# Timely Broadcasting in Erasure Networks: Age-Rate Tradeoffs

Xingran Chen, Renpu Liu*, Shaochong Wang*, Shirin Saeedi Bidokhti

*Abstract*—The interplay between timeliness and rate efficiency is investigated in packet erasure broadcast channels with feedback. A scheduling framework is proposed in which coding actions, as opposed to users, are scheduled to attain desired tradeoffs between rate and age of information (AoI). This tradeoff is formalized by an upper bound on AoI as a function of the target rate constraints and two lower bounds: one as a function of the communication rate and one as a function of the arrival rate. Simulation results show that (i) coding can be beneficial in reducing AoI in the regime of moderate arrival rates even without rate constraints and the benefit increases with the number of users, and (ii) AoI increases with both the target rate constraint and the arrival rate when either is kept fixed, but decreases with them when they are set to be equal.

## I. INTRODUCTION

The technology of Internet of Things (IoT) provides a vision for integrating intelligence into cyber-physical systems using real-time applications. Timeliness is key for such applications and it has therefore emerged as a communication design criteria. There are, however, tradeoffs between timeliness and rate which we aim to investigate in broadcast networks.

Timeliness is measured using the metric of *Age of Information (AoI)*, as introduced in [1]. AoI captures, at the receiving side, how much time has passed since the generation time of the latest received packet. In the past decade, Age of information has been extensively investigated for status update systems [2]–[7]. From the aspect of scheduling, optimal transmission policies were proposed in [8]–[12] to optimize the overall age in wireless networks. The reader is referred to [13], [14] for a survey on the topic.

Rate efficiency is often provided by channel coding schemes over multiple realizations of the network and it comes at the cost of large delays. It is, therefore, not clear a-priori what types of tradeoffs exist between rate and timeliness. In coding theory, previous works have mainly studied point to point channels [15]–[17]. In point to point erasure channels, [18] proves that when the source alphabet and channel input alphabet have the same size, a Last-Come First-Serve (LCFS) policy with no coding is optimal. This is in contrast to channel coding schemes that provide rate efficiency by block coding. Considering erasure channels with FCFS M/G/1 queues, [19]

Xingran Chen, Shaochong Wang, Renpu Liu and Shirin Saeedi Bidokhti are with the Department of Electrical and System Engineering, University of Pennsylvania, PA, 19104. E-mail: {xingranc, tainerl, shaowang, saeedi}@seas.upenn.edu

*: contributed equally

finds an optimal block length for channel coding to minimize the average age and average peak age. In the context of broadcast packet erasure channels (BPECs) with feedback, coding is shown to be beneficial for age efficiency with two users [20]. In related work, [21], [22] design optimal precoding schemes to minimize AoI in a MIMO broadcast channel with multiple senders and receivers under FIFO channels without packet management. Reference [23] analyzes the AoI in a multicast network with network coding.

In this work, we consider erasure networks and devise broadcast strategies that are efficient both in AoI and rate. The inherent tradeoff can be explained as follows. On the one hand, 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), hence may correspond to a smaller AoI. On the other hand, to achieve high rates with coding, we have to incur delay by waiting for the arrival/generation of other packets for the purpose of coding as well as prioritizing their transmission, and this leads to larger AoI. To shed light on the above tradeoff, we build on our previous work [20] and consider an erasure wireless network with $M$ users. Motivated by the success of age-based scheduling in wireless networks, we propose a scheduling framework where we schedule various useful coding actions as opposed to scheduling the users. Within this framework, we can capture both rate efficiency as well as age efficiency. In particular, we design deterministic policies that minimize the expected weighted sum of AoI (EAoI) under given target rate constraints.

The contributions of the work are summarized as follows: (i) We propose a novel framework of network AoI on the broadcast channels under transmission mechanism with coding (Section II). (ii) (Near-)optimal coding policies with uncoded and coded caching are proposed (Section III). Two general lower bounds and an upper bound are derived on EAoI for any transmission policy (Section V, Theorem 2). The bounds are functions of generation rates, erasure probabilities and target rate constraints. (iii) Simulation results reveal that (a) coding is beneficial, and the benefits increase with the number of users; (b) a good approximation of proposed policies is obtained based on maximum clique size of information graph; (c) the tradeoff between rate and AoI exists, which implies that the system has to sacrifice AoI to achieve a higher rate.

## A. Notation

We use the notations $\mathbb{E}(\cdot)$ and $\Pr(\cdot)$ for expectation and probability, respectively. We denote scalars with lower case letters, e.g. $s$. Denote vectors as lower case letters with underline, e.g., $\underline{s}$. Random variables are denoted by capital letters, e.g. $S$. Sets are denoted by calligraphy letters, e.g. $\mathcal{S}$. We use $M$ to denote the number of users, $K$ to denote the time horizon. $[n]$ denotes the set $\{1, 2, \cdots, n\}$. For two sets $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{A} \subset \mathcal{B}$ represents that $\mathcal{A}$ is a subset of $\mathcal{B}$.

## II. System Model

The system model extends that of [20, Section VI] to $M$ users. In particular, transmission occurs in a wireless network which we model by a Broadcast Packet Erasure Channel (BPEC) with $M$ users. In the beginning of time slot $k$, a packet intended for user $i$ is generated with probability $\theta_i$. Let $G_i(k) = 1$ represent that a new packet is generated (for user $i$) in time slot $k$. So $\Pr\left(G_i(k) = 1\right) = \theta_i$ for $k = 1, 2, \ldots$.

Every broadcast packet is received at user $i$ with probability $1 - \epsilon_i$, $0 \le \epsilon_i < 1$, and lost with probability $\epsilon_i$. Erasure events at multiple users can be dependent in general. The *transmission delay* is assumed fixed and equal to one time slot. After each transmission, the transmitter receives ACK/NACK feedback from all receivers and can thus calculate and track the aging of information at each user. Let $d_i(k) = 1$ if user $i$ decodes a packet of type $i$ in time $k$, and $d_i(k) = 0$ otherwise.

If a packet is not received at its intended user, it can be cached by other user(s) that have received it. The cached packets can act as side information. Using the available feedback, the encoder can track the cached packets and exploit them as side information in the code design to form more efficient coded packets that are simultaneously useful for multiple users. Such code designs have also appeared in [24]–[27] for rate-efficiency.

We call a packet *a coded* if it is formed by combining more than one packets; otherwise we call it an *uncoded packet*. Consider a coded packet $x$. If user $i$ can instantaneously *decode* a packet that is intended for him upon successful delivery of $x$ (possibly using its locally cached packets), we call user $i$ a *destination* for packet $x$. A coded packet can be *fully decoded* by user $i$ if user $i$ extracts every uncoded packets combined within it upon successful delivery.

Depending on the available caching and coding capabilities, we can consider three class of policies: (i) policies that benefit from coding by caching uncoded packets, (ii) policies that benefit from coding by caching general (potentially coded) packets, and (iii) policies that schedule different users and perform no caching/coding [10], [11] (time-sharing policies). We investigate the first class in Section III and refer them as coding policies with uncoded caching. The second class, referred by coding policies with coded caching, is investigated in Section IV. Time-sharing policies from benchmarks in our

simulations in Appendix A. In this work, we consider coding policies with uncoded caching and linear network coding through XOR operations only. Coding over larger finite fields may impose larger decoding delay and is often practically less desirable.

## A. A Virtual Network of Queues

The idea of caching and coding on the fly is to cache overheard packets at the users and track them using feedback at the encoder through a *network of virtual queues*.

*1) Virtual Queues in Uncoded Caching:* Let $Q_i$ denote the queue of incoming packets for user $i$. If a packet chosen from $Q_i$ is transmitted and received by its intended user, it is removed from the queue. If it is not received by its intended user $i$, but received by some other users, then the packet will be cached in the cache of those users (as side information) and tracked at a virtual queue at the encoder. The buffer size of each virtual queue is assumed infinite. Define $Q_{i,\mathcal{S}}$ as the virtual queue that tracks, *at the encoder*, uncoded packets for user $i$ that are received *only* by the users in $\mathcal{S}$, where $\mathcal{S} \subset [M] \backslash i$. Note that $Q_i \left(= Q_{i,\varnothing}\right)$ is some sort of $Q_{i,\mathcal{S}}$. Queue $Q_{i,\mathcal{S}}$ contains two types of packets: packets from $Q_i$ that are cached (received or decoded) by the users in $\mathcal{S}$, and/or uncoded packets combined within coded packets which are fully decoded. The queues $Q_{i,\mathcal{S}}$ are defined so that the set of packets in them are disjoint.

