Benefits of Coding on Age of Information in Broadcast Networks

Xingran Chen and Shirin Saeedi Bidokhti
Department of Electrical and Systems Engineering
University of Pennsylvania
{xingranc, saeedi}@seas.upenn.edu

Abstract

Age of Information (AoI) is studied in two-user broadcast networks with feedback, and lower and upper bounds are derived on the expected weighted sum AoI of the users. In particular, a class of simple coding actions is considered and within this class, randomized and deterministic policies are devised. Explicit conditions are found for symmetric dependent channels under which coded randomized policies strictly outperform the corresponding uncoded policies. Similar behavior is numerically shown for deterministic policies.

Index Terms—**Age of Information, Network Coding, Randomized Policy, Max-Weight Policy, Feedback, Broadcast Packet Erasure Channels**

I. Introduction

Sending status updates in a timely manner has significant importance in the Internet of Things (IoT) applications. In practice, it is not always effective to update the information as fast as possible for it may cause further delay in the network queues. To measure the timeliness of information at a remote system, the concept of Age of Information (AoI) was introduced in [1]–[3]. AoI measures, at the receiving side, how much time has passed since the generation time of the latest received packet. In [3], a single source and server setup were considered under First-Come First-Serve (FCFS) queue management and it was shown that there is an optimal update rate that minimizes time-average AoI. Further extensions to networks of multiple sources and servers with and without packet management were studied in [4]–[8]. More recently, AoI has been studied as a performance metric in various contexts such as source and channel coding [9]–[11], caching [12]–[14], energy harvesting [15]–[17], sampling [18]–[20] and scheduling [21]–[29].

In coding theory, previous work has mainly studied point to point channels. For example, [30]–[32] consider erasure channels, propose coding schemes, and analyze the resulting average or peak AoI in various setups. More recently, [33] proves that when the source alphabet and channel input alphabet have the same size, a Last-Come First-Serve (LCFS) with no buffer policy is optimal. Considering erasure channels with FCFS M/G/1 queues, [34] finds an optimal block length for channel coding to minimize the average age and average peak age of information.

AoI has also been investigated in network management and scheduling. In particular, [21] proposes scheduling policies to optimize the overall age in a wireless network. Maintaining equally up-to-date and synchronized information from multiple sources is studied in [22]. In [26], scheduling algorithms are designed to minimize AoI in wireless broadcast channels. [25] devises scheduling policies to minimize average AoI under throughput constraints in wireless multi-access networks. The minimum age of time-varying wireless channels with interference constraints is obtained in [27]–[29] with and without channel state information.

In this work, we aim to shed light on the interplay between AoI and (channel/network) coding in the context of broadcast packet erasure channels (BPECs) with feedback. BPECs and their variants have been investigated in previous work such as [35]–[39] and rate-optimal coding algorithms are designed using (network) coding ideas. The key idea is due to [35] where it is shown that the entire capacity region of two-user BPECs with feedback can be attained by XOR-ing overheard packets. More precisely, suppose a packet p, intended for user 1, is broadcasted and only received at user 2. Scheduling algorithms re-transmit this packet because it is not received at its intended receiver. However, one may be able to exploit coding opportunities by tracking such packets (and this is possible through the available feedback). For example, in a similar manner, a packet q, intended for user 2, may get transmitted and received only at receiver 1. Now instead of re-transmitting p and q in two uses of the network, one can transmit the XOR packet $p \oplus q$ which is simultaneously useful for both users.

As opposed to the aforementioned literature, in this work, we seek *efficiency in terms of age as opposed to rate*. The underlying challenge is as follows. On the one hand, the highest rate of communication in BPECs can be attained when coding is employed across packets of different users [35]. A higher rate effectively corresponds to a smaller delay (both in the sense that the queues get emptied faster and in the sense that fewer uses of the network are needed in total to transmit a fixed number of information bits). On the other hand, to achieve high rates with coding, we have to incur delay by waiting for the arrival of other packets for the purpose of coding as well as prioritizing their transmission. So it is not clear apriori when coding is beneficial. We will devise scheduling policies that *schedule different coding actions*, as opposed to traditional schemes that schedule different users, and show the benefit of coding in terms of average age over

uncoded schemes such as those proposed in [26].

Motivated by the capacity achieving coding scheme of [35], in this work we restrict attention to a class of coding algorithms consisting of three actions: uncoded transmission for user 1, uncoded transmission for user 2, and coded (XOR-ed) transmission for both users. We consider a discrete time model as in [26] and study the expected weighted sum of AoI (EWSAoI) at the users. The first contribution of the paper is a general lower bound on the achievable EWSAoI. As opposed to previous lower bounds (e.g. [26]) that hold only in the class of traditional scheduling algorithms, the new lower bound is valid for any coding scheme. The second contribution of the paper is the devise and analysis of EWSAoI or an upper bound on it for (i) stationary randomized policies and (ii) deterministic Max-Weight (MW) policies. In the class of randomized policies, for symmetric channels, we find conditions under which coded policies perform strictly better than their corresponding uncoded policies. For MW policies, we numerically compare the performance of coded and uncoded MW policies and show gains of coding.

The remainder of this paper is organized as follows. The problem setup and notation are introduced in Section II. Section III presents a general lower bound on EWSAoI. In Section IV, we devise a randomized policy based on three coding actions and find a closed-form expression for the resulting EWSAoI. We further study the special case of symmetric BPECs and find conditions under which coded randomized policies strictly outperform uncoded randomized policies in terms of age. In Section V, we propose a Max-Weight policy and derive an upper bound on the resulting EWSAoI. Simulation results and the comparison of uncoded vs. coded schemes are presented in Section VI.

II. SYSTEM MODEL

We consider a model where time is slotted. At the beginning of every time slot, new packets are generated for the users and they replace any undelivered packets from previous time slots. Our model is similar to [25], [26], where users are scheduled with the goal of minimizing age.

Transmission occurs on a noisy network which we model by a broadcast packet erasure channel with two users. In each time slot k, the input $X(k)$ to the channel is a packet. The packet is successfully delivered to user i with probability $1-\epsilon_i$, $0 \le \epsilon_i < 1$, and lost with probability ϵ_i. Let $Z_i(k)$ be a random variable modeling erasure at user $i \in \{1, 2\}$ in time slot $k \in \{1, 2, \ldots\}$. We assume that the channel is memoryless and hence $\{Z_i(k)\}_{k=1}^{\infty}$ is an iid Bernoulli process with probability $1 - \epsilon_i$. The output of the channel at user i in slot k is:

$$Y_i(k) = \begin{cases} X(k) & \text{if } Z_i(k) = 1 \\ \Delta & \text{otherwise} \end{cases}$$

where Δ is a symbol denoting erasure. Note that the pairs $\{(Z_1(k), Z_2(k))\}_k$ are independent across time (over $k =$
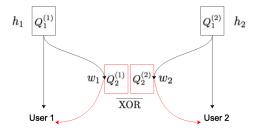

Fig. 1: the virtual network of queues at the encoder

$1, 2, \ldots$) but potentially correlated across (Z_1, Z_2). In addition, the feedback is available at the encoder after each transmission. Define $\epsilon_1, \epsilon_2, \epsilon_{12}$ as

$$\epsilon_1 := \Pr(Z_1 = 0)$$
$$\epsilon_2 := \Pr(Z_2 = 0)$$
$$\epsilon_{12} := \Pr(Z_1 = 0, Z_2 = 0),$$

and hence we have

$$\Pr(Z_1 = 1, Z_2 = 0) = \epsilon_2 - \epsilon_{12}$$
$$\Pr(Z_1 = 0, Z_2 = 1) = \epsilon_1 - \epsilon_{12}.$$

The statistics of (Z_1, Z_2) that describes the channel is assumed fixed and given and can be characterized by $(\epsilon_1, \epsilon_2, \epsilon_{12})$.

We consider a simple class of coding algorithms that consists of three actions, including a network coding action. The encoder is modeled by a network of virtual queues. Let $Q_1^{(i)}$ denote the queue of incoming packets for user i and $Q_2^{(i)}$ denote the queue of packets that are intended for user i but are received only by the other user. The encoder can track such packets using the available feedback. For $i = \{1, 2\}$, we use the notation $\backslash i$ as short for $\{1, 2\}\backslash i$. The packets in $Q_2^{(i)}$ are not received at their intended receivers, but are received at the other receiver and act as side information for it – this can be exploited in the code design at the encoder. In particular, the encoder can XOR packets in $Q_2^{(1)}$ with $Q_2^{(2)}$ and form more efficient coded packets for transmission.

