends on the history only through the augmented system states and can be constructed with the optimal stationary policy for the augmented MDP.

A. Reduction of a Risk-Aware MDP to an Augmented MDP

According to [15, Equation (6.69)], we know that, for any $\alpha, \beta \in (0,1]$, the coherent risk measure $\rho(Z) = (1-\beta)\mathbb{E}[Z] + \beta \text{CVaR}_{\alpha}(Z)$ has the dual representation in the form of (4), where the risk envelop is $\Xi = \{\xi : \xi(\omega) \in [1-\beta, 1+\beta(1/\alpha - 1)], \forall \omega \in \Omega, \text{ and } \sum_{\omega \in \Omega} \xi(\omega) P(\omega) = 1\}$. Then, we can transform $\bar{A}^{\gamma}_{r,\pi}(\boldsymbol{A}) + \eta \rho^{\gamma}_{\pi}(\boldsymbol{A})$ in the objective function of (9) to a coherent risk measure:

$$\begin{aligned}
&\bar{A}^{\gamma}_{r,\pi}(\boldsymbol{A}) + \eta \rho^{\gamma}_{\pi}(\boldsymbol{A}) \\
&= (1+\eta)\left((1 - \frac{\eta}{1+\eta}) \bar{A}^{\gamma}_{r,\pi}(\boldsymbol{A}) + \frac{\eta}{1+\eta} \rho^{\gamma}_{\pi}(\boldsymbol{A}) \right) \\
&\triangleq (1+\eta) \rho_{\delta,\phi}\Big(\limsup_{T \to \infty} \sum_{t=0}^{T} \gamma^t A_{r,t} | \boldsymbol{A}_0 = \boldsymbol{A}, \pi \Big),
\end{aligned} \tag{10}$$

where $\rho_{\delta,\phi}(\cdot)$ is a coherent risk measure with a risk envelop: $\Xi(\delta, \phi, P) = \{\xi : \xi(\omega) \in [\delta, 1/\phi], \forall \omega \in \Omega \text{ and } \sum_{\omega \in \Omega} \xi(\omega) P(\omega) = 1\}$, $\delta = \frac{1}{1+\eta} \in (0,1]$ and $\phi = \frac{\alpha(1+\eta)}{\alpha+\eta} \in (0,1]$. (δ, ϕ) are the risk levels of $\rho_{\delta,\phi}(\cdot)$.

Now, we present the key temporal decomposition property of the coherent risk measure. First, for each $k = 0, 1, \cdots$, we define $\bar{\pi}_k = (\bar{\mu}_t)_{t=k,k+1,\cdots}$ as a k-th tail history-dependent policy, where the action $\bar{\mu}_t$ at slot $t \geq k$ is a mapping from $\mathcal{H}_{k,t}$ to the control action space $\mathcal{S}$. Here, $\mathcal{H}_{k,t}$ denotes the set of all histories from slot k to slot t, satisfying $\mathcal{H}_{k,t+1} \triangleq \mathcal{H}_{k,t} \times \mathcal{S} \times \mathcal{A}$ for $t \geq k+1$ and $\mathcal{H}_{k,k} \triangleq \boldsymbol{A}$. A generic element $\boldsymbol{h}_{k,t}$ of $\mathcal{H}_{k,t}$ takes the form $\boldsymbol{h}_{k,t} \triangleq (\boldsymbol{A}_k, s_k, \cdots, \boldsymbol{A}_{t-1}, s_{t-1}, \boldsymbol{A}_t)$. From Definition 2, we know that $\bar{\pi}_0 = \pi$. Then, we have the following temporal decomposition of $\rho_{\delta,\phi}(\cdot)$, based on Theorem 2.6.1 in [20].