Packets stored in the virtual queues at the encoder can form efficient coded packets that are simultaneously useful for multiple users. In this work, we consider linear network coding through XOR operations only. This is because for broadcast erasure channel with multiple unicast traffic, using simple coding operations leads to low decoding delay and is also practically desirable [28]. For example, consider a packet in $p_1$ in $Q_{1,\{2\}}$ and a packet $p_2$ in $Q_{1,\{2\}}$. the XOR packet $x = p_1 + p_2$ is useful for both users 1 and 2 because user 2 has cached packet $p_1$ and user 1 has cached packet $p_2$ and they can therefore recover their desired packets by XORing packet $x$ with their respective cached packet. More generally, consider a set of non-empty queues $\{Q_{\tau_i, \mathcal{S}_{\tau_i}}\}_{i=1}^{\ell}$ where $\tau_i$ is a user index ($\tau_i \in [M]$) and $\mathcal{S}_{\tau_i}$ is a subset of $[M] \backslash \tau_i$. Suppose the following condition holds:

$$\mathcal{S}_{\tau_i} \supset \{\cup_{j=1, j \ne i}^{\ell} \tau_j\} \qquad \forall i = 1, \ldots \ell. \tag{1}$$

Then XORing packets

$$p_i \in Q_{\tau_i, \mathcal{S}_{\tau_i}}$$

forms a coded packet $x$ as follows

$$x = \bigoplus_{i=1}^{\ell} p_i \tag{2}$$

which is simultaneously decodable at all users $\{\tau_1, \ldots, \tau_\ell\}$. To view condition (1) alternatively, draw a side information graph $\mathcal{G}$ with nodes $V = \{1, \ldots, M\}$. Add an edge between
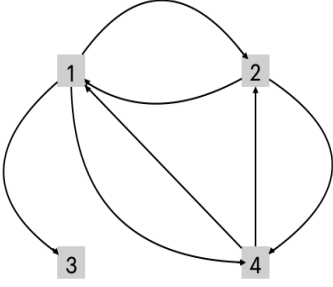
Fig. 1: Each clique corresponds to a coded packet.

nodes $(i,j)$ if $Q_{i,\mathcal{S}_i}$ is non-empty for some set $\mathcal{S}_i$ that has $j$ as an element. On this graph, condition (1) corresponds to the subgraph induced by nodes $\{\tau_1, \ldots, \tau_\ell\}$ forming a clique of size $\ell$. An example with a clique of size 3 is shown in Figure 1 with 4 users. Consider packets $p_1 \in Q_{1,\{2,3,4\}}$, $p_2 \in Q_{2,\{14\}}$ and $p_4 \in Q_{4,\{12\}}$. Let clique $1 \leftrightarrow 2 \leftrightarrow 4 \leftrightarrow 1$ corresponds to the coded packet $x = p_1 \oplus p_2 \oplus p_4$.

The *coding actions* we consider in this section correspond to cliques on the side information graph (which has to be updated on the fly after each transmission). In this class, maximal cliques are sufficient to consider among all cliques because sending a coded packet that corresponds to a subset of cliques is at most as useful (in terms of the users at which coded packets are decodable) as a coded packet that corresponds to a maximal clique. Among all possible maximal cliques (the number of which can generally be on the order of $3^{\frac{M}{3}}$ [29]) we aim to choose (schedule) one that leads to a coding action with the most benefit in terms of information freshness and rate.

*2) Virtual Queues in Coded Caching:* Under coding policies with coded caching, coded packets can be cached by users. Besides queues $Q_{i,\mathcal{S}}$, we introduce a new class of queues, denoted by $Q_{\mathcal{I},\mathcal{K}}$. In $Q_{\mathcal{I},\mathcal{K}}$, coded packets are formed by uncoded packets which are intended for users in $\mathcal{I}$ only, and cached by users in $\mathcal{K}$ only. Note that the intersection of $\mathcal{I}$ and $\mathcal{K}$ can be non-empty. The coding situation is extremely sophisticated when the number of users is large. Then, to simplify the analysis, we specialize our coding policies with coded caching by the following assumptions,

**Assumption 1.** (1) $\mathcal{I} \cap \mathcal{K} = \varnothing$.
(2) If $p \in Q_{\mathcal{I},\mathcal{K}}$ and $x = \oplus_{j \in [l]} p_j$, then $p_1, p_2, \cdots, p_l$ are *intended for distinct users in $\mathcal{I}$.*

Assumption 1 (1) implies that user $i$ can not cache a coded packet which contains a uncoded packet intended for oneself. In other words, if user $i$ receives a coded packet containing an uncoded packet intended for oneself, he can decode this uncoded packet. Under Assumption 1 (2), two uncoded packets which are intended for an identical user can not be encoded in a coded packet. Define the *length* of a coded packet as the number of uncoded packets encoded within it. From Assumption 1 (2), the length of a coded packet is at most $M$.

Similar to Section II-A1, consider a set of non-empty queues $\{Q_{\tau_i, \mathcal{S}_{\tau_i}}\}_{i=1}^{\ell_1}$ and $\{Q_{\mathcal{I}_j, \mathcal{K}_j}\}_{j=1}^{\ell_2}$, suppose the following conditions hold:

$$\mathcal{S}_{\tau_m} \supset \{\tau_i\}_{i=1, i \neq m}^{\ell_1} \cup \{\mathcal{I}_j\}_{j=1}^{\ell_2}, \quad \forall m = 1, \ldots \ell_1$$
$$\mathcal{K}_m \supset \{\tau_i\}_{i=1}^{\ell_1} \cup \{\mathcal{I}_j\}_{j=1, j \neq m}^{\ell_2}, \quad \forall m = 1, \ldots \ell_2 \quad (3)$$
$$\tau_i \cap \mathcal{I}_j = \varnothing, \quad \mathcal{I}_i \cap \mathcal{I}_j = \varnothing.$$

The last condition in (3) comes from Assumption 1 (2) directly. Then XORing packets

$$p_i \in Q_{\tau_i, \mathcal{S}_{\tau_i}}, \quad x_j \in Q_{\mathcal{I}_j, \mathcal{K}_j}$$

forms a coded packet $x$ as follows

$$x = \bigoplus_{i=1}^{\ell_1} p_i \bigoplus_{j=1}^{\ell_2} x_j \quad (4)$$

which is simultaneously decodable at all users $\{\tau_1, \ldots, \tau_{\ell_1}\} \cup \{\mathcal{I}_j\}_{j=1}^{\ell_2}$. To view condition (3) alternatively, draw a side information graph $\mathcal{G}'$ with nodes $V' = \{1, \ldots, M\} \cup \{\mathcal{I}_i\}_i$[1]. Add an edge between nodes $(i, m)$ (respectively, $(\mathcal{I}_j, m)$) if $Q_{i,\mathcal{S}_i}$ (respectively, $Q_{\mathcal{I}_j, \mathcal{K}_j}$) is non-empty for some set $\mathcal{S}_i$ (respectively, $\mathcal{K}_j$) that has $m$ as an element. On this graph, condition (3) corresponds to the subgraph induced by a clique of size $\ell_1 + \ell_2$.

*B. Age and Rate Efficiency*

To capture the freshness of information, we use the metric of *Age of Information* (AoI) defined in [13]. Denote $h_i(k)$ as the AoI of user $i$ in time slot $k$. The age function $h_i(k)$ increases linearly in time when no delivery for user $i$ occurs and drops with every delivery to a value that represents how old the received packet is. If an outdated packet (for user $i$) is received (meaning that a more recently generated packet is previously received at user $i$) then the outdated packet does not offer age reduction and $h_i(k)$ keeps increasing linearly.

**Definition 1.** *Denote the generation time of the packet received by user $i$ in time slot $k$ as $v_i(k)$. Assuming the initial state $h_i(0) = 1$, the age function $h_i(k)$ evolves as follows:*

$$h_i(k) = \begin{cases} \min\{h_i(k-1) + 1, k - v_i(k)\} & d_i(k) = 1 \\ h_i(k-1) + 1 & d_i(k) = 0. \end{cases}$$

The expected weighted sum of AoI (EAoI) at the users is thus given by $\mathbb{E}[J_K^\pi]$ and

$$J_K^\pi := \frac{1}{MK} \sum_{k=1}^{K} \sum_{i=1}^{M} \alpha_i h_i^\pi(k) \quad (5)$$

where $\alpha_1, \alpha_2, \cdots, \alpha_M$ are weights and the superscript $\pi$ represents the communication policy. We are interested in minimizing EAoI under some constraints on the rate of communications. We define the *communication rate* to user $i$ as the number of decoded packets (intended for user $i$) per time

---

[1]$\mathcal{I}_i$ can be regarded as a "big" node.

slot in the limit of time. The larger the rate, the fewer packet in the network of virtual queues at the encoder.