In each time slot k, the encoder decides between the following three actions, denoted by $A(k) \in \{1, 2, 3\}$ and defined below:

- $A(k) = 1$: a packet is transmitted from $Q_1^{(1)}$;
- $A(k) = 2$: a packet is transmitted from $Q_1^{(2)}$;
- $A(k) = 3$: a coded packet is transmitted from $Q_2^{(1)}, Q_2^{(2)}$.

Definition 1 (Age of Information [1]). *Consider a source-destination pair. Let $\{t_k\}_k$ be the times at which packets are generated and $\{t'_k\}_k$ be the times at which packets are received at the destination. At any time ξ, denote $N(\xi) = \max\{k|t'_k \le \xi\}$, and $u(\xi) = t_{N(\xi)}$. The Age of Information (AoI) at the destination is $\Delta(t) = t - u(t)$.*

Using Definition 1, let h_i be the positive real number that represents the age at user i. The age h_i increases linearly in time when there is no delivery of packets to user i and
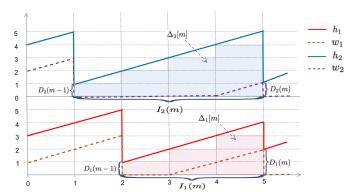
Fig. 2: a sample path of the channel state (h_1, h_2, w_1, w_2) which initial state $(h_1, h_2, w_1, w_2) = (3, 4, 1, 2)$

drops with every delivery to a value that represents how old the received packet is.

Lemma 1. *To attain the optimal age in the above class of $3-$action coding algorithms, we can assume, without loss of generality, that all queues are of buffer size 1.*

Proof. The proof is presented in Appendix A. □

To capture the evolution of h_i in the class of $3-$action algorithms described, we proceed as follows. First, define $w_i(k)$ as the (current) age of information at $Q_2^{(i)}$ in slot k. If the packet in $Q_2^{(i)}$ is successfully delivered at user i by time k, then it is removed from $Q_2^{(i)}$ and $w_i(k)$ is defined to be zero; if a packet in $Q_2^{(i)}$ is replaced by a new packet p in slot k, then $w_i(t)$ is the age of the new packet when $t > k$. More precisely, suppose packet p is generated at time t_p. At time t, while p is in the queue $Q_2^{(i)}$, the age $w(t)$ is $t - t_p$. Finally, once a packet is delivered successfully at user i from $Q_1^{(i)}$, then the existing packet in $Q_2^{(i)}$ (which is necessarily older) becomes obsolete and hence we remove it from $Q_2^{(i)}$ and define $w_i(k+1)$ to be 0. Thus, the recursion of $w_i(k)$ is

$$w_i(k+1) = \begin{cases} 0 & \text{if } A(k) \in \{i, 3\},\ Z_i(k) = 1 \\ 1 & \text{if } A(k) = i,\ (Z_i(k), Z_{\setminus i}(k)) = (0, 1) \ . \\ (w_i(k)+1) \cdot 1_{\{w_i(k)>0\}} & \text{otherwise} \end{cases} \quad (1)$$

Based on $w_i(k)$, the age function $h_i(k)$ evolves as follows:

$$h_i(k+1) = \begin{cases} 1 & \text{if } A(k) = i,\ Z_i(k) = 1 \\ w_i(k) + 1 & \text{if } A(k) = 3,\ Z_i(k) = 1 \ . \\ h_i(k) + 1 & \text{otherwise} \end{cases} \quad (2)$$

Remark 1. *Using the mathematical recursions in (1) and (2), we conclude that $w_i(k) \leq h_i(k) - 1$.*

A. A Sample Path

A sample path for the evolution of w_1, h_1 and w_2, h_2 is shown in Fig. 2. The initial state is $(h_1, h_2, w_1, w_2) = (3, 4, 1, 2)$, and the actions and the channels are as follows:

k	1	2	3	4	5
$A(k)$	2	1	1	2	3
$(Z_1(k), Z_2(k))$	(1,1)	(1,0)	(0,1)	(1,0)	(1,1)

Now consider a general sample path associated with a transmission policy and a finite time-horizon T. For this sample path, let $N_i(T)$ be the total number of packets delivered to user i up to and including time slot T, and $I_i(m)$ be the number of time slots between the $(m-1)$th and mth deliveries to user i, i.e., the inter delivery times of user i. Denote the age of user i after delivery of the mth packet by $D_i(m)$ and let L_i be the number of remaining time slots after the last packet delivery to the same user. With this notation, the time-horizon can be written as $T = \sum_{m=1}^{N_i(T)} I_i(m) + L_i$ with $i \in \{1, 2\}$.

Next, consider the sum of the instantaneous ages in the interval corresponding to $I_i(m)$, $m \geq 2$, denoted by $\Delta_i[m]$. As shown in Figure 2, $\Delta_i[m]$ is equal to the area underneath the age curve in the corresponding interval minus $I_i(m)$ small triangle areas; i.e.,

$$\Delta_i[m] = \sum_{k \text{ in between delivery of } m-1^{th} \text{ and } m^{th} \text{ packets}} h_i[k] \quad (3)$$

$$= \frac{(D_i(m-1) + I_i(m))^2}{2} - \frac{(D_i(m-1))^2}{2} - \frac{I_i(m)}{2} \quad (4)$$

$$= \frac{I_i^2(m)}{2} + D_i(m-1)I_i(m) - \frac{I_i(m)}{2}. \quad (5)$$

B. The Expected Weighted Sum AoI

Both $w_i(k)$ and $h_i(k)$ and their evolution depend on the policy that we choose and, hence, we sometimes write them as $w_i^\pi(k)$ and $h_i^\pi(k)$. Moreover, we denote the vector $(h_1(1), h_2(1), w_1(1), w_2(1))$ by $\vec{s}(1)$.

We aim to find policies π that minimize the following EWSAoI at the users:

$$\mathbb{E}\left[\frac{1}{2T} \sum_{k=1}^{T} \sum_{i=1}^{2} \alpha_i h_i^\pi(k) \middle| \vec{s}(1) \right] \quad (6)$$

where α_1 and α_2 are weights associated to users 1 and 2, respectively. We assume $\alpha_i \geq 0$ and $\alpha_1 + \alpha_2 = 1$. For notational simplicity, we omit $\vec{s}(1)$ hereafter, and hence the minimum age is given by the following optimization problem.

$$\min_{\pi \in \Pi} \mathbb{E}[J_T^\pi], \quad \text{where} \quad J_T^\pi = \frac{1}{2T} \sum_{k=1}^{T} \sum_{i=1}^{2} \alpha_i h_i^\pi(k). \quad (7)$$

Since we aim to minimize the EWSAoI in the long run, we define

$$J^\pi = \lim_{T \to \infty} J_T^\pi.$$

III. A LOWER BOUND

We prove a lower bound on EWSAoI as stated below.

Theorem 1. *For any communication policy π, we have:*

$$\mathbb{E}[J^\pi] \geq \frac{1}{4} \left(\frac{(\sum_{i=1}^{2} \sqrt{\alpha_i(2 - \epsilon_{12} - \epsilon_{\setminus i})})^2}{(1 - \epsilon_{12})(2 - \epsilon_1 - \epsilon_2)} + 1 \right). \quad (8)$$

Remark 2. *The lower bound of Theorem 1 holds in general and is not restricted to the class of* $3-$*action coded algorithms that we introduced in Section II.*

Proof. Consider a sample path associated with a transmission policy and a finite time-horizon T (see Section II-A). The EWSAoI as defined in (6) can be re-written in terms of $\Delta_i(m)$'s:

$$J_T^\pi = \frac{1}{2T}\sum_{k=1}^{T}\sum_{i=1}^{2}\alpha_i h_i(k)$$

$$= \frac{1}{2}\sum_{i=1}^{2}\frac{\alpha_i}{T}\left(\sum_{m=1}^{N_i(T)}\Delta_i(m) + \frac{1}{2}L_i^2 + D_i(N_i(T))L_i - \frac{1}{2}L_i\right). \quad (9)$$

Since $D_i(m) \geq 1$ for all $1 \leq m \leq N_i(T)$, we can lower bound (5) and hence (9) by substituting $D_i(m) = 1$. Using similar steps as [26, Eqns. (9) - (14)], we find

$$J_T^\pi \geq \frac{1}{4}\sum_{i=1}^{2}\alpha_i\frac{T}{N_i(T)} + \frac{1}{4}. \quad (10)$$

We remark that so far, the lower bound on J_T^π is the same as [26, Eqn. (7)]. We now depart from [26] by allowing for the general class of coding and scheduling schemes. Recall that $N_i(T)$ is the total number of packets received by user i, $i = 1, 2$. In the limit of $T \to \infty$, $\frac{N_i(T)}{T}$ is the throughput of user i. We further know the capacity of two-user broadcast packet erasure channels from [35]. In particular, any non-negative rate pair (R_1, R_2) is achievable if and only if it satisfies the following conditions.