Lemma 1: Given slot k, system AoI state $\boldsymbol{A}_k \in \mathcal{A}$, control action $s_k \in \mathcal{S}$, and risk levels $(\delta_k, \phi_k) = (\delta, \phi) \in (0,1]^2$, for any $(k+1)$-th tail history-dependent policy $\bar{\pi}_{k+1}$, we have the following temporal decomposition property of the conditional coherent risk measure of $\rho_{\delta,\phi}(\cdot)$:

$$\begin{aligned}
&\rho_{\delta,\phi}(Z_{k+1} | \boldsymbol{A}_k, s_k, \bar{\pi}_{k+1}) = \max_{\xi \in \Xi(\delta,\phi,\Pr(\cdot|\boldsymbol{A}_k,s_k))} \\
&\mathbb{E}\left[\xi(\boldsymbol{A}_{k+1}) \rho_{\delta/\xi(\boldsymbol{A}_{k+1}), \phi\xi(\boldsymbol{A}_{k+1})}(Z_{k+1} | \boldsymbol{A}_{k+1}, \bar{\pi}_{k+1}) | \boldsymbol{A}_k, s_k \right],
\end{aligned}$$

where $Z_{k+1} \triangleq \limsup_{T \to \infty} \sum_{t=0}^{T} \gamma^t A_{r,t+k+1}$ denotes the (random) total discounted AoI at the receiver from time $k+1$ such that the system AoI state evolves under policy $\bar{\pi}_{k+1}$ via the AoI dynamics in (1) and (2) conditioned on $(\boldsymbol{A}_k, s_k)$, and the expectation is taken with respect to the probability distribution of $\boldsymbol{A}_{k+1}$ conditioned on $(\boldsymbol{A}_k, s_k)$.

Note that, the difference between Lemma 1 and Theorem 2.6.1 in [20] is that we remove the dependency on the history prior to time k. This is because $\boldsymbol{A}_k$, s_k, and (δ_k, ϕ_k) are given, Z_{k+1} is conditioned on $\boldsymbol{A}_{k+1}$, and the system AoI state is Markov. Based on the temporal decomposition of the coherent risk measure in Lemma 1, by following the state space augmentation approach in [20, Chapter 2], we augment the system AoI state space $\mathcal{A}$ to include additional two dimensional state space $\mathcal{X} \times \mathcal{Y} = (0,1]^2$, which correspond to the two risk levels (δ, ϕ). We refer to as $\mathcal{A} \times \mathcal{X} \times Y$ as the augmented system state space. The dynamics of the augmented system state $(\boldsymbol{A}, x, y) \in \mathcal{A} \times \mathcal{X} \times Y$ are as follows: The system AoI states $\{\boldsymbol{A}_t\}_{t=0,1,\cdots}$ still evolve as per the AoI dynamics in (1) and (2) as well as the transition probability in (5), and the evolution does not depend on the risk levels. The risk levels $\{x_t, y_t\}_{t=0,1,\cdots}$ evolve deterministically according to $x_{t+1} = x_t / \xi^*(\boldsymbol{A}_t, x_t, y_t, s_t)$ and $y_{t+1} = y_t \xi^*(\boldsymbol{A}_t, x_t, y_t, s_t)$, where $\xi^*(\cdot)$ is a known deterministic function that will be specified in (13). Now, we introduce a new class of policies with the augmented system state space.

Definition 3: An *augmented stationary updating control policy* $\tilde{\pi}$ is a sequence of decision rules for each time slot, i.e., $\tilde{\pi} = (\tilde{\mu}, \tilde{\mu}, \cdots)$, where $\tilde{\mu}$ is a mapping from the augmented system state space $\mathcal{A} \times \mathcal{X} \times Y$ to the control action space $\mathcal{S}$, i.e., $s = \tilde{\mu}(\boldsymbol{A}, x, y)$. Let $\tilde{\Pi}_S$ be the set of all augmented stationary policies $\tilde{\pi}$.