Let $q_i$ be a strictly positive real value that represents the minimum rate requirement of node $i$. Without loss of generality, we assume that $\underline{q} = (q_1, q_2, \cdots, q_M)$ is in the capacity region. Similar to [9], we define the *long-term rate* of node $i$ when policy $\pi$ is employed as

$$r_i^\pi := \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\big[d_i^\pi(k)\big]. \tag{6}$$

Then, we express the *minimum rate constraint* of each individual node as

$$r_i^\pi \geq q_i, i = 1, 2, \cdots, M. \tag{7}$$

Ultimately, we seek to schedule the coding actions in order to achieve a judicious tradeoff between the EAoI and communication rate, as outlined below. Combining (5), (6) and (7), the objective is given by the following optimization problem:

$$
\begin{aligned}
J(\underline{q}) := \min_{\pi} \quad & \lim_{K \to \infty} \mathbb{E}[J_K^\pi] \\
s.t \quad & r_i^\pi \geq q_i, i = 1, 2, \cdots, M.
\end{aligned}
\tag{8}
$$

## III. SCHEDULING CODING ACTIONS IN UNCODED CACHING

In this section, we consider coding policies with uncoded caching, i.e., all cached packets are uncoded. We develop and analyze max-weight policies that schedule the coding actions to optimize (8). Our results can be generalized to the case where we allow the caching of coded packets as outlined in Section IV.

Each coding action can be described by a set of queues, each storing multiple packets. We allow packet management in choosing which packets of the chosen queues to use to form coded packets as it reduces age without impacting rate.

In order to optimize for age, we first define the AoI of queues $Q_{i,\mathcal{S}}$ in the virtual network of queues (at the encoder) and explain their time evolution. Recall that $Q_{i,\mathcal{S}}$ is the queue that contains those packets of user $i$ that are decodes only by the users in $\mathcal{S}$. Thus, if $a \in Q_{i,\mathcal{S}}$, for any $\mathcal{S}' \subset \mathcal{S}$ and $\mathcal{S}' \neq \mathcal{S}$, $a \notin Q_{i,\mathcal{S}'}$. In addition, if $a \in Q_{i,\varnothing}$, then $a \notin Q_{i,\mathcal{S}}$ for all $\mathcal{S} \neq \varnothing$. So the map from packets to queues is a surjection. From Section II-A, the encoder decides among the following actions, denoted by $A(k)$, and defined below:

- $A(k) = Q_{i,\varnothing}$: a packet is transmitted from $Q_{i,\varnothing}$;
- $A(k) = \oplus_{j=1}^{l} Q_{\tau_j, \mathcal{S}_{\tau_j}}$: a coded packet is transmitted that is formed by an XOR of $l$ packets, one from each of the queues $Q_{\tau_1, \mathcal{S}_{\tau_1}}, Q_{\tau_2, \mathcal{S}_{\tau_2}}, \cdots, Q_{\tau_l, \mathcal{S}_{\tau_l}}$, where $S_{\tau_l} \not\ni \tau_l$ and users $\tau_1, \tau_2, \cdots, \tau_l$ form a *maximal* clique on the side information graph.

### A. Encoder's Age of Information

To capture the aging of information at the encoder, we define a notion of AoI of each virtual queue. The following Lemma is proved in Appendix B

**Lemma 1.** *If $p_j \in Q_{i,\mathcal{S}}$ has the generation time $k_j$, $j \in \{1, 2\}$, and $k_2 > k_1$, then (encoding and) transmitting $p_2$ can not be worse than (encoding and) transmitting $p_1$ in terms of AoI.*

If $\mathcal{S} = \varnothing$, denote the AoI of $Q_{i,\varnothing}$ by $w_{i,\varnothing}(k)$, and the generation time of latest packet by $k'$. Based on Lemma 1, we define

$$w_{i,\varnothing}(k) = \min\{k - k', h_i(k)\}$$

and $w_{i,\varnothing}(0) = h_i(0)$. This is to capture the fact that if $k - k' > h_i(k)$, then packets in $Q_{i,\varnothing}$ are older than the latest one recovered by user $i$, so packets in $Q_{i,\varnothing}$ are obsolete in terms of AoI in time slot $k$. The evolution of the AoI at the queue $Q_{i,\varnothing}$ is as follows: $w_{i,\varnothing}(k)$ drops to $0$ if a new packet is generated; otherwise it increases by $1$. Thus, the recursion of $w_{i,\varnothing}(k)$ is

$$
w_{i,\varnothing}(k+1) = \begin{cases} 0 & G_i(k) = 1 \\ \min\{w_{i,\varnothing}(k)+1, h_i(k)+1\} & G_i(k) = 0 \end{cases}.
\tag{9}
$$

Before defining the AoI of $Q_{i,\mathcal{S}}$, let $t_{i,\mathcal{S}}(k)$ be an indicator function as follows: $t_{i,\mathcal{S}}(k) = 1$ if the latest packet in $Q_{i,\mathcal{S}}$ is encoded and transmitted in time slot $k$, and is $t_{i,\mathcal{S}}(k) = 0$ otherwise. In Figure 1, if the encoder transmits $p = a_1 \oplus a_2 \oplus a_4$ in time slot $k$, then $t_{1,\{2,3,4\}}(k) = 1$, $t_{2,\{1,4\}}(k) = 1$ and $t_{4,\{1,2\}}(k) = 1$. Now we consider the AoI of $Q_{i,\mathcal{S}}$ with $\mathcal{S} \neq \varnothing$. Denote the AoI of $Q_{i,\mathcal{S}}$ as $w_{i,\mathcal{S}}(k)$. Let the generation time of the latest packet in $Q_{i,\mathcal{S}}$ be $k'$. We define

$$w_{i,\mathcal{S}}(k) = \min\{k - k', h_i(k)\}$$

and $Q_{i,\mathcal{S}}(0) = h_i(0)$. Then, $w_{i,\mathcal{S}}(k)$ increases by $1$ unless $Q_{i,\mathcal{S}}$ is updated with a fresher packet. The content of $Q_{i,\mathcal{S}}$ change when packets move in other virtual queues at the encoder. For example, If packet $a \in Q_{i,\mathcal{S}}$ is recovered by other users in $\mathcal{I}$, $\mathcal{I} \cap \mathcal{S} = \varnothing$, then $a \in Q_{i,\mathcal{I} \cup \mathcal{S}}$ and $a \notin Q_{i,\mathcal{S}}$. Let $d_{\mathcal{S}}(k) = 1$ represent that only users in $\mathcal{S}$ receive a (coded) packet. Let $\mathcal{S}' \subset \mathcal{S}$. The recursion of $w_{i,\mathcal{S}}(k)$ is

$$
w_{i,\mathcal{S}}(k+1) = \begin{cases} \min\{w_{i,\varnothing}(k)+1, h_i(k)+1\} & \mathcal{P}_{i,\mathcal{S}}^{(1)}(k) \\ \min\{w_{i,\mathcal{S}'}(k)+1, h_i(k)+1\} & \mathcal{P}_{i,\mathcal{S}}^{(2)}(k) \\ \min\{w_{i,\mathcal{S}}(k)+1, h_i(k)+1\} & \text{otherwise} \end{cases}.
\tag{10}
$$

where

$$\mathcal{P}_{i,\mathcal{S}}^{(1)}(k) = \{d_i(k) = 0, d_{\mathcal{S}}(k) = 1, t_{i,\varnothing}(k) = 1\}$$

and

$$\mathcal{P}_{i,\mathcal{S}}^{(2)}(k) = \{\mathcal{S}' \subset \mathcal{S}, d_i(k) = 0, d_{\mathcal{S} \setminus \mathcal{S}'}(k) = 1, t_{i,\mathcal{S}'}(k) = 1\}.$$

From (9) and (10), the recursion of $h_i(k)$ is

$$
h_i(k+1) = \begin{cases} w_{i,\mathcal{S}}(k) + 1 & t_{i,\mathcal{S}}(k) = 1, d_i(k) = 1 \\ h_i(k) + 1 & \text{otherwise} \end{cases}.
\tag{11}
$$

## B. Age-Rate Max-Weight Scheduling

It is well established that coding actions can enhance the communication rate of broadcast channels [24], and may incur additional delays. To seek efficiency both in AoI and communication rate, similar to [9], [11], [20], we propose Age-Rate Max-Weight (ARM) policies to minimize EAoI in (8) under rate constraints.

We define the *age-gain* of queue $Q_{i,\mathcal{S}}$ (for user $i$), where $\mathcal{S} \subset [M]\backslash i$ as follows:

$$\delta_{i,\mathcal{S}}(k) = h_i(k) - w_{i,\mathcal{S}}(k). \tag{12}$$

The term $\delta_{i,\mathcal{S}}(k)$ quantifies how much the instantaneous user's age of information reduces upon successful delivery from the encoder's virtual queue $Q_{i,\mathcal{S}}$. If $Q_{i,\mathcal{S}}$ is empty or contains old packets, then by the definition of $w_{i,\mathcal{S}}(k)$, $\delta_{i,\mathcal{S}}(k) = 0$.