$$\frac{R_1}{1-\epsilon_1} + \frac{R_2}{1-\epsilon_{12}} \leq 1 \quad (11)$$

$$\frac{R_1}{1-\epsilon_{12}} + \frac{R_2}{1-\epsilon_2} \leq 1. \quad (12)$$

Hence, from (11) and (12), we have

$$\lim_{T\to\infty}\frac{\frac{N_1(T)}{T}}{1-\epsilon_1} + \frac{\frac{N_2(T)}{T}}{1-\epsilon_{12}} \leq 1 \quad (13)$$

$$\lim_{T\to\infty}\frac{\frac{N_1(T)}{T}}{1-\epsilon_{12}} + \frac{\frac{N_2(T)}{T}}{1-\epsilon_2} \leq 1. \quad (14)$$

For notational simplicity, denote $x = \lim_{T\to\infty}\frac{N_1(T)}{T}$ and $y = \lim_{T\to\infty}\frac{N_2(T)}{T}$. Substituting x and y into (13) and (14), we obtain

$$\frac{x}{1-\epsilon_1} + \frac{y}{1-\epsilon_{12}} \leq 1$$

$$\frac{x}{1-\epsilon_{12}} + \frac{y}{1-\epsilon_2} \leq 1.$$

Re-arranging terms, we obtain

$$(1-\epsilon_{12})x + (1-\epsilon_1)y \leq (1-\epsilon_{12})(1-\epsilon_1)$$

$$(1-\epsilon_2)x + (1-\epsilon_{12})y \leq (1-\epsilon_{12})(1-\epsilon_2)$$

and summing the above two inequalities, we get

$$k_1 x + k_2 y \leq k_3 \quad (15)$$

where

$$k_1 = 2 - \epsilon_2 - \epsilon_{12}$$
$$k_2 = 2 - \epsilon_1 - \epsilon_{12}$$
$$k_3 = (1-\epsilon_{12})(2 - \epsilon_1 - \epsilon_2).$$

We can now use (15) along with the Cauchy-Schwarz inequality and write

$$\left(\frac{\alpha_1}{x} + \frac{\alpha_2}{y}\right)k_3 \geq \left(\frac{\alpha_1}{x} + \frac{\alpha_2}{y}\right)(k_1 x + k_2 y)$$

$$\geq \left(\sqrt{\alpha_1 k_1} + \sqrt{\alpha_2 k_2}\right)^2.$$

Re-arranging terms and replacing for k_1, k_2, k_3, we find

$$\left(\frac{\alpha_1}{x} + \frac{\alpha_2}{y}\right) \geq \frac{(\sum_{i=1}^{2}\sqrt{\alpha_i(2 - \epsilon_{\backslash i} - \epsilon_{12})})^2}{(1-\epsilon_{12})(2 - \epsilon_1 - \epsilon_2)}$$

and hence, from (10), we obtain

$$J^\pi \geq \lim_{T\to\infty}\frac{1}{4}\sum_{i=1}^{2}\alpha_i\frac{T}{N_i(T)} + \frac{1}{4}$$

$$\geq \frac{1}{4}\left(\frac{(\sum_{i=1}^{2}\sqrt{\alpha_i(2 - \epsilon_{12} - \epsilon_{\backslash i})})^2}{(1-\epsilon_{12})(2 - \epsilon_1 - \epsilon_2)} + 1\right).$$

$\square$

IV. CODED RANDOMIZED POLICIES

Consider a stationary randomized policy where each action is chosen with a fixed probability in each time slot, independent of the system's status. Denote by μ_i the probability of action $A(k) = i$, $i \in \{1, 2, 3\}$, $k \in \{1, \ldots\}$, where $\mu_1 + \mu_2 + \mu_3 = 1$. Denote the EWSAoI for randomized policies (in the long run) as $\mathbb{E}[J^R]$.

A. Age Analysis

We will first find the exact EWSAoI of the coded randomized policy. We start with $\Delta_i(m)$ derived in (5). Consider $i = 1$. The expectation of $\Delta_1(m)$ is

$$\mathbb{E}[\Delta_1(m)|\vec{s}(1)] \quad (16)$$

$$= \mathbb{E}\left[\frac{I_1^2(m)}{2} + D_1(m-1)I_1(m) - \frac{I_1(m)}{2}\Big|\vec{s}(1)\right] \quad (17)$$

$$= \frac{\mathbb{E}[I_1^2]}{2} + \mathbb{E}[D_1]\mathbb{E}[I_1] - \frac{\mathbb{E}[I_1]}{2}. \quad (18)$$

Equality (18) holds because of the following observations: (i) the processes $\{I_1(m)\}_m$ and $\{D_1(m)\}_m$ are each iid and not dependent on $\vec{s}(1)$ (so we use I_1 and D_1 to denote the underlying random variables, respectively), and (ii) $I_1(m)$ and

$D_1(m-1)$ are independent (while $I_1(m)$ and $D_1(m)$ may be dependent). Thus, the EWSAoI of user 1 is

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{T}\mathbb{E}[h_1(k)|\vec{s}(1)] \qquad (19)$$

$$=\lim_{T\to\infty}\frac{1}{T}\sum_{m=1}^{N_1(T)}\mathbb{E}[\Delta_1(m)|\vec{s}(1)] \qquad (20)$$

$$=\lim_{T\to\infty}\frac{N_1(T)}{T}\left(\frac{\mathbb{E}\left[I_1^2\right]}{2}+\mathbb{E}[D_1]\mathbb{E}[I_1]-\frac{\mathbb{E}[I_1]}{2}\right) \qquad (21)$$

$$=\frac{\mathbb{E}\left[I_1^2\right]}{2\mathbb{E}[I_1]}+\mathbb{E}[D_1]-\frac{1}{2} \qquad (22)$$

where (22) holds because the arrival process is a renewal process [40, Section 3.3, Theorem 3.3.1] and hence $\lim_{T\to\infty}\frac{T}{N_1(T)}=\mathbb{E}[I_1]$.

Then, we consider the statistics of D_1, and find the probability of $\Pr(D_1=d)$ for $d=1,2,\ldots$. For each slot k,

$$P(h_1(k)=1)=\mu_1(1-\epsilon_1)$$
$$P(h_1(k)=d)=\mu_1\mu_3(1-\epsilon_1)(\epsilon_1-\epsilon_{12}) \qquad (23)$$
$$\times(\mu_1\epsilon_{12}+\mu_2+\mu_3\epsilon_1)^{d-2}\quad d\geq 2.$$

To find the probability distribution of D_1, condition the above probabilities on the event that a packet is delivered to user 1 at time slot k. The probability of this event can be found by summing (23) over all $d\geq 1$:

$$P^1_{\text{delivery}}=\frac{\mu_1(1-\epsilon_1)(\mu_1+\mu_3)(1-\epsilon_{12})}{1-\mu_1\epsilon_{12}-\mu_2-\mu_3\epsilon_1}.$$

We thus find

$$P(D_1=d)=\begin{cases}\mu_1(1-\epsilon_i)/P^1_{\text{delivery}} & d=1\\ \frac{\mu_1\mu_3(\epsilon_1-\epsilon_{12})(1-\epsilon_1)(\mu_1\epsilon_{12}+\mu_2+\mu_3\epsilon_1)^{d-2}}{P^1_{\text{delivery}}} & d\geq 2\end{cases}$$

and the expectation of D_1 is equal to

$$E(D_1)=1+\frac{\mu_3(\epsilon_1-\epsilon_{12})}{(\mu_1+\mu_3)(1-\epsilon_{12})(1-\mu_1\epsilon_{12}-\mu_2-\mu_3\epsilon_1)}. \qquad (24)$$

The distribution of I_1 can be found by treating I_1 and D_1 jointly. First of all, we have

$$P(I_1=1)=P(I_1=1,D_1=1)=\mu_1(1-\epsilon_1).$$

Then we look at the event of $I_1=\ell$ and $D_1=d$ for $\ell\geq 2, d\leq\ell$ (otherwise, if $d>\ell$, then $\Pr(I_1=\ell,D_1=d)=0$ because the packet in the queue $Q_2^{(1)}$ becomes obsolete once a new packet is delivered to user 1). So we assume $d\leq\ell$ and consider the following cases: (i) If $d=1$, then a packet was delivered by action i at slot ℓ. (ii) For $d\geq 2$, the delivered packet was moved to $Q_2^{(1)}$ at slot $\ell-d+1$, stayed there, and got received at user 1 at slot ℓ. Now consider slots 1 to $\ell-d+1$. Denote by t the first slot in which a packet is received in $Q_2^{(1)}$, $1\leq t\leq\ell-d+1$. Then we have the two sub-cases: (ii-1) If t exists, then $Q_2^{(1)}$ is empty before t, and the delivered packet (another packet different from the delivered packet) moves to $Q_2^{(1)}$ at t when $t=\ell-d+1$ (when $t<\ell-d+1$) and

from that point $Q_2^{(1)}$ is non-empty. (ii-2) If there is no such slot t, which may happen for $d=1$, then $Q_2^{(1)}$ is empty for the entire duration of ℓ slots. Considering the above cases, for $\ell=1,2,\ldots$, we can show the following lemma.