Given an augmented stationary policy $\tilde{\pi}$, an initial augmented system state $(\boldsymbol{A}, x, y)$, and a discounted factor γ, we define $\bar{A}^{\gamma}_{r,\tilde{\pi}}(\boldsymbol{A}, x, y)$, $\bar{C}^{\gamma}_{\tilde{\pi}}(\boldsymbol{A}, x, y)$, and $\rho^{\gamma}_{\tilde{\pi}}(\boldsymbol{A}, x, y)$, in the same manner as in (6)-(8), respectively, and formulate the corresponding augmented MDP as follows:

$$V^*(\boldsymbol{A}, x, y) \triangleq \min_{\tilde{\pi} \in \tilde{\Pi}_S} V_{\tilde{\pi}}(\boldsymbol{A}, x, y), \tag{11}$$

where $V_{\tilde{\pi}}(\boldsymbol{A}, x, y) \triangleq \bar{A}^{\gamma}_{r,\tilde{\pi}}(\boldsymbol{A}, x, y) + \eta \rho^{\gamma}_{\tilde{\pi}}(\boldsymbol{A}, x, y) + \nu \bar{C}^{\gamma}_{\tilde{\pi}}(\boldsymbol{A}, x, y)$. Next, we show that the optimal history-dependent updating control policy $\pi \in \Pi_H$ in Definition 2 for the risk-aware MDP in (9) can be constructed by obtaining the optimal augmented stationary updating control policy $\tilde{\pi} \in \tilde{\Pi}_S$ in Definition 3 for the augmented MDP in (11).

B. Optimality Equations

According to (10) and Lemma 1, for any function $V : \mathcal{A} \times \mathcal{X} \times Y \to \mathbb{R}$, we define the following risk-aware Bellman operator $T : \mathcal{A} \times \mathcal{X} \times Y \to \mathcal{A} \times \mathcal{X} \times Y$ on V as follows:

$$\begin{aligned} T[V](\boldsymbol{A}, x, y) = \min_{s \in \mathcal{S}} \Bigg[& (1+\eta) A_r + \nu s C \\ & + \gamma \max_{\xi \in \Xi(x, y, \Pr(\cdot|\boldsymbol{A}, s))} \sum_{\boldsymbol{A}' \in \mathcal{A}} \Big(\xi(\boldsymbol{A}') \\ & \times V(\boldsymbol{A}', x/\xi(\boldsymbol{A}'), y\xi(\boldsymbol{A}')) \Pr[\boldsymbol{A}'|\boldsymbol{A}, s] \Big) \Bigg]. \end{aligned} \tag{12}$$

Here, from (12), we introduce $\xi^*(\cdot)$ as follows:

$$\begin{aligned} \xi^*(\boldsymbol{A}, x, y, s) \triangleq \arg \max_{\xi \in \Xi(x, y, \Pr(\cdot|\boldsymbol{A}, s))} & \sum_{\boldsymbol{A}' \in \mathcal{A}} \Big(\xi(\boldsymbol{A}') \\ \times V(\boldsymbol{A}', x/\xi(\boldsymbol{A}'), y\xi(\boldsymbol{A}')) & \Pr[\boldsymbol{A}'|\boldsymbol{A}, s] \Big), \forall \boldsymbol{A}, x, y. \end{aligned} \tag{13}$$

We denote T^k by the composition of the mapping T with itself k times, i.e., $T^k[V](\boldsymbol{A}, x, y) \triangleq T[T^{k-1}[V]](\boldsymbol{A}, x, y)$. According the definition of the risk envelop of the coherent risk measure $\rho_{\delta,\phi}$, we can show that the risk-aware Bellman operator $T[V]$ has the monotonicity and constant shift properties [18, Chapter 1.1.2].[1] Given the definition of $T[V]$ in (12), we next provide the expression of $T^k[V]$ for $k = 1, 2, \cdots$.