Let $x_i(k)$ be the throughput debt associated with node $i$ at the beginning of slot $k$ [9]. It evolves as follows:

$$x_i(k+1) = kq_i - \sum_{\tau=1}^{k} d_i^{\pi}(\tau). \tag{13}$$

The value of $kq_i$ is the minimum average number of packets that node $i$ should have decoded by slot $k+1$ and $\sum_{\tau=0}^{k} d_i^{\pi}(\tau)$ is the total number of recovered packets in the same interval. In fact, strong stability of the process $x_i^+(k)$ is sufficient to establish that the minimum rate constraint, $r_i^{\pi} \geq q_i$, is satisfied [9], [30, Theorem 2.8].

Define the encoder's state in time slot $k$ as

$$S(k) = \Big(\{h_i(k)\}_i, \{w_{i,\mathcal{S}}(k)\}_{i,\mathcal{S}}, \{x_i(k)\}_i\Big),$$

and the Lyapunov function $L\big(S(k)\big)$ as

$$L\big(S(k)\big) = \sum_{i=1}^{M} \beta_i h_i(k) + \lambda \sum_{i=1}^{M} \big(x_i^+(k)\big)^2 \tag{14}$$

where $\beta_i, \lambda > 0$. Here, the quadratic function for $x_i(k)$ is to maximize the rate [8], [9], [11], and the linear function for $h_i(k)$ is to simplify the derivation. The one-slot Lyapunov Drift is defined as

$$\Theta(k) = \mathbb{E}\Big[L\big(S(k+1)\big) - L\big(S(k)\big)|S(k)\Big]. \tag{15}$$

Define the *rate-gain* of user $i$ in time slot $k$ as follows:

$$f_i(k) = \Big(\big(x_i(k) + q_i\big)^+\Big)^2 - \Big(\big(x_i(k) + q_i - 1\big)^+\Big)^2. \tag{16}$$

**Definition 2.** *In each slot $k$, the ARM policy chooses the action that has the maximum weight in Table of Fig. 2.*

**Remark 1.** *When AoI is the only metric in decision making (i.e., $q_i = 0$ for all $i$), only the latest packets matters (see Lemma 1). We can thus assume that the buffer size of every queue is 1 and the stability region is $\{\theta_i \leq 1, i = 1, 2, \cdots, M\}$.*

| $A(k)$ | Weights |
|---|---|
| $Q_{i,\varnothing}$ | $(1 - \epsilon_i)\Big(\beta_i \delta_{i,\varnothing}(k) + \lambda f_i(k)\Big)$ |
| $\oplus_{u \in [l]} Q_{\tau_u, \mathcal{S}_{\tau_u}}$ | $\sum_{u=1}^{l} \beta_{\tau_u} \delta_{\tau_u, \mathcal{S}_{\tau_u}}(k)(1 - \epsilon_{\tau_u})$ $+ \lambda \sum_{u=1}^{l}(1 - \epsilon_{\tau_u})f_{\tau_u}(k)$ |

Fig. 2: Coding actions and their weights.

**Remark 2.** *We have observed in simulations that a good approximation of the above ARM policy is obtained by choosing the maximal clique size $l$ be 2.*

**Theorem 1.** *The ARM policy defined in Definition 2 minimizes the one-slot Lyapunov Drift in each slot.*

*Proof.* The proof of Theorem 1 is given in Appendix C. $\square$

Let $C^{uncoded}$ be the set of all tuples $\underline{q}$ for which $x_i^+(k)$ is strongly stabilized using the considered coding actions. Upper bounding $J(\underline{q})$ for the ARM policy with a specific choice of $\beta$, we prove the following result in Appendix D.

**Theorem 2.** *For any $\underline{q} \in C^{uncoded}$, we have the following upper bound on $J(\underline{q})$:*

$$J(\underline{q}) \leq \frac{1}{M} \sum_{i=1}^{M} \Big(\frac{\alpha_i}{\theta_i} + \frac{\alpha_i}{q_i(1 - \epsilon_i)}\Big) + \lambda. \tag{17}$$

## IV. SCHEDULING CODING ACTIONS IN CODED CACHING

In this section, we consider coded caching, where coded packets can be cached in caches. Caching coded packets can bring more coding opportunity. However, the number of coding actions increases exponentially due to caching coded packets. For clarity of ideas, we explain and derive for $M = 3$, one can extend in a straightforward manner. Denote the packets intended for user 1, user 2 and user 3 as $\{a_t\}_{t \geq 1}$, $\{b_t\}_{t \geq 1}$ and $\{c_t\}_{t \geq 1}$, respectively. From Assumption 1, the length of coded packets is at most 3; if coded packets have length 3, then they must have forms as $a_{t_1} \oplus b_{t_2} \oplus c_{t_3}$. Thus, a coded packet which is cached must have length 2.

Let $\mathcal{X} = \big\{\{1,2\}, \{1,3\}, \{2,3\}\big\}$, $\mathcal{I} \in \mathcal{X}$ and $u = [3]\backslash\mathcal{I}$. In queue $Q_{\mathcal{I},u}$, coded packets for users in $\mathcal{I}$ are cached by user $u$. It is easy to show that if $u_1 \neq u_2$, then the encoder would not encode packets from $Q_{\mathcal{I}_1,u_1}$, $Q_{\mathcal{I}_2,u_2}$. In fact, suppose that $x_1 \in Q_{\mathcal{I}_1,u_1}$, $x_2 \in Q_{\mathcal{I}_2,u_2}$, then the length of coded packet $x = x_1 \oplus x_2$ is 4, which contradicts with Assumption 1 (2). Therefore, in each time slot $k$, the encoder decides among the following actions, denoted by $A(k)$, and defined below:

- $A(k) = Q_i$: a packet is transmitted from $Q_{i,\varnothing}$;
- $A(k) = \oplus_{j=1}^{l} Q_{\tau_j, \mathcal{S}_{\tau_j}}$: a coded packet is transmitted from $Q_{\tau_1, \mathcal{S}_{\tau_1}}, Q_{\tau_2, \mathcal{S}_{\tau_2}}, \cdots, Q_{\tau_l, \mathcal{S}_{\tau_l}}$, where users $\tau_1, \tau_2, \cdots, \tau_l$ form a maximal clique on the side information graph, $l = 2, 3$.
- $A(k) = Q_{\tau_i, \mathcal{S}_{\tau_i}} \oplus Q_{\mathcal{I},u}$: a coded packet is transmitted from $Q_{\tau_i, \mathcal{S}_{\tau_i}}, Q_{\mathcal{I},u}$, where $\tau_i, \mathcal{I}$ form a (general) maximal clique on the side information graph.

Then, we consider packets in $Q_{\mathcal{I},u}$. Without loss of generality, let $\mathcal{I} = \{1, 2\}$, $u = 3$, and $x_1, x_2 \in Q_{\{1,2\},3}$, $x_1 = a_{t_1} \oplus b_{t_2}$, $x_2 = a_{t_3} \oplus b_{t_4}$. Denote the generation time of $a_{t_1}$, $a_{t_2}$, $b_{t_2}$, $b_{t_4}$ as $u(t_1)$, $u(t_3)$, $v(t_2)$, $v(t_4)$, respectively.