Lemma 2. *The probability distribution of the inter delivery random variable I_1 is given by*

$$P(I_1=\ell)=\delta_1 x_1^{\ell-1}+\beta_1 y_1^{\ell-1} \qquad (25)$$

where

$$x_1=\mu_1\epsilon_{12}+\mu_2+\mu_3$$

$$y_1=\mu_1\epsilon_1+\mu_2+\mu_3\epsilon_1$$

$$\delta_1=\mu_1(1-\epsilon_{12})+\frac{\mu_1^2(\epsilon_1-\epsilon_{12})(1-\epsilon_{12})}{-\mu_1(\epsilon_1-\epsilon_{12})+\mu_3(1-\epsilon_1)}$$

$$\beta_1=-\frac{\mu_1(\epsilon_1-\epsilon_{12})(1-\epsilon_1)}{-\mu_1(\epsilon_1-\epsilon_{12})+\mu_3(1-\epsilon_1)}\left(\mu_1+\mu_3\right).$$

The proof of Lemma 2 is in Appendix B. Using Lemma 2, we find $\mathbb{E}[I_1]$ and $\mathbb{E}[I_1^2]$:

$$\mathbb{E}[I_1]=\frac{\delta_1}{(1-x_1)^2}+\frac{\beta_1}{(1-y_1)^2} \qquad (26)$$

$$\mathbb{E}[I_1^2]=\frac{\delta_1(1+x_1)}{(1-x_1)^3}+\frac{\beta_1(1+y_1)}{(1-y_1)^3}. \qquad (27)$$

Finally, substituting (24), (26), (27) into (22), and replacing for the values of x_i, y_i, z_i and δ_i, β_i, we find the EWSAoI as given by the following theorem.

Theorem 2. *The $EWSAoI$ of Randomized policy is characterized by*

$$\mathbb{E}[J^R]=\frac{1}{2}\sum_{i=1}^{2}\alpha_i\left(\frac{\frac{\mu_3(1-\epsilon_i)}{(\mu_i(1-\epsilon_{12}))^2}+\frac{-\mu_i(\epsilon_i-\epsilon_{12})}{((\mu_i+\mu_3)(1-\epsilon_i))^2}}{\frac{\mu_3(1-\epsilon_i)}{\mu_i(1-\epsilon_{12})}+\frac{-\mu_i(\epsilon_i-\epsilon_{12})}{(\mu_i+\mu_3)(1-\epsilon_i)}}\right.$$
$$\left.+\frac{\mu_3(\epsilon_i-\epsilon_{12})}{(\mu_i+\mu_3)(1-\epsilon_{12})(\mu_i(1-\epsilon_{12})+\mu_3(1-\epsilon_i))}\right). \qquad (28)$$

Remark 3. *To find an optimal coded randomized policy with respect to age, we have to choose the probability vector $(\mu_1^*, \mu_2^*, \mu_3^*)$ such that $\mathbb{E}[J^R]$ is minimized.*

Remark 4. *Setting $\mu_3=0$, we recover the EWSAoI of [26] which corresponds to uncoded randomized policies.*

B. Symmetric BPECs

Consider the class of symmetric BPECs. Let the erasure probabilities of channels to users 1 and 2 be equal to ϵ and ϵ_{12} be the probability of simultaneous erasure at both users. So $\epsilon>\epsilon_{12}$. Note that ϵ_{12} is either a function of ϵ or a constant, thus we rewrite ϵ_{12} as $\epsilon_{12}(\epsilon)$. Denote the probabilities of choosing action 1 and 2 by μ, hence the probability of choosing action 3 is $1-2\mu$ with $0\leq\mu\leq 1/2$. For simplicity, let $\alpha_1=\alpha_2=\alpha$. We find regimes of operation where optimal coded randomized policies strictly improve the EWSAoI over uncoded randomized policies such as [26].

From Theorem 2, we find the EWSAoI as follows

$$\mathbb{E}[J^R] = \frac{\alpha}{(1-\epsilon_{12}(\epsilon))(1-\mu)}$$
$$\times \left(\frac{(1-\epsilon)^3(1-2\mu)(1-\mu)^2 - (1-\epsilon_{12}(\epsilon))^2(\epsilon-\epsilon_{12}(\epsilon))\mu^3}{\left((1-2\mu)(1-\epsilon)^2(1-\mu) - \mu^2(1-\epsilon_{12}(\epsilon))(\epsilon-\epsilon_{12}(\epsilon))\right)\mu} \right.$$
$$\left. + \frac{(1-2\mu)(\epsilon-\epsilon_{12}(\epsilon))}{(2\epsilon-1-\epsilon_{12}(\epsilon))\mu + 1 - \epsilon} \right). \tag{29}$$

First, we get optimal randomized policies (i.e., optimal probabilities) of the symmetric case by the following lemma.

Lemma 3. *For any given ϵ,*

$$\mu^* = \begin{cases} \dfrac{\sqrt{1-\epsilon}}{\sqrt{1-\epsilon}+\sqrt{\epsilon-\epsilon_{12}(\epsilon)}} & \epsilon_{12}(\epsilon) - 2\epsilon + 1 < 0 \\ 1/2 & \epsilon_{12}(\epsilon) - 2\epsilon + 1 \geq 0. \end{cases} \tag{30}$$

The proof of Lemma 3 is in Appendix C. From Lemma 3, if $\epsilon_{12}(\epsilon) - 2\epsilon + 1 \geq 0$, then the optimal probability of action 1 (action 2) is $\mu^* = 1/2$. In this case, no packets reach the XOR, i.e., action 3 is never chosen. So coded randomized policies degenerate to uncoded randomized algorithms. Thus coded and uncoded randomized policies reach the same age. If $\epsilon_{12}(\epsilon) - 2\epsilon + 1 < 0$, since μ^* is the *unique* minimum point of (29) as shown in Appendix C, then

$$\mathbb{E}[J^R]\big|_{\{\mu^*\}} < \mathbb{E}[J^R]\big|_{\{\mu=1/2\}}$$

which implies that the age of optimal coded randomized policies is strictly lower than that of optimal uncoded randomized policies. So we arrive at the following result.

Theorem 3. *Optimal coded and uncoded randomized policies achieve the same EWSAoI if and only if $\epsilon_{12}(\epsilon) - 2\epsilon + 1 \geq 0$. Otherwise, optimal coded randomized policies strictly outperform uncoded randomized policies.*

Remark 5. *When the channels are independent, i.e., $\epsilon_{12}(\epsilon) = \epsilon^2$, coded and uncoded randomized policies have the same performance with respect to age.*

V. MAX-WEIGHT POLICIES

A. The Algorithm

In this section, we devise deterministic policies using techniques from Lyapunov Optimization. Denote the EWSAoI for Max-Weight policies (in the long run) as $\mathbb{E}[J^{MW}]$. Denote $\vec{h}(k) = (h_1(k), h_2(k))$ and $\vec{s}(k) = (h_1(k), h_2(k), w_1(k), w_2(k))$. Define the Lyapunov function

$$L(\vec{h}(k)) = \frac{1}{2}\sum_{i=1}^{2}\alpha_i h_i^2(k), \tag{31}$$

and the one-slot Lyapunov Drift

$$\Theta(\vec{h}(k)) = \mathbb{E}[L(\vec{h}(k+1)) - L(\vec{h}(k))|\vec{s}(k)]. \tag{32}$$

We devise the Max-Weight (MW) policy such that it minimizes the one-slot Lyapunov drift:

Definition 2. *In each slot k, the MW policy chooses the action that has the maximum weight as shown in following Table:*

$A(k)$	Weights
1	$\frac{\alpha_1(1-\epsilon_1)}{2}h_1(k)(h_1(k)+2)$
2	$\frac{\alpha_2(1-\epsilon_2)}{2}h_2(k)(h_2(k)+2)$
3	$\frac{1}{2}\sum_i \alpha_i(1-\epsilon_i)1_{\{w_i(k)>0\}}(h_i^2(k)+2h_i(k)-w_i^2(k)-2w_i(k))$

Theorem 4. *The MW policy defined in Definition 2 minimizes the one-slot Lyapunov Drift in each slot.*

Proof. The proof of Theorem 4 can be found in Appendix D. $\square$

B. The Upper Bound of EWSAoI

While the exact analysis of the resulting EWSAoI is difficult for the above MW policy, we derive an upper bound on it in Theorem 5. The proof is deferred to Appendix E.