Lemma 2: For any $(\boldsymbol{A}, x, y) \in \mathcal{A} \times \mathcal{X} \times Y$ and any function $V : \mathcal{A} \times \mathcal{X} \times Y \to \mathbb{R}$, we have

$$\begin{aligned} & T^k[V](\boldsymbol{A}, x, y) \\ & = \min_{\tilde{\pi} \in \tilde{\Pi}_S} \rho_{x,y} \Bigg((1+\eta) \sum_{t=0}^{k-1} \gamma^t A_{r,t} + \gamma^k V(\boldsymbol{A}, x, y | \boldsymbol{A}_0 = \boldsymbol{A}, \tilde{\pi} \Bigg) \\ & \quad + \nu \mathbb{E} \Bigg[\sum_{t=0}^{k-1} \gamma^t s_t C | \boldsymbol{A}_0 = \boldsymbol{A}, \tilde{\pi} \Bigg], \ \forall k = 1, 2, \cdots, \end{aligned} \tag{14}$$

where the action s_t is induced by $\tilde{\pi}(\boldsymbol{A}_t, x_t, y_t)$.

Based on Lemma 2, we can obtain the optimal augmented stationary updating control policy $\tilde{\pi}^*$:

Theorem 1: For any given $(\boldsymbol{A}, x, y) \in \mathcal{A} \times \mathcal{X} \times Y$, the optimal function $V^*(\cdot)$ in (11) satisfies that:

$$V^*(\boldsymbol{A}, x, y) = T[V^*](\boldsymbol{A}, x, y). \tag{15}$$

Moreover, $V^*(\cdot)$ is the unique solution to (15) within the class of bounded functions.

Proof Sketch: We first show that, for any bounded functions $V : \mathcal{A} \times \mathcal{X} \times Y \to \mathbb{R}$,

$$V^*(\boldsymbol{A}, x, y) = \lim_{N \to \infty} T^N[V](\boldsymbol{A}, x, y), \tag{16}$$

holds for all $\boldsymbol{A}, x, y$. To prove (16), we break $V_{\tilde{\pi}}(\boldsymbol{A}, x, y)$ into the portions incurred over the first N stage and over the remaining stages. Then, by using the monotonocity and the subadditivity of the coherent risk measure $\rho_{x,y}(\cdot)$, the fact $\rho_{x,y}(c) = c$ for a constant c, the upper-limit $\hat{A}_r$ of the AoI at the receiver, and Lemma 2, we can obtain

$$\begin{aligned} & |V_{\tilde{\pi}}(\boldsymbol{A}, x, y) - T^N[V](\boldsymbol{A}, x, y)| \\ & \leq \frac{\gamma^N}{1-\gamma} \big((1+\eta) \hat{A}_r + \nu C + \max_{\boldsymbol{A}, x, y} |V(\boldsymbol{A}, x, y)| \big), \end{aligned} \tag{17}$$

based on which, we can prove (16).

Next, considering a zero function $V_0(\cdot)$ such that $V_0(\boldsymbol{A}, x, y) = 0$ for all $\boldsymbol{A}, x, y$, and by the monotonicity and constant shift properties of $T[V]$, we can show that $V^* = T[V^*]$.

Finally, the uniqueness of the solution to (15) can be proved by the monotonicity property of $T[V]$ and (16). ∎

Now, we have the optimal augmented stationary policy $\tilde{\pi}^*$ to the augmented MDP in (11), given by:

$$\begin{aligned} \tilde{\pi}^*(\boldsymbol{A}, x, y) = \arg\min_{s \in \mathcal{S}} \Bigg[& (1+\eta) A_r + \nu s C \\ & + \gamma \max_{\xi \in \Xi(x, y, \Pr(\cdot|\boldsymbol{A}, s))} \sum_{\boldsymbol{A}' \in \mathcal{A}} \Big(\xi(\boldsymbol{A}') \\ & \times V^*(\boldsymbol{A}', x/\xi(\boldsymbol{A}'), y\xi(\boldsymbol{A}')) \Pr[\boldsymbol{A}'|\boldsymbol{A}, s] \Big) \Bigg]. \end{aligned} \tag{18}$$

We now show that $\tilde{\pi}^*$ can be used to construct the optimal history-dependent policy π for the risk-aware MDP in (9).