**Lemma 2.** *If $\alpha_1 u(t_1) + \alpha_2 v(t_2) > \alpha_1 u(t_3) + \alpha_2 v(t_4)$, then (encoding and) transmitting $x_2$ can not be worse than (encoding and) transmitting $x_1$ in terms of AoI.*

*Proof.* The proof of Lemma 2 is given in Appendix E. □

Let $\mathcal{I} = \{i, j\}$. For any $x \in Q_{\mathcal{I},u}$, denote the corresponding generation time vector as $(u_i, v_j)$. Suppose that packet $x_0$ has the minimum weighted sum of generation time[2], i.e.,

$$x_0 = \min\{x \in Q_{\mathcal{I},u} | \alpha_i u_i + \alpha_j v_j\}.$$

Denote the corresponding generation time vector of $x_0$ as $(u_{i,0}, v_{j,0})$. From Lemma 2, we define the AoI of $Q_{\mathcal{I},l}$ as

$$w_{\mathcal{I},u}(k) = (w_{\mathcal{I},u}^{(i)}(k), w_{\mathcal{I},u}^{(j)}(k)) \tag{18}$$

where

$$w_{\mathcal{I},u}^{(i)}(k) = \min\{k - v_{i,0}, h_i(k-1) + 1\}$$
$$w_{\mathcal{I},u}^{(j)}(k) = \min\{k - v_{j,0}, h_j(k-1) + 1\}$$

and $w_{\mathcal{I},u}(0) = (h_i(0), h_j(0))$. Similar with (12), we can define *age-gain of $Q_{\mathcal{I},u}$* as

$$\delta_{\mathcal{I},u}(k) = h_{\mathcal{I}}(k) - w_{\mathcal{I},u}(k) \triangleq (\delta_{\mathcal{I},u}^{(i)}(k), \delta_{\mathcal{I},u}^{(j)}(k)) \tag{19}$$

where $h_{\mathcal{I}}(k) = (h_i(k), h_j(k))$. Define

$$\vec{b}(k) = \Big( \{h_i(k)\}_i, \{w_{i,\mathcal{S}}(k)\}_{i,\mathcal{S}}, \{w_{\mathcal{I},u}(k)\}_u, \{x_i(k)\}_i \Big).$$

The Lyapunov function $L(\vec{h}(k), \vec{Y}(k))$ is defined as

$$L(b(k)) = \sum_{i=1}^{M} \beta_i h_i(k) + \lambda \sum_{i=1}^{M} (x_i^+(k))^2, \tag{20}$$

where $\lambda, \beta_i > 0$ is a positive number. We also define the one-slot Lyapunov Drift as

$$\Theta(k) = \mathbb{E}\Big[ L(b(k+1)) - L((b(k))|b(k) \Big]. \tag{21}$$

**Definition 3.** *In each slot $k$, the ARM policy chooses the action that has the maximum weight as follows: $A(k) = Q_i, \oplus_j Q_{j,\mathcal{S}_j}$ have the same weights as that in Definition 2 and*

| $A(k)$ | Weights |
|---|---|
| $Q_{u,i} \oplus Q_{u,\mathcal{I}}$ | $\beta_u \delta_{u,i}(k)(1 - \epsilon_u) + \lambda(1 - \epsilon_u) f_u(k)$ |
| | $+\beta_i \delta_{\mathcal{I},u}^{(i)}(k)(1 - \epsilon_i) + \lambda(1 - \epsilon_i) f_i(k)$ |
| $Q_{u,\mathcal{I}} \oplus Q_{\mathcal{I},u}$ | $\beta_u \delta_{u,\mathcal{I}}(k)(1 - \epsilon_u)$ |
| | $+\sum_{\mathcal{I}:i \in \mathcal{I}} \beta_i \delta_{\mathcal{I},u}^{(i)}(k)(1 - \epsilon_i)$ |
| | $+\lambda \sum_{i \in [3]} (1 - \epsilon_i) f_i(k)$ |

*where $f_i(k)$ is defined in (16).*

[2]Packet $x_0$ is not always the latest packet in $Q_{\mathcal{I},u}$.

**Theorem 3.** *The ARM policy defined in Definition 3 minimizes the one-slot Lyapunov Drift in each slot.*

*Proof.* The proof of Theorem 3 is given in Appendix F. □

Let $C^{coded}$ be the set of all tuples $\underline{q}$ for which $x_i^+(k)$ is strongly stabilized using the considered coding actions. Upper bounding $J(\underline{q})$ for the ARM policy with a specific choice of $\beta$,

**Theorem 4.** *If $(q_1, q_2, \cdots, q_M) \in C^{coded}$, then the strong stability of the process $x_i^+(k)$ satisfies. EAoI has an upper bound (17).*

*Proof.* The proof is similar to that of Theorem 2. □

## V. LOWER BOUND

In prior works [9], [20], [31], lower bounds were found on AoI as a function of the communication rate. Similar to [31, Section III], we derive two lower bounds on the achievable age. The first lower bound is derived by assuming that there is always a fresh packet to be delivered. The second one assumes that all packets are delivered instantaneously upon arrivals.

**Theorem 5.** *For any policy $\pi$ with communication rate $r_i^\pi$, we have the following lower bounds on $J^\pi(\underline{q})$ in (8):*

$$J^\pi(\underline{q}) \geq \frac{M}{2\sum_{i=1}^{M} r_i^\pi / \alpha_i} + \sum_{i=1}^{M} \frac{\alpha_i}{2M} \tag{22}$$

$$J^\pi(\underline{q}) \geq \frac{1}{M} \sum_{i=1}^{M} \frac{\alpha_i}{\theta_i} \tag{23}$$

*Proof.* The proof of Theorem 5 is given in Appendix G. □

**Corollary 1.** *For symmetric networks with independent erasure events, the lower bound in (22) leads to*

$$J^\pi(\underline{q}) \geq \frac{M}{2\epsilon(M) \sum_{i=1}^{M} 1/\alpha_i} + \sum_{i=1}^{M} \frac{\alpha_i}{2M}.$$

*where $\epsilon(M) = \sum_{j=1}^{M} 1/(1 - \epsilon^j)$.*

*Proof.* The proof of Corollary 1 is given in Appendix H. □

## VI. NUMERICAL RESULTS AND DISCUSSION

Finally, we seek to answer the questions that we raised in Section I through simulations. We assume a symmetric networks with $\epsilon_i = \epsilon$, $q_i = q$, and $\theta_i = \theta$ for $i \in [M]$.
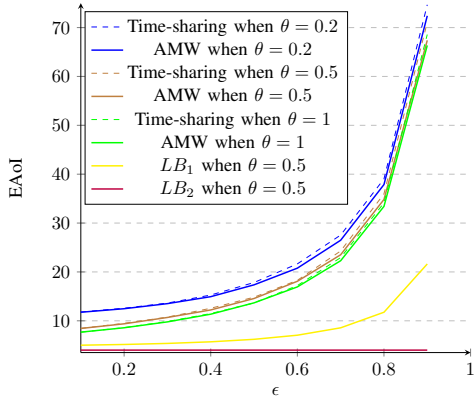
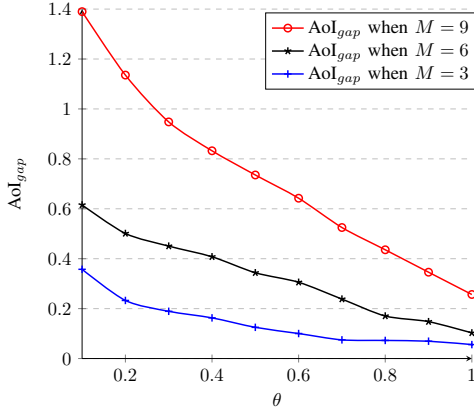Fig. 3: EAoI as a function of $\epsilon$ and $\theta$ when $M = 6$, the upper bound, and the lower bounds.



Fig. 4: $\text{AoI}_{gap}$ v.s $\theta$ when $\epsilon = 0.6$.
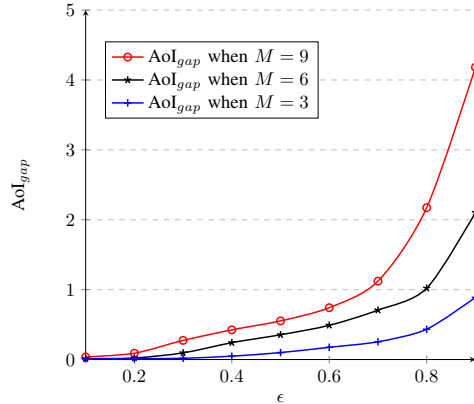


Fig. 5: $\text{AoI}_{gap}$ as a function of $\epsilon$ when $\theta = 0.2$.

### A. Benefits of Coding

We first consider the benefits of coding. The ARM policy and the Time-sharing policy are compared in Figure 3 - Figure 5. To eliminate the impact of rate, we consider the case defined in Remark 1, i.e., the buffer size of every queue is 1 and the stability region is $\{\theta \le 1\}$. We have set $\lambda = 0$ and $\beta_i = \min\{i, \max\{0, i - 3\}\}$. Figure 3 plots the EAoI for $M = 6$

users under the ARM and time-sharing policies, and againts the lower bound in (22). We observe that coding is indeed beneficial when the erasure probability $\epsilon$ is relatively large ($\ge 0.6$) and/or the arrival rate $\theta$ is relatively small ($\le 0.5$). When $\theta$ is fixed, EAoI increases with $\epsilon$.

Next, we define $\text{AoI}_{gap}$ as the gap between the EAoI under the ARM and time-sharing policies. The relationship between $\text{AoI}_{gap}$ and $\theta$ (resp. $\epsilon$) is provided in Figure 4 (resp. Figure 5). In Figure 4, we set $\epsilon = 0.6$. We observethat $\text{AoI}_{gap}$ (the benfit of coding) decreases with the arrival rate $\theta$. This is because the (expected) number of newly incoming packets increases with $\theta$ and the availability of fresh uncoded packets weakens the impact of coding actions.