Theorem 5. *The EWSAoI achieved by the proposed Max-Weight policy is upper bounded by*

$$\mathbb{E}[J^{MW}] \leq \sqrt{\frac{1}{2}\sum_{i=1}^{2}\frac{\alpha_i}{\mu_i(1-\epsilon_i)}\sum_{i=1}^{2}\alpha_i\Psi_i} + \frac{1}{2}\sum_{i=1}^{2}\alpha_i\Phi_i$$

where

$$\Phi_i = \frac{1 - \mu_3 P_{ne}^i(1-\epsilon_i) - \mu_i(1-\epsilon_i)}{\mu_i(1-\epsilon_i)}$$

$$\Psi_i = 1 - \mu_3 P_{ne}^i(1-\epsilon_i) + \frac{\left(1 - \mu_i(1-\epsilon_i) - \mu_3 P_{ne}^i(1-\epsilon_i)\right)^2}{\mu_i(1-\epsilon_i)}$$

$$P_{ne}^i = \frac{\mu_i(\epsilon_i - \epsilon_{12})}{\mu_3(1-\epsilon_i) + \mu_i(\epsilon_i - \epsilon_{12}) - \mu_3\mu_i(1-\epsilon_i)(\epsilon_i - \epsilon_{12})}.$$

Remark 6. *Setting $\mu_3 = 0$, we recover the upper bound of [26] which corresponds to uncoded randomized policies.*

VI. NUMERICAL RESULTS AND DISCUSSION

In this section, we compare the performance of the proposed coded algorithms with the uncoded algorithms in [26]. In Figure 3, we plot the EWSAoI of optimal coded and uncoded randomized policies, coded and uncoded Max-Weight policies, as well as the lower bound of Theorem 1. We have chosen $\alpha_1 = 0.3$, $\alpha_2 = 0.7$, and $\epsilon_1 = \epsilon_2 = \epsilon$, where ϵ varies from 0.5 to 0.9. We consider a dependent channel of $\epsilon_{12} = \epsilon^2/5$. We see that for the proposed policies, coding is beneficial when the channel erasure is larger than a threshold. Although the gain is small for MW policies, we believe the gain will be more significant over networks with many users. In addition, we also consider Dynamic Programming (DP) policies in the class of 3-action coding schemes, the EWSAoI performance of DP policies is very close to that of MW policies for all ϵ.
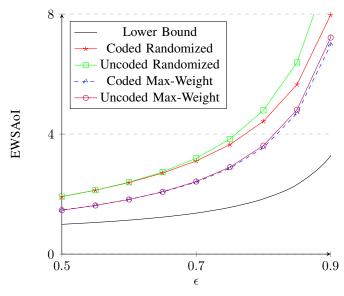
Fig. 3: EWSAoI as a function of erasure probability for a class of dependent channels with $\epsilon_{12} = \epsilon^2/5$.

APPENDIX A
PROOF OF LEMMA 1

Proof. We consider two types of policies: policies with buffer size 1, denote by π_1, and policies with buffer size larger than 1, denote by π_2. π_1 works on buffers of size 1 while π_2 works on buffers of infinite size. To differentiate the two policies and their corresponding queues, we label the packets inside the queues by *new* and *old*. A new packet in a queue means that this packet has not been previously received at its intended user. A packet in a queue is labeled old if there is a newer packet in the same queue or if the packet (or a fresher packet) is already received at its intended user. We refer to the freshest old packet as the old packet. At a given time slot, denote the new packet and the old packet for user 1 (resp. user 2) as p_{new} and p_{old} (resp. q_{new} and q_{old}) in $Q_1^{(i)}$ (resp. $Q_2^{(i)}$), $i = 1, 2$, respectively.

Consider a fixed erasure pattern for the channels to the users and an arbitrary time slot k. Denote the generation time of the new packet and the old packet for user 1 (resp. user 2) in $Q_1^{(i)}$ (resp. $Q_2^{(i)}$) as k_{pn} and k_{po} (resp. k_{qn} and k_{qo}), $i = 1, 2$. We measure the efficiency of the policies at each slot k by their resulting age at both users. To this end, we show that no matter what policy π_2 chooses, there is a policy of type π_1 that is at least as efficient as π_2 in terms of age at both users.

1) π_2 chooses action i, $i = 1, 2$, at slot k. In this case, π_2 transmits a packet from $Q_1^{(i)}$. By sending the new packet from $Q_1^{(i)}$, policy π_1 attains a smaller age for user i at slot $k + 1$ and the age remains the same for the other user. See also [33, Theorem 2] for a similar argument.

2) π_2 chooses action 3 at slot k. First of all, note that it is always strictly better to choose action i over action 3 if queue $Q_2^{(\backslash i)}$ is empty, and then a similar analysis with the previous case can be done. Therefore, we consider the case that both queues $Q_2^{(1)}$ and $Q_2^{(2)}$ are non-empty and a coded packet of the form $p_{old} \oplus q_{old}$ is sent by π_2. To compare with policies of type π_1, we need to consider four possible cases depending on if there are any new packets in $Q_2^{(1)}$ and $Q_2^{(2)}$. Note that type π_1 policies are effectively oblivious to old packets.

   - There are packets of type p_{old} as well as p_{new} in $Q_1^{(2)}$ and similarly q_{old}, q_{new} in $Q_2^{(2)}$. In this case, π_1 can choose action 3 and send the coded packet $p_{new} \oplus q_{new}$. This way, if there is a delivery at user 1 (resp. user 2), the age of user 1 (resp. user 2) will drop to $k + 1 - k_{po}$ (resp. $k + 1 - k_{qo}$) under π_2 and to $k + 1 - k_{pn}$ (resp. $k + 1 - k_{qn}$) under π_1 at time $k + 1$. Since $k_{pn} > k_{po}$ and $k_{qn} > k_{qo}$, π_1 is strictly more efficient than π_2 at slot $k + 1$.
   - There are no packets of type p_{new} in $Q_1^{(2)}$ but there are packets of type q_{old} as well as q_{new} in $Q_2^{(2)}$. Since there is no packets of type new in $Q_1^{(2)}$, the existing old packet (or a fresher packet) must have been previously received at user 1. So the current age at user 1 is at most $k - k_{po}$ and will increase at most up to $k + 1 - k_{po}$ at slot $k + 1$. The best that a policy of type π_2 can do by action 3 is to drop the age at user 1 to $k + 1 - k_{po}$ and drop the age at user 2 to $k + 1 - k_{qo}$ at slot $k + 1$. π_1 can choose action 2 and perform strictly better than π_2 at slot $k + 1$ because the newest packet in $Q_1^{(2)}$ is fresher than any new packet in $Q_2^{(2)}$.
   - There are no packets of type q_{new} in $Q_2^{(2)}$ but there are packets of type p_{old} as well as p_{new} in $Q_2^{(1)}$. This case is similar to the previous case.
   - There are only packets of type p_{old} in $Q_1^{(2)}$ and q_{old} in $Q_2^{(2)}$. Since there is no packet of type new in the queues, it means that the old packets (or fresher ones) are previously received at the users under π_1. So under π_1, the current age at user 1 (resp. user 2) is at most $k - k_{po}$ (resp. $k - k_{qo}$) at this time and will increase at most by 1 at time $k + 1$. The age under π_2, however, can only become $k + 1 - k_{po}$ (resp. $k + 1 - k_{qo}$) if there is a delivery. Hence π_1 is at least as efficient as π_2.

So at each time, the average age of policy π_1 is smaller than or equal to that of policy π_2.