Theorem 2: For any $\boldsymbol{A} \in \mathcal{A}$, $x = \frac{1}{1+\eta}$, and $y = \frac{\alpha(1+\eta)}{\alpha+\eta}$, the optimal function $V^*(\cdot)$ in (11) equals to the optimal solution of the risk-aware MDP in (9), i.e.,

$$V^*(\boldsymbol{A}, x, y) = \min_{\pi \in \Pi_H} \bar{A}^{\gamma}_{r,\pi}(\boldsymbol{A}) + \eta \rho^{\gamma}_{\pi}(\boldsymbol{A}) + \nu \bar{C}^{\gamma}_{\pi}(\boldsymbol{A}). \tag{19}$$

Moreover, the optimal history-dependent policy $\pi^* = (\mu_0^*, \mu_1^*, \cdots)$ of (9) is given by:

$$\mu_t^*(\boldsymbol{h}_t) = \tilde{\pi}^*(\boldsymbol{A}_t, x_t, y_t), \tag{20}$$

with the initial system AoI state $\boldsymbol{A}_0 = \boldsymbol{A}$ and risk levels $(x_0, y_0) = (x, y)$. Here, the dynamics of $\boldsymbol{A}_t$ are given by (1) and (2), and the dynamics of (x_t, y_t) are given by:

$$x_{t+1} = x_t / \xi^*(\boldsymbol{A}_t, x_t, y_t, \tilde{\pi}^*(\boldsymbol{A}_t, x_t, y_t)), \tag{21}$$

[1] All proofs are omitted due to space limitations.

$$y_{t+1} = y_t \xi^*(\boldsymbol{A}_t, x_t, y_t, \tilde{\pi}^*(\boldsymbol{A}_t, x_t, y_t)), \tag{22}$$

where $\xi^*(\cdot)$ is given by (13).

Proof Sketch: First, we show that the optimal solution of the risk-aware MDP in (9) is also a solution to (17), by exploiting the k-th tail history dependent policy and the temporal decomposition of $\rho_{x,y}(\cdot)$. Then, by the uniqueness of the solution to (17), we can immediately have (19). Then, we show that, the history-dependent policy constructed with the associated augmented stationary policy, is optimal, by using similar approaches as in Proposition 1.2.5 in [18]. ■

From Theorems 1 and 2, we observe that, although the original risk-aware MDP in (9) is defined over the intractable space of history-dependent updating policies, we only need to focus on finding the optimal augmented stationary policy defined in Definition 3, which depends on the original system AoI state $\boldsymbol{A}$ and two additional risk levels (x, y). Moreover, from the dynamics of the two risk levels (x_t, y_t) in (21) and (22), it can be seen that the values of (x_t, y_t) contain the historical information that is necessary to make the optimal decision, and thus can be seen as a certain kind of sufficient statistics. Furthermore, the optimal history-dependent updating control policy $\pi^* \in \Pi_H$ can be derived, by first obtaining the optimal augmented stationary $\tilde{\pi}^* \in \tilde{\Pi}_S$ in (18) and then using the construction procedure in Theorem 2. Here, to derive derive $\tilde{\pi}^*$, we can apply the value iteration algorithm [18] to obtain $V^*(\cdot)$. Let V_k be the value function at iteration k which is updated according to $V_k(\boldsymbol{A}, x, y) = T[V_{k-1}](\boldsymbol{A}, x, y)$. By (16), under any initialization of a bounded $V_0(\cdot)$, the generated sequence $\{V_k(\boldsymbol{A}, x, y)\}$ converges to $V^*(\boldsymbol{A}, x, y)$, i.e., $V^*(\boldsymbol{A}, x, y) = \lim_{k\to\infty} V_k(\boldsymbol{A}, x, y)$.

In a nutshell, we have proposed a novel approach that explicitly accounts for rare events with very large AoI in a IoT status updating system and developed a dynamic programming based solution to obtain the optimal history-dependent updating policy.