In Figure 5, we set $\theta = 0.2$. $\text{AoI}_{gap}$ increases with $\epsilon$. This is because erased packets can be cached and provide more coding opportunities. $\text{AoI}_{gap}$ increases slowly when $\epsilon$ is small, and sharply when $\epsilon$ is large. In addition, from Figure 4 and Figure 5, the benefits of coding increase with $M$.

### B. Impact of Maximal Clique Size

The impact of maximal clique size is captured in Figure 6. Let the buffer size of all (virtual) queues be 1 and set $M = 6$, $\lambda = 0$, $\epsilon = 0.6$ and $\beta_i = \min\{i, \max\{0, i - 3\}\}$. ARM in Definition 2 with maximal clique sizes $= 2, 3, 4$ are compared in Figure 6. From Figure 6, we can see that the ARM policy with maximal clique size of 2 is a good approximation. This is useful as it significantly reduces the number of coding actions.
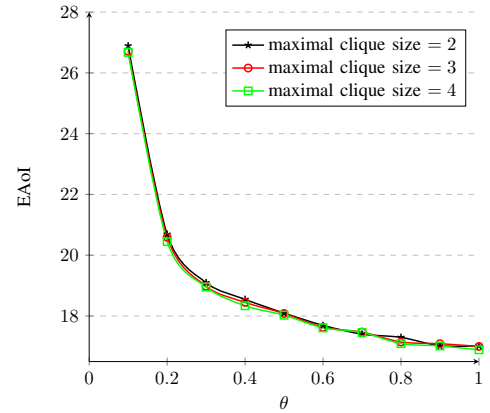


Fig. 6: EAoI under ARM policies under the.

### C. Tradeoff between Age and Rate

We finally investigate the tradeoff between the AoI and rate. Set $M = 3$. The maximum sum-rate achievable with uncoded caching is around $0.44$, and the channel capacity is around $0.46$. Setting $\beta_i = 3$ and $\epsilon = 0.6$ in the ARM policy, we first investigate the relationship between $q$ and EAoI (the red star curve in Figure 7). Now set $\theta = 0.14$, $\lambda = 10$, $q \in [0, 0.1368]$. EAoI increases with $q$ implying that if the minimum required

throughput becomes larger, the system has to sacrifice EAoI to satisfy the rate constraints. Next, the relationship between $\lambda$ and EAoI is investigated (the black circle curve in Figure 7). Let $\theta = 0.14$, $q = 0.1368$, $\lambda \in [0, 10]$. EAoI increases with $\lambda$. In other words, if the rate constraints become more important, then EAoI increases.

Finally, in Figure 8, the EAoI is plotted as a function of the communication rate under the time-sharing policy as well as the ARM policy with uncoded and coded caching. This plot is obtained by setting $\beta_i = \min\{i, \max\{0, i-3\}\}$, $\theta = q$, and $\lambda = 1$. We observe that EAoI decreases as rate increases. The three policies have similar performances up to the rate they support. It appears that ARM with coded caching outperforms for rates close to the boundary of the capacity region.
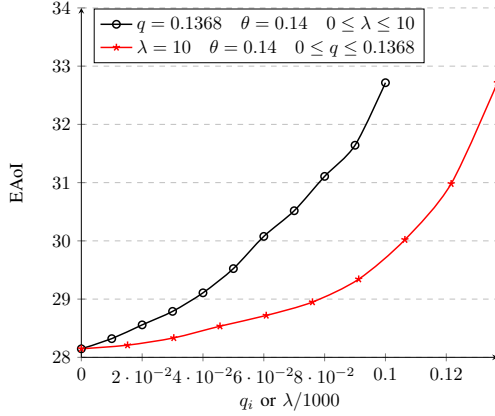


Fig. 7: EAoI vs. $q$ and $\lambda$ when $M = 3$, $\epsilon = 0.6$
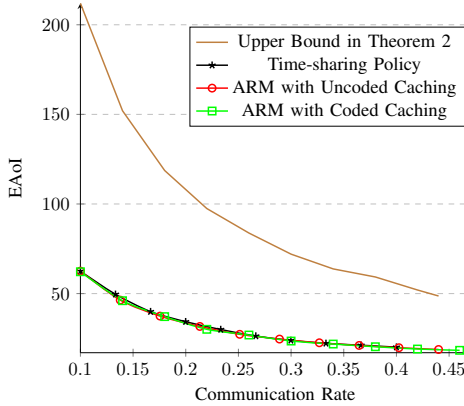


Fig. 8: EAoI v.s rate (when $M = 3$) under time-sharing, ARM with uncoded and coded caching polices when $\epsilon = 0.6$.

## VII. CONCLUSION

We investigated the benefit of coding in memoryless broadcast channel with $M$ users. A scheduling framework is proposed in which coding actions, as opposed to users, are scheduled to attain desired tradeoffs between rate and age of information (AoI). Two general lower bounds and the upper bound for the proposed MW policies are obtained. Simulation showed

that (i) coding is beneficial, and the benefits increase with the number of users; (ii) The tradeoff between rate and AoI exists, EAoI increases with the target rate constraint; (iii) A good approximation of ARM based on maximum clique size was proposed.

## REFERENCES

[1] S. Kaul, M. Gruteser, V. Rai and J. Kenny, "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. Gruteset, "Real-time status: How often should one update?" in *2012 Proceedings IEEE INFOCOM*, 2012, pp. 2731 – 2735.

[3] ——, "Status updates through queues," in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, 2012, pp. 1 – 6.

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

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

[6] 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, 2016.

[7] R. Yates and S. Kaul, "The Age of Information: Real-Time Status Updating by Multiple Sources," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1807 – 1827, 2019.

[8] I. Kadota and E. Modiano, "Minimizing the age of information in wireless networks with stochastic arrivals," *IEEE Transactions on Mobile Computing*, vol. Early Access, 2019.

[9] I. Kadota, A. Sinha, E. Modiano, "Scheduling Algorithms for Optimizing Age of Information in Wireless Networks with Throughput Constraints," *IEEE/ACM Transactions on Networking*, vol. 27, no. 4, pp. 1359 – 1372, 2019.

[10] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling Policies for Minimizing Age of Information in Broadcast Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 26, pp. 2637 – 2650, 2018.

[11] I. Kadota, A. Sinha, E. Modiano, "Optimizing Age of Information in Wireless Networks with Throughput Constraints," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 1844 – 1852.

[12] 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)*, 2018, pp. 2564 – 2568.

[13] A. Kosta, N. Pappas and V. Angelakis, "Age of information: A new concept, metric, and tool," *Foundations and Trends in Networking*, vol. 12, no. 3, 2017.

[14] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano and S. Ulukus, "Age of information: An introduction and survey," arXiv: 2007.08564, 2020.

[15] K. Chen and L. Huang, "Age-of-Information in the Presence of Error," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016.

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

[17] 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)*, 2017, pp. 1 – 6.

[18] E. Najm, E. Telater and R. Nasser, "Optimal Age over Erasure Channels," arXiv: 1901.01573, 2019.

[19] 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)*, 2018, pp. 1 – 5.

[20] X. Chen and S. Saeedi Bidokhti, "Benefits of Coding on Age of Information in Broadcast Networks," in *2019 IEEE Information Theory Workshop (ITW)*, 2019.

[21] S. Feng and J. Yang, "Procoding and Scheduling for AoI Minimization in MIMO Broadcast Channels," arXiv: 2009.00171, 2020.

[22] ——, "Adaptive Coding for Information Freshness in a Two-user Broadcast Erasure Channel," arXiv: 1905.00521, 2019.

[23] M. Costa and Y. Sagduyu, "Age of information with network coding," *Ad Hoc Networks*, vol. 86, pp. 15 – 22, 2019.

[24] M. Gatzianas, L. Georgiadis, L. Tassiulas, "Multiuser Broadcast Erasure Channel with Feedback—Capacity and Algorithms," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5779 – 5804, 2013.

[25] S. Saeedi Bidokhti, M. Gatzianas and C. Fragouli, "A class of feedback-based coding algorithms for broadcast erasure channels with degraded message sets," http://infoscience.epfl.ch/record/175823, 2012.

[26] 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, 2018.

[27] S. C. Lin, I. H. Wang and A. Vahid, "No feedback, no problem: Capacity of erasure broadcast channels with single-user delayed csi," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019.

[28] S. Athanasiadou, M. Gatzianas, L. Georgiadis, and L. Tassiulas, "XOR-based coding for the 3-user broadcast erasure channel with feedback," in *2012 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 2012.

[29] J. M. Moon and L. Moser, "On cliques in graphs," *Israel Journal of Mathematics*, vol. 3, pp. 23 – 28, 1965.