$\square$

APPENDIX B
PROOF OF LEMMA 2

Proof. From the analysis above, the probability of Case (i) is

$$P_1 = (\mu_1\epsilon_{12} + \mu_2 + \mu_3)^{\ell-1}\mu_1(1-\epsilon_1).$$

The probability of Case (ii-1) is

$$\begin{aligned}
P_2 =& \sum_{t=1}^{\ell-1}(\mu_1\epsilon_{12}+\mu_2+\mu_3)^{t-1}(\mu_1\epsilon_1+\mu_2+\mu_3\epsilon_1)^{\ell-1-t}\\
&\times\mu_1(\epsilon_1-\epsilon_{12})\mu_1(1-\epsilon_1)\\
&+\sum_{2\le d\le\ell}\sum_{t=1}^{\ell-d}\Big((\mu_1\epsilon_{12}+\mu_2+\mu_3)^{t-1}(\mu_1\epsilon_1+\mu_2+\mu_3\epsilon_1)^{\ell-d-t}\\
&\times(\mu_1(\epsilon_1-\epsilon_{12}))^2(\mu_2+\mu_1\epsilon_{12}+\mu_3\epsilon_1)^{d-2}\mu_3(1-\epsilon_1)\Big).
\end{aligned}$$

The probability of Case (ii-2) is

$$\begin{aligned}
P_3 =& \sum_{2\le d\le\ell}(\mu_1\epsilon_{12}+\mu_2+\mu_3)^{\ell-d}\\
&\times\mu_1(\epsilon_1-\epsilon_{12})(\mu_2+\mu_1\epsilon_{12}+\mu_3\epsilon_1)^{d-2}\mu_3(1-\epsilon_1).
\end{aligned}$$

Denote

$$\begin{aligned}
x_1 &= \mu_1\epsilon_{12}+\mu_2+\mu_3\\
y_1 &= \mu_1\epsilon_1+\mu_2+\mu_3\epsilon_1\\
z_1 &= \mu_1\epsilon_{12}+\mu_2+\mu_3\epsilon_1,
\end{aligned}$$

thus

$$\begin{aligned}
\Pr(I_1=\ell) &= \sum_{d\le\ell}\Pr(I_1=\ell,D_1=d)\\
&= P_1+P_2+P_3 = eq_1+eq_2+eq_3+eq_4.
\end{aligned}$$

where

$$\begin{aligned}
eq_1 =& x_1^{\ell-1}\mu_1(1-\epsilon_1)\\
eq_2 =& \mu_1^2(\epsilon_1-\epsilon_{12})(1-\epsilon_1)\frac{x_1^{l-1}-y_1^{l-1}}{x_1-y_1}\\
eq_3 =& \frac{(\mu_1(\epsilon_1-\epsilon_{12}))^2\mu_3(1-\epsilon_1)}{x_1-y_1}\\
&\times\left(\frac{x_1^{l-1}-z_1^{l-1}}{x_1-z_1}-\frac{y_1^{l-1}-z_1^{l-1}}{y_1-z_1}\right)\\
eq_4 =& \mu_1(\epsilon_1-\epsilon_{12})\mu_3(1-\epsilon_1)\frac{x_1^{l-1}-z_1^{l-1}}{x_1-z_1}.
\end{aligned}$$

By calculation, for $\ell\ge 2$,

$$\Pr(I_1=\ell) = \delta_1 x_1^{\ell-1}+\beta_1 y_1^{\ell-1} \qquad (33)$$

where

$$\begin{aligned}
\delta_1 &= \mu_1(1-\epsilon_{12})+\frac{\mu_1^2(\epsilon_1-\epsilon_{12})(1-\epsilon_{12})}{-\mu_1(\epsilon_1-\epsilon_{12})+\mu_3(1-\epsilon_1)}\\
\beta_1 &= -\frac{\mu_1(\epsilon_1-\epsilon_{12})(1-\epsilon_1)}{-\mu_1(\epsilon_1-\epsilon_{12})+\mu_3(1-\epsilon_1)}\left(\mu_1+\mu_3\right).
\end{aligned}$$

Now we check $\Pr(\ell=1)$. Substituting $\ell=1$ into (33), we have $P(I_1=\ell)=\delta_1+\beta_1=\mu_1(1-\epsilon_1)$, therefore

$$P(I_1=\ell) = \delta_1 x_1^{\ell-1}+\beta_1 y_1^{\ell-1}$$

holds for all $\ell\ge 1$. □

APPENDIX C
PROOF OF LEMMA 3

Proof. Taking the first derivative of (29) with respect to μ and letting it equal to zero, we have

$$\begin{aligned}
\mu^{(1)} &= \frac{1-\epsilon+\sqrt{(\epsilon-\epsilon_{12}(\epsilon))(1-\epsilon)}}{\epsilon_{12}(\epsilon)-2\epsilon+1}\\
\mu^{(2)} &= \frac{1-\epsilon-\sqrt{(\epsilon-\epsilon_{12}(\epsilon))(1-\epsilon)}}{\epsilon_{12}(\epsilon)-2\epsilon+1}.
\end{aligned}$$

If $\epsilon_{12}(\epsilon)-2\epsilon+1=0$, then substituting $\epsilon_{12}(\epsilon)=2\epsilon-1$ into (29), taking the derivative and equating to zero, we obtain $\mu^{(1)}=\mu^{(2)}=1/2$.

First, we simplify $\mu^{(2)}$. If $\epsilon_{12}(\epsilon)-2\epsilon+1\ne 0$, then

$$\mu^{(2)} = \frac{1-\epsilon-\sqrt{(\epsilon-\epsilon_{12}(\epsilon))(1-\epsilon)}}{\epsilon_{12}(\epsilon)-2\epsilon+1} \qquad (34)$$

$$= \frac{\sqrt{1-\epsilon}}{\epsilon_{12}(\epsilon)-2\epsilon+1}(\sqrt{1-\epsilon}-\sqrt{\epsilon-\epsilon_{12}(\epsilon)}) \qquad (35)$$

$$= \frac{\sqrt{1-\epsilon}}{\epsilon_{12}(\epsilon)-2\epsilon+1}\cdot\frac{\epsilon_{12}(\epsilon)-2\epsilon+1}{\sqrt{1-\epsilon}+\sqrt{\epsilon-\epsilon_{12}(\epsilon)}} \qquad (36)$$

$$= \frac{\sqrt{1-\epsilon}}{\sqrt{1-\epsilon}+\sqrt{\epsilon-\epsilon_{12}(\epsilon)}}. \qquad (37)$$

Note that (37) still holds when $\epsilon_{12}(\epsilon)-2\epsilon+1=0$. From (37), we have $\mu^{(2)}>0$.

Consider the following three cases:

(1) If $\epsilon_{12}(\epsilon)-2\epsilon+1<0$, then $\mu^{(1)}<0$. Also,

$$\begin{aligned}
\mu^{(2)} &= \frac{\sqrt{1-\epsilon}}{\sqrt{1-\epsilon}+\sqrt{\epsilon-\epsilon_{12}(\epsilon)}}\\
&< \frac{\sqrt{1-\epsilon}}{\sqrt{1-\epsilon}+\sqrt{1-\epsilon}} < 1/2.
\end{aligned}$$

$\mathbb{E}[J^R]$ in (29) decreases when $\mu\in[0,\mu^{(2)}]$ and increases when $\mu\in(\mu^{(2)},1/2]$, so the unique minimal point is obtained at $\mu^*=\mu^{(2)}$.

(2) If $\epsilon_{12}(\epsilon)-2\epsilon+1=0$, then $\mu^{(1)}=\mu^{(2)}=1/2$, and $\mathbb{E}[J^R]$ in (29) decreases when $\mu\in[0,1/2]$. So the unique minimum point is obtained at $\mu^*=1/2$.

(3) If $\epsilon_{12}(\epsilon)-2\epsilon+1>0$, then

$$\mu^{(1)} > \mu^{(2)} > 0.$$

Also,

$$\begin{aligned}
\mu^{(2)} &= \frac{\sqrt{1-\epsilon}}{\sqrt{1-\epsilon}+\sqrt{\epsilon-\epsilon_{12}(\epsilon)}}\\
&> \frac{\sqrt{1-\epsilon}}{\sqrt{1-\epsilon}+\sqrt{1-\epsilon}} > 1/2.
\end{aligned}$$

$\mathbb{E}[J^R]$ decreases when $\mu\in[0,1/2]$, so the unique minimum point is obtained at $\mu^*=1/2$. □

Proof. We first denote by $d_i(k) \in \{0,1\}$ and $t_i(k) \in \{0,1\}$, the number of packets delivered to user i during slot k from $Q_1^{(i)}$ and $Q_2^{(i)}$, respectively. We have

$$d_i(k) \leq 1$$
$$t_i(k) \leq 1$$
$$d_i(k) + t_i(k) \leq 1.$$

Thus from (1) and (2),

$$
\begin{aligned}
h_i(k+1) =& d_i(k) + t_i(k)(w_i(k)+1) \\
&+ (1 - d_i(k) - t_i(k))(h_i(k)+1).
\end{aligned}
\tag{38}
$$