Remark 1: The proposed solution framework is significant, as it can be applied to design *optimal solutions* for risk-aware AoI minimization in other IoT scenarios, in which the optimal policy should be history-dependent, and, thus, is generally intractable. Moreover, the kernel of the proposed solution is dynamic programming, which further allows for the design of more efficient algorithms by levering advanced machine learning in real-time IoT monitoring systems.

V. Conclusion

In this paper, we have studied the optimal process update policy that minimizes the AoI at the receiver, the CVaR of the AoI at the receiver, and the energy cost. We have formulated this stochastic optimization problem as an infinite horizon discounted risk-aware MDP. To obtain the optimal history-dependent policy of the risk-aware MDP, we first reduce it to a standard MDP with an augmented system state space consisting of the original system AoI state space and the state space of two additional risk levels. For the augmented MDP, we have shown that the optimal stationary policy can be derived through dynamic programming based on a risk-aware Bellman operator. Then, we have shown that the optimal history-dependent policy of the risk-aware MDP depends on the history only through the augmented system states and can be constructed, by first obtaining the optimal stationary policy of the augmented MDP and then using a special construction procedure. The proposed solution is shown to be computationally tractable and can be applied in real-time IoT monitoring systems to minimize the AoI.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, pp. 1–9, 2019.

[2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, Orlando, FL, USA, March 2012.

[3] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov 2017.

[4] 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.

[5] B. Wang, S. Feng, and J. Yang, "When to preempt? age of information minimization under link capacity constraint," *arXiv preprint arXiv:1812.05670*, 2018.

[6] B. Zhou and W. Saad, "Minimum age of information in the Internet of Things with non-uniform status packet sizes," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2019.

[7] E. T. Ceran, D. Gndz, and A. Gyrgy, "Reinforcement learning to minimize age of information with an energy harvesting sensor with HARQ and sensing cost," in *Proc. of IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Paris, France, April 2019.

[8] S. Feng and J. Yang, "Age of information minimization for an energy harvesting source with updating erasures: With and without feedback," *arXiv preprint arXiv:1808.05141*, 2018.

[9] M. A. Abd-Elmagid, H. S. Dhillon, and N. Pappas, "A reinforcement learning framework for optimizing age-of-information in RF-powered communication systems," *arXiv preprint arXiv:1908.06367*, 2019.

[10] R. Devassy, G. Durisi, G. C. Ferrante, O. Simeone, and E. Uysal, "Reliable transmission of short packets through queues and noisy channels under latency and peak-age violation guarantees," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 721–734, April 2019.

[11] M. K. Abdel-Aziz, S. Samarakoon, C. Liu, M. Bennis, and W. Saad, "Optimized age of information tail for ultra-reliable low-latency communications in vehicular networks," *IEEE Trans. Commun.*, pp. 1–1, 2019.

[12] C. Liu and M. Bennis, "Taming the tail of maximal information age in wireless industrial networks," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2442–2446, Dec 2019.

[13] R. T. Rockafellar, S. Uryasev *et al.*, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.

[14] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203–228, 1999.

[15] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, 2009.

[16] G. C. Pflug and A. Pichler, "Time-consistent decisions and temporal decomposition of coherent risk functionals," *Math. Oper. Res.*, vol. 41, no. 2, pp. 682–699, 2016.

[17] K. Deb, "Multi-objective optimization," in *Search methodologies*. Springer, 2014, pp. 403–449.

[18] D. P. Bertsekas, *Dynamic programming and optimal control, 4th edition, volume II*. Belmont, MA: Athena Scientific, 2012.

[19] D. A. Iancu, M. Petrik, and D. Subramanian, "Tight approximations of dynamic risk measures," *Math. Oper. Res.*, vol. 40, no. 3, pp. 655–682, 2015.

[20] Y. Chow, "Risk-sensitive and data-driven sequential decision making," Ph.D. dissertation, Stanford University, 2017.