[30] M. J. Neely, *Stochastic Network Optimization With Application to Communication and Queueing Systems*. Morgan and Claypool Publishers, 2010.

[31] X. Chen, K. Gatsis, H. Hassani, and S. Saeedi Bidokhti, "Age of Information in Random Access Channels," arXiv:1912.01473, 2019.

## APPENDIX A
### TIME-SHARING POLICIES

We devise deterministic policies without coding using techniques from Lyapunov Optimization. Denote the EAoI for Max-Weight policies (in the long run) as $\mathbb{E}[J]$. Denote

$$S(k) = \Big( \{h_i(k)\}_i, \{w_{i,\varnothing}(k)\}_i, \{x_i(k)\}_i \Big).$$

Define the Lyapunov function

$$L\big(S(k)\big) = \sum_{i=1}^{M} \beta_i h_i(k) + \lambda \sum_{i=1}^{M} \big(x_i^+(k)\big)^2, \quad (24)$$

where $\beta_i, \lambda > 0$, and the one-slot Lyapunov Drift is defined as

$$\Theta(k) = \mathbb{E}\Big[ L\big(S(k+1)\big) - L\big(S(k)\big) | S(k) \Big]. \quad (25)$$

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

**Definition 4.** *Let $f_i(k)$ be defined in (16). In each slot $k$, the MW policy chooses the action that has the maximum weight as shown in following Table:*

| $A(k)$ | Weights |
|---|---|
| $Q_{i,\varnothing}$ | $\beta_i \delta_{i,\varnothing}(k)(1-\epsilon_i) + \lambda(1-\epsilon_i)f_i(k)$ |

**Theorem 6.** *The MW policy defined in Definition 4 minimizes the one-slot Lyapunov Drift in each slot.*

*Proof.* Based on the definition of $t_{i,\varnothing}(k)$, we have $t_{i,\varnothing}(k) \in \{0,1\}$ and

$$\sum_{i=1}^{M} t_{i,\varnothing}(k) = 1.$$

Let $\Theta(k) = \Theta_1(k) + \lambda \Theta_2(k)$, where

$$\Theta_1(k) = \mathbb{E}\Big[ \sum_{i=1}^{M} \beta_i h_i(k+1) - \sum_{i=1}^{M} \beta_i h_i(k) | S(k) \Big]$$

$$\Theta_2(k) = \mathbb{E}\Big[ \sum_{i=1}^{M} \big(x_i^+(k+1)\big)^2 - \sum_{i=1}^{M} \big(x_i^+(k)\big)^2 | S(k) \Big].$$

We first consider $\Theta_1(k)$. From (11),

$$h_i(k+1) = t_{i,\varnothing}(k)d_i(k)\big(w_{i,\varnothing}(k) + 1\big) + \big(1 - t_{i,\varnothing}(k)d_i(k)\big)\big(h_i(k) + 1\big). \quad (26)$$

Similar with (30), we have $d_i(k) = H_i t_{i,\varnothing}(k)$. Using (32), we can re-write the Lyapunov Drift as follows:

$$\Theta_1(k) = \sum_{i=1}^{M} \beta_i \mathbb{E}\Big[ H_i t_{i,\varnothing}(k)\big(w_{i,\varnothing}(k) - h_i(k)\big) + 1 | S(k) \Big]$$

$$= -\sum_{i=1}^{M} \beta_i \mathbb{E}\big[ H_i t_{i,\varnothing}(k)\delta_{i,\varnothing}(k) | S(k) \big] + \sum \beta_i$$

$$= -\sum_{i=1}^{M} \beta_i(1-\epsilon_i)\mathbb{E}\big[ t_{i,\varnothing}(k) | S(k) \big]\delta_{i,\varnothing}(k) + \sum \beta_i.$$

Therefore,

$$\Theta_1(\vec{h}(k)) = \sum_{i=1}^{M} \beta_i - \sum_{i=1}^{M} 1_{\{A(k)=Q_{i,\varnothing}\}} \beta_i \delta_{i,\varnothing}(k)(1-\epsilon_i). \quad (27)$$

Then, we consider $\Theta_2(k)$. Given $S(k)$, it is sufficient to consider

$$\tilde{\Theta}_2(k) = \mathbb{E}\Big[ \sum_{i=1}^{M} \big(x_i^+(k+1)\big)^2 | S(k) \Big].$$

Note that

$$\mathbb{E}[(x_i^+(k+1))^2|S(k)] = \mathbb{E}\Big[\big((x_i(k)+q_i-d_i(k))^+\big)^2|S(k)\Big]$$

$$= \begin{cases} \big((x_i(k)+q_i)^+\big)^2 & \text{if } t_{i,\varnothing}(k)=0 \\ \big((x_i(k)+q_i-1)^+\big)^2(1-\epsilon_i) \\ \quad + \big((x_i(k)+q_i)^+\big)^2\epsilon_i & \text{if } t_{i,\varnothing}(k)=1 \end{cases}.$$
(28)

If $A(k) = Q_{i,\varnothing}$, then

$$\tilde{\Theta}_2(k) = (1-\epsilon_i)\Big(\big((x_i(k)+q_i-1)^+\big)^2 - \big((x_i(k)+q_i)^+\big)^2\Big)$$
$$+ \sum_{j=1}^M \big((x_j(k)+q_j)^+\big)^2.$$

From (27) and (35), minimizing $\Theta(k)$ is equivalent to maximize

$$\sum_{i=1}^M \mathbf{1}_{\{A(k)=Q_{i,\varnothing}\}}\Big(\beta_i\delta_{i,\varnothing}(k)(1-\epsilon_i) + \lambda(1-\epsilon_i)f_i(k)\Big).$$

$\square$

## APPENDIX B
## PROOF OF LEMMA 1

For any $\mathcal{S} \subset [M]\backslash i$, we first consider $\mathcal{S} = \varnothing$ (Step 1), then consider $\mathcal{S} \neq \varnothing$ (Step 2).

**Step 1**. Consider $\mathcal{S} = \varnothing$. If $p_j$ is delivered to user $i$ in time slot $k$, then from Definition 1, the AoI of user $i$ is $\min\{k-k_j, h_i(k-1)+1\}$. Note that $k-k_1 > k-k_2$, hence

$$\min\{k-k_1, h_i(k-1)+1\} \geq \min\{k-k_2, h_i(k-1)+1\},$$

which implies $p_1$ provides a larger AoI reduction (for user $i$) than that of $p_2$. Then, transmitting $p_2$ can not be worse than transmitting $p_1$ in terms of AoI. If there is no delivery, then transmitting $p_2$ is again not worse that transmitting $p_1$.

**Step 2**. Consider $\mathcal{S} \neq \varnothing$. Without loss of generality, suppose that $i, j_1, j_2, \cdots, j_l$ form a maximal clique. Denote the corresponding coded packet as $c_\tau$, $\tau \in \{1,2\}$, respectively. Note that $c_1$ and $c_2$ provide the same AoI reduction to users $j_1, j_2, \cdots, j_l$. We only consider the case where $c_1, c_2$ can be fully decoded at user $i$; otherwise the AoI of user $i$ does not change after recovering $c_1, c_2$.

**(1)** If the coded packet is delivered to user $i$, then from Definition 1, the AoI of user $i$ is $\min\{k-k_j, h_i(k-1)+1\}$ when recovering $c_j$. Note that $k-k_1 > k-k_2$, so $\min\{k-k_1, h_i(k-1)+1\} \geq \min\{k-k_2, h_i(k-1)+1\}$, $c_2$ provides smaller AoI for user $i$. Then, encoding $p_2$ can not be worse than encoding $p_1$ in terms of AoI.

**(2)** In the future time slot, suppose that $p_2$ has been recovered, and $i, j'_1, j'_2, \cdots, j'_m$ forms another maximal clique, and the corresponding coded packet is denoted by $c'$. If $c'$ is delivered to user $i$, by Definition 1, $c'$ can not provide AoI reduction for user $i$. Therefore, the coded packet encoded by uncoded packets from $j'_1, j'_2, \cdots, j'_m$ provides the same AoI reduction (as $c'$) for users $j'_1, j'_2, \cdots, j'_m$.

Thus, from **(1)** and **(2)**, encoding $p_2$ can not be worse than encoding $p_1$.