Using (38), we can re-write the Lyapunov Drift as follows:

$$
\begin{aligned}
\Theta(\vec{h}(k)) =& \mathbb{E}\big[L(\vec{h}(k+1)) - L(\vec{h}(k))|\vec{s}(k)\big] \\
=& \frac{1}{2}\sum_{i=1}^{2} \alpha_i \mathbb{E}\big[\big((1 - d_i(k) - t_i(k))(h_i(k)+1) \\
&+ d_i(k) + t_i(k)(w_i(k)+1)\big)^2 - h_i^2(k)|\vec{s}(k)\big] \\
\overset{(a)}{=}& \frac{1}{2}\sum_{i=1}^{2} \alpha_i \mathbb{E}\big[\big((1 - d_i(k) - t_i(k))^2(h_i(k)+1)^2 \\
&+ d_i^2(k) + t_i^2(k)(w_i(k)+1)^2 - h_i^2(k)|\vec{s}(k)\big)\big] \\
\overset{(b)}{=}& \frac{1}{2}\sum_{i=1}^{2} \alpha_i \mathbb{E}\big[\big((1 - d_i(k) - t_i(k)) \cdot (h_i(k)+1)^2 \\
&+ d_i(k) + t_i(k)(w_i(k)+1)^2 - h_i^2(k)|\vec{s}(k)\big)\big] \\
=& \frac{1}{2}\sum_{i=1}^{2} \alpha_i \Big(\mathbb{E}[t_i(k)|\vec{h}(k)]\big(w_i^2(k) + 2w_i(k) - h_i^2(k) - 2h_i(k)\big) \\
&- \mathbb{E}[d_i(k)|\vec{h}(k)](h_i^2(k) + 2h_i(k)) + 2h_i(k) + 1\Big).
\end{aligned}
$$

(a) holds because from the definition of $d_i(k)$ and $t_i(k)$, at each slot only one of $d_i(k), t_i(k), 1 - d_i(k) - t_i(k)$ equals to 1, and the rest are zero. (b) holds because $d_i^2(k) = d_i(k)$, $t_i^2(k) = t_i(k)$, and $(1 - d_i(k) - t_i(k))^2 = 1 - t_i(k) - d_i(k)$.

Minimizing $\Theta(\vec{h}(k))$ is equivalent to minimize

$$
\begin{aligned}
\tilde{\Theta}(\vec{h}(k)) =& \frac{1}{2}\sum_{i=1}^{2} \alpha_i \Big(- \mathbb{E}[d_i(k)|\vec{s}(k)]\big(h_i^2(k) + 2h_i(k)\big) \\
&+ \mathbb{E}[t_i(k)|\vec{s}(k)]\big(w_i^2(k) + 2w_i(k) - h_i^2(k) - 2h_i(k)\big)\Big)
\end{aligned}
\tag{39}
$$

since $h_i(k)$ is a constant once we know $\vec{s}(k)$. Moreover, from the definition of MW policies, in each slot, $P(A(k) = i) \in \{0,1\}$ and $\sum_{i=1}^{3} P(A(k) = i) = 1$. Therefore, we have

$$\mathbb{E}[d_i(k)|\vec{s}(k)] = 1_{\{A(k)=i\}}(1 - \epsilon_i) \tag{40}$$
$$\mathbb{E}[t_i(k)|\vec{s}(k)] = 1_{\{A(k)=3\}}1_{\{w_i(k)>0\}}(1 - \epsilon_i). \tag{41}$$

Substituting (40) and (41) into (39), we obtain the desired results. □

Before obtaining the upper bound of Max-Weight policies, we consider any randomized policy with probability μ_1,

μ_2, and μ_3. Now we calculate the probability of $Q_2^{(i)} = \varnothing$. This can be found as a corollary of the following lemma:

Lemma 4. *Consider a queue Q of size 1. Suppose that packets arrive with a Bernoulli random process of rate λ and leave with another (independent) Bernoulli random process of rate μ. Then the probability of empty queue is*

$$P(Q = \varnothing) = \frac{\mu(1-\lambda)}{\mu + \lambda - \mu\lambda}.$$

Proof. Since the capacity of the system is 1, then only two state exists, one is "an empty queue", and the other is "a non-empty queue". Denote "an empty queue" as state 0 and "a non-empty queue" as state 1, and consider a two-state Markov chain. Then, we find the transition probabilities

$$
\begin{aligned}
P_{0\to1} =& P\{\text{a arrival occurs}\} = \lambda \\
P_{1\to0} =& P\{\text{a departure occurs but no arrivals occur}\} \\
=& \mu(1-\lambda).
\end{aligned}
$$

Thus the transition probability matrix is

$$P = \begin{bmatrix} 1 - \lambda & \lambda \\ \mu(1-\lambda) & 1 - \mu + \mu\lambda \end{bmatrix}.$$

Denote the stationary distribution as $\pi = (\pi_0, \pi_1)$, therefore

$$\pi = \pi P \Rightarrow \pi_0 = \frac{\mu(1-\lambda)}{\mu + \lambda - \mu\lambda}.$$

□

Corollary 1. *The probability of empty queue in $Q_2^{(i)}$ is*

$$P_{empty}^i := \frac{\mu_3(1-\epsilon_i)(1 - \mu_i(\epsilon_i - \epsilon_{12}))}{\mu_3(1-\epsilon_i) + \mu_i(\epsilon_i - \epsilon_{12}) - \mu_3\mu_i(1-\epsilon_i)(\epsilon_i - \epsilon_{12})}. \tag{42}$$

Proof. Since $Q_2^{(i)}$ is the queue in which packets are erased by user i but received user $\backslash i$, so the arrival rate is

$$\lambda = \mu_i(\epsilon_i - \epsilon_{12}),$$

and the departure rate

$$\mu = \mu_3(1 - \epsilon_i).$$

From Lemma 4,

$$
\begin{aligned}
P(Q_2^{(i)} = \varnothing) =& \frac{\mu_3(1-\epsilon_i)(1 - \mu_i(\epsilon_i - \epsilon_{12}))}{1 - (1 - \mu_3(1-\epsilon_i))(1 - \mu_i(\epsilon_i - \epsilon_{12}))} \\
=& \frac{\mu_3(1-\epsilon_i)(1 - \mu_i(\epsilon_i - \epsilon_{12}))}{\mu_3(1-\epsilon_i) + \mu_i(\epsilon_i - \epsilon_{12}) - \mu_3\mu_i(1-\epsilon_i)(\epsilon_i - \epsilon_{12})}.
\end{aligned}
$$

□

To obtain the upper bound of MW policies, we manipulate the expression of the one-slot Lyapunov Drift. From the proof of Theorem 4, we have

$$
\begin{aligned}
\Theta(\vec{h}(k)) =& \frac{1}{2}\sum_{i=1}^{2} \alpha_i \Big(- \mathbb{E}[d_i(k)|\vec{s}(k)]\big(h_i^2(k) + 2h_i(k)\big) + 2h_i(k) + 1 \\
&+ \mathbb{E}[t_i(k)|\vec{s}(k)]\big(w_i^2(k) + 2w_i(k) - h_i^2(k) - 2h_i(k)\big)\Big).
\end{aligned}
$$

Consider any randomized policy with probability (μ_1, μ_2, μ_3) in Section IV, and denote the probability of the non-empty

queue in $Q_2^{(i)}$ as $P_{ne}^i = 1 - P_{empty}^i$. From the definition of Max-Weight policies, and substituting μ_1, μ_2, μ_3 and P_{ne}^i into $\Theta(\vec{h}(k))$, we find

$$\Theta(\vec{h}(k)) \leq \frac{1}{2} \sum_{i=1}^{2} \alpha_i \big(-\mu_i(1-\epsilon_i)(h_i^2(k) + 2h_i(k)) + 2h_i(k) + 1$$
$$+ \mu_3 P_{ne}^i(1-\epsilon_i)(w_i^2(k) + 2w_i(k) - h_i^2(k) - 2h_i(k))\big).$$

From Remark 1, since $w_k(k) \leq h_i(k) - 1$, then

$$\Theta(\vec{h}(k)) \leq \frac{1}{2} \sum_{i=1}^{2} \alpha_i \big(-\mu_i(1-\epsilon_i)(h_i(k) - \Phi_i)^2 + \Psi_i \big)$$

where

$$\Phi_i = \frac{1 - \mu_3 P_{ne}^i(1-\epsilon_i) - \mu_i(1-\epsilon_i)}{\mu_i(1-\epsilon_i)}$$
$$\Psi_i = 1 - \mu_3 P_{ne}^i(1-\epsilon_i) + \frac{\big(1 - \mu_i(1-\epsilon_i) - \mu_3 P_{ne}^i(1-\epsilon_i)\big)^2}{\mu_i(1-\epsilon_i)}.$$

By Cauchy-Schwarz inequality,

$$\sum_{i=1}^{2} \alpha_i \mu_i(1-\epsilon_i)(h_i(k) - \Phi_i)^2 \big(\sum_{i=1}^{2} \frac{\alpha_i}{\mu_i(1-\epsilon_i)} \big)$$
$$\geq \big(\sum_{i=1}^{2} \alpha_i |h_i(k) - \Phi_i| \big)^2.$$

Therefore

$$\Theta(\vec{h}(k)) \leq -\frac{1}{2} \big(\sum_{i=1}^{2} \frac{\alpha_i}{\mu_i(1-\epsilon_i)} \big)^{-1} \big(\sum_{i=1}^{2} \alpha_i |h_i(k) - \Phi_i| \big)^2$$
$$+ \frac{1}{2} \sum_{i=1}^{2} \alpha_i \Psi_i,$$

which implies

$$\frac{1}{2} \big(\sum_{i=1}^{2} \alpha_i |h_i(k) - \Phi_i| \big)^2$$
$$\leq -\big(\sum_{i=1}^{2} \frac{\alpha_i}{\mu_i(1-\epsilon_i)} \big) \Theta(\vec{h}(k)) + \frac{1}{2} \sum_{i=1}^{2} \frac{\alpha_i}{\mu_i(1-\epsilon_i)} \sum_{i=1}^{2} \alpha_i \Psi_i.$$

Summing over $k \in \{1, 2, \cdots, T\}$, taking expectation and dividing by T results in

$$\frac{1}{2T} \Big(\sum_{k=1}^{T} \mathbb{E} \big(\sum_{i=1}^{2} \alpha_i |h_i(k) - \Phi_i| \big)^2 \Big)$$
$$\leq -\big(\sum_{i=1}^{2} \frac{\alpha_i}{\mu_i(1-\epsilon_i)} \big) \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}[\Theta(\vec{h}(k))] + \frac{1}{2} \big(\sum_{i=1}^{2} \frac{\alpha_i}{\mu_i(1-\epsilon_i)} \big) \sum_{i=1}^{2} \alpha_i \Psi_i.$$

By Jensen's inequality,

$$\frac{1}{2} \Big(\frac{1}{T} \sum_{k=1}^{T} \mathbb{E} \big(\sum_{i=1}^{2} \alpha_i |h_i(k) - \Phi_i| \big) \Big)^2$$
$$\leq -\big(\sum_{i=1}^{2} \frac{\alpha_i}{\mu_i(1-\epsilon_i)} \big) \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}[\Theta(\vec{h}(k))] + \frac{1}{2} \sum_{i=1}^{2} \frac{\alpha_i}{\mu_i(1-\epsilon_i)} \sum_{i=1}^{2} \alpha_i \Psi_i.$$

Since

$$-\frac{1}{T} \sum_{k=1}^{T} \mathbb{E}[\Theta(\vec{h}(k))] \leq \frac{\mathbb{E}[L(\vec{h}(1))]}{T}$$

and $\mathbb{E}[L(\vec{h}(1))]$ is a constant, then

$$\lim_{T \to \infty} -\frac{1}{T} \sum_{k=1}^{T} \mathbb{E}[\Theta(\vec{h}(k))] = 0.$$

Thus,

$$\mathbb{E}[J^{MW}] \leq \sqrt{\frac{1}{2} \sum_{i=1}^{2} \frac{\alpha_i}{\mu_i(1-\epsilon_i)} \sum_{i=1}^{2} \alpha_i \Psi_i} + \frac{1}{2} \sum_{i=1}^{2} \alpha_i \Phi_i$$

where

$$\Phi_i = \frac{1 - \mu_3 P_{ne}^i(1-\epsilon_i) - \mu_i(1-\epsilon_i)}{\mu_i(1-\epsilon_i)}$$
$$\Psi_i = 1 - \mu_3 P_{ne}^i(1-\epsilon_i) + \frac{\big(1 - \mu_i(1-\epsilon_i) - \mu_3 P_{ne}^i(1-\epsilon_i)\big)^2}{\mu_i(1-\epsilon_i)}.$$

REFERENCES

[1] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, June 2011, pp. 350–358.

[2] S. Kaul, R. Yates, and M. Gruteser, "On piggybacking in vehicular networks," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, Dec 2011, pp. 1–5.

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

[4] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, March 2012, pp. 1–6.

[5] C. Kam, S. Kompella, and A. Ephremides, "Effect of message transmission diversity on status age," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 2411–2415.

[6] C. Kam, S. Kompella, G. D. Nguyen, and A. Ephremides, "Effect of message transmission path diversity on status age," *IEEE Transactions on Information Theory*, vol. 62, no. 3, pp. 1360–1374, March 2016.

[7] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1897–1910, April 2016.

[8] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1807–1827, March 2019.

[9] J. Zhong, R. D. Yates, and E. Soljanin, "Timely lossless source coding for randomly arriving symbols," in *2018 IEEE Information Theory Workshop (ITW)*, Nov 2018, pp. 1–5.

[10] P. Mayekar, P. Parag, and H. Tyagi, "Optimal lossless source codes for timely updates," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 1246–1250.

[11] R. Devassy, G. Durisi, G. C. Ferrante, O. Simeone, and E. Uysal-Biyikoglu, "Delay and peak-age violation probability in short-packet transmissions," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 2471–2475.

[12] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Information freshness and popularity in mobile caching," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 136–140.

[13] R. D. Yates, P. Ciblat, A. Yener, and M. Wigger, "Age-optimal constrained cache updating," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 141–145.

[14] S. Zhang, J. Li, H. Luo, J. Gao, L. Zhao, and X. S. Shen, "Towards fresh and low-latency content delivery in vehicular networks: An edge caching aspect," in *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2018, pp. 1–6.

[15] A. Arafa, J. Yang, and S. Ulukus, "Age-minimal online policies for energy harvesting sensors with random battery recharges," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.

[16] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Online timely status updates with erasures for energy harvesting sensors," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2018, pp. 966–972.

[17] S. Feng and J. Yang, "Minimizing age of information for an energy harvesting source with updating failures," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 2431–2435.

[18] J. Yun and C. Joo and A. Eryilmaz, "Optimal Real-Time Monitoring of an Information Source Under Communication Costs," in *2018 IEEE Conference on Decision and Control (CDC)*, Dec 2018, pp. 4767–4772.

[19] Roy D. Yates, "The Age of Information in Networks: Moments, Distributions, and Sampling," arXiv:1806.03487, 2018.

[20] Tasmeen Zaman Ornee and Yin Sun, "Sampling for Remote Estimation through Queues: Age of Information and Beyond," arXiv:1902.03552, 2019.

[21] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 5381–5394, July 2018.

[22] C. Joo and A. Eryilmaz, "Wireless scheduling for information freshness and synchrony: Drift-based design and heavy-traffic analysis," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2556–2568, Dec 2018.

[23] B. Li, R. Li, and A. Eryilmaz, "Throughput-optimal scheduling design with regular service guarantees in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 23, no. 5, pp. 1542–1552, Oct 2015.

[24] B. Li , R. Li and A. Eryilmaz, "Wireless scheduling design for optimizing both service regularity and mean delay in heavy-traffic regimes," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1867–1880, June 2016.

[25] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, April 2018, pp. 1844–1852.

[26] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," arXiv:1801.01803, 2018.

[27] R. Talak, I. Kadota, S. Karaman, and E. Modiano, "Scheduling policies for age minimization in wireless networks with unknown channel state," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 2564–2568.

[28] R. Talak, S. Karaman, and E. Modiano, "Optimizing age of information in wireless networks with perfect channel state information," in *2018 16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2018, pp. 1–8.

[29] R. Talak, S. Karaman and E. Modiano, "Distributed scheduling algorithms for optimizing information freshness in wireless networks," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.

[30] K. Chen and L. Huang, "Age-of-information in the presence of error," arXiv:1605.00559, 2016.

[31] R. D. Yates, E. Najm, E. Soljanin, and J. Zhong, "Timely updates over an erasure channel," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 316–320.

[32] P. Parag, A. Taghavi, and J. Chamberland, "On real-time status updates over symbol erasure channels," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2017, pp. 1–6.

[33] E. Najm, E. Telatar, and R. Nasser, "Optimal age over erasure channels," arXiv:1901.01573, 2019.

[34] H. Sac, T. Bacinoglu, E. Uysal-Biyikoglu, and G. Durisi, "Age-optimal channel coding blocklength for an M/G/1 queue with HARQ," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.

[35] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback - capacity and algorithms," in *2009 Workshop on Network Coding, Theory, and Applications*, 54-61, Ed., June 2009.

[36] C. Wang, "Capacity of 1-to-k broadcast packet erasure channels with channel output feedback," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2010, pp. 1347–1354.

[37] M. Gatzianas, S. Saeedi Bidokhti, and C. Fragouli, "Feedback-based coding algorithms for broadcast erasure channels with degraded message sets," in *2012 International Symposium on Network Coding (NetCod)*, June 2012, pp. 108–113.

[38] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser broadcast erasure channel with feedback—capacity and algorithms," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5779–5804, Sep. 2013.

[39] M. Heindlmaier and S. Saeedi Bidokhti, "Capacity regions of two-receiver broadcast erasure channels with feedback and memory," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 5042–5069, July 2018.

[40] S. M. Ross, *STOCHASTIC PROCESSES, 2ND ED.* Wiley India Pvt. Limited, 2008.