## APPENDIX C
## PROOF OF THEOREM 1

Based on the definition of $t_{i,\mathcal{S}_i}(k)$, we have $t_{i,\mathcal{S}_i}(k) \in \{0,1\}$, and if $A(k) = Q_{i,\mathcal{S}_i} \oplus_{j \in [l]} Q_{\tau_j,\mathcal{S}_{\tau_j}}$, then $t_{i,\mathcal{S}_i}(k) = 1$ and $t_{i,\tilde{\mathcal{S}}_i}(k) = 0$ where $\tilde{\mathcal{S}}_i \neq \mathcal{S}_i$. Then,

$$\sum_{\mathcal{S}_i \subset [M]\backslash i} t_{i,\mathcal{S}_i}(k) \leq 1.$$
(29)

Let $\Theta(k) = \Theta_1(k) + \lambda\Theta_2(k)$, where

$$\Theta_1(k) = \mathbb{E}\Big[\sum_{i=1}^M \beta_i h_i(k+1) - \sum_{i=1}^M \beta_i h_i(k)|S(k)\Big]$$

$$\Theta_2(k) = \mathbb{E}\Big[\sum_{i=1}^M \big(x_i^+(k+1)\big)^2 - \sum_{i=1}^M \big(x_i^+(k)\big)^2|S(k)\Big].$$

The relationship between $d_i(k)$ and $t_{i,\mathcal{S}}$ is

$$d_i^\pi(k) = E_i \sum_{\mathcal{S} \subset [M]\backslash i} t_{i,\mathcal{S}}^\pi(k),$$
(30)

where $E_i$ is a binary random variable independent of $t_{i,\mathcal{S}}^\pi(k)$ for all $\mathcal{S}$ and $k$, $\Pr(E_i=1) = 1-\epsilon_i$ and $\Pr(E_i=0) = \epsilon_i$. Thus,

$$\mathbb{E}[d_i^\pi(k)] = (1-\epsilon_i) \sum_{\mathcal{S} \subset [M]\backslash i} \mathbb{E}[t_{i,\mathcal{S}}^\pi(k)].$$
(31)

We first consider $\Theta_1(k)$. Using (30), (11) and (29),

$$h_i(k+1) = \sum_{\mathcal{S} \subset [M]\backslash i} t_{i,\mathcal{S}}(k)E_i\big(w_{i,\mathcal{S}}(k)+1\big)$$
$$+ \big(1 - \sum_{\mathcal{S} \subset [M]\backslash i} t_{i,\mathcal{S}}(k)E_i\big)\big(h_i(k)+1\big).$$
(32)

Recall that $t_{i,\mathcal{S}_i}(k) \in \{0,1\}$ and $\sum_{\mathcal{S}_i \subset [M]\backslash i} t_{i,\mathcal{S}_i}(k) \leq 1$, so $t_{i,\mathcal{S}_i}^2(k) = t_{i,\mathcal{S}_i}(k)$ and $t_{i,\mathcal{S}_i}(k)t_{i,\tilde{\mathcal{S}}_i}(k) = 0$ if $\mathcal{S}_i \neq \tilde{\mathcal{S}}_i$. Using (32), we can re-write the Lyapunov Drift as follows:

$$\Theta_1(k) = \sum_{i=1}^M \beta_i \mathbb{E}\Bigg[\sum_{\mathcal{S}_i \subset [M]\backslash i} E_i t_{i,\mathcal{S}_i}(k)\big(w_{i,\mathcal{S}_i}(k) - h_i(k)\big) + 1|S(k)\Bigg]$$

$$= -\sum_{i=1}^M \beta_i \mathbb{E}\Big[\sum_{\mathcal{S}_i \subset [M]\backslash i} E_i t_{i,\mathcal{S}_i}(k)\delta_{i,\mathcal{S}_i}(k)|S(k)\Big] + \sum \beta_i$$

$$= -\sum_{i=1}^M \beta_i(1-\epsilon_i) \sum_{\mathcal{S}_i \subset [M]\backslash i} \mathbb{E}\big[t_{i,\mathcal{S}_i}(k)|S(k)\big]\delta_{i,\mathcal{S}_i}(k) + \sum \beta_i.$$

Therefore,

$$
\begin{aligned}
\Theta_1(\vec{h}(k)) = &\sum_{i=1}^{M} \beta_i - \sum_{i=1}^{M} 1_{\{A(k)=Q_{i,\varnothing}\}} \beta_i \delta_{i,\varnothing}(k)(1-\epsilon_i) \\
&- \sum_{l=2}^{M} \sum_{\tau_1,\tau_2,\cdots,\tau_l} 1_{\{A(k)=\oplus_{u\in[l]}Q_{\tau_j,\mathcal{S}_{\tau_u}}\}} \\
&\times \sum_{u=1}^{l} \beta_{\tau_u} \delta_{i,\mathcal{S}_{\tau_u}}(k)(1-\epsilon_{\tau_u}).
\end{aligned}
\tag{33}
$$

Then, we consider $\Theta_2(k)$. Given $S(k)$, it is sufficient to consider

$$
\tilde{\Theta}_2(k) = \mathbb{E}\Big[ \sum_{i=1}^{M} \big(x_i^+(k+1)\big)^2 | S(k) \Big].
$$

Note that

$$
\begin{aligned}
&\mathbb{E}[(x_i^+(k+1))^2 | S(k)] = \mathbb{E}\Big[\big((x_i(k)+q_i-d_i(k))^+\big)^2 | S(k)\Big] \\
&= \begin{cases} \big((x_i(k)+q_i)^+\big)^2 & \text{if } \sum_{\mathcal{S}_i\subset[M]\backslash i} t_{i,\mathcal{S}_i}(k)=0 \\ \big((x_i(k)+q_i-1)^+\big)^2(1-\epsilon_i) \\ \quad + \big((x_i(k)+q_i)^+\big)^2 \epsilon_i & \text{if } \sum_{\mathcal{S}_i\subset[M]\backslash i} t_{i,\mathcal{S}_i}(k)=1 \end{cases}.
\end{aligned}
\tag{34}
$$

Then, we consider the following two cases:

**Case 1**. If $A(k)=Q_{i,\varnothing}$, then

$$
\begin{aligned}
\tilde{\Theta}_2(k) =&(1-\epsilon_i)\Big( \big((x_i(k)+q_i-1)^+\big)^2 - \big((x_i(k)+q_i)^+\big)^2 \Big) \\
&+ \sum_{j=1}^{M} \big((x_j(k)+q_j)^+\big)^2.
\end{aligned}
$$

**Case 2**. If $A(k)=\oplus_{j\in[l]}Q_{\tau_j,\mathcal{S}_{\tau_j}}$, then

$$
\begin{aligned}
\tilde{\Theta}_2(k) =&\sum_{u}^{l}(1-\epsilon_{\tau_u})\Big( \big((x_{\tau_u}(k)+q_{\tau_u}-1)^+\big)^2 - \big((x_{\tau_u}(k)+q_{\tau_u})^+\big)^2 \Big) \\
&+ \sum_{j=1}^{M} \big((x_j(k)+q_j)^+\big)^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\tilde{\Theta}_2(k) =&\sum_{j=1}^{M} \big((x_j(k)+q_j)^+\big)^2 - \sum_{j} 1_{\{A(k)=Q_{j,\varnothing}\}} f_j(k) \\
&- \sum_{l=2}^{M} \sum_{\tau_1,\tau_2,\cdots,\tau_l} 1_{\{A(k)=\oplus_{u\in[l]}Q_{\tau_u,\mathcal{S}_{\tau_u}}\}} \sum_{u=1}^{l} f_{\tau_u}(k)
\end{aligned}
\tag{35}
$$

From (33) and (35), minimizing $\Theta(k)$ is equivalent to maximize

$$
\begin{aligned}
&\sum_{i=1}^{M} 1_{\{A(k)=Q_{i,\varnothing}\}}(1-\epsilon_i)\Big( \beta_i\delta_{i,\varnothing}(k) + \lambda f_i(k) \Big) \\
&+ \sum_{l=2}^{M} \sum_{\tau_u,u\in[l]} 1_{\{A(k)=\oplus_{u\in[l]}Q_{\tau_u,\mathcal{S}_{\tau_u}}\}} \\
&\times \sum_{u=1}^{l}(1-\epsilon_{\tau_u})\Big( \beta_{\tau_u}\delta_{i,\mathcal{S}_{\tau_u}}(k) + \lambda f_{\tau_u}(k) \Big).
\end{aligned}
$$

## APPENDIX D
## PROOF OF THEOREM 2.

*In the proof, we consider $3$ users. The process can be easily generalized to $M$ users.*

We first consider a stationary randomized policy,

$$
\begin{aligned}
\Pr(A(k)=Q_{i,\varnothing}) &= \mu_{i,\varnothing} \\
\Pr\{A(k)=\oplus_{j\in[l]}Q_{\tau_j,\mathcal{S}_{\tau_j}}\} &= \mu_{(\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l})}.
\end{aligned}
$$

Then,

$$
\sum_{j=1}^{M} \mu_{i,\varnothing} + \sum_{l=2}^{M} \sum_{\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l}} \mu_{(\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l})} = 1.
\tag{36}
$$

As defined before, let $\sigma(\mathcal{I})$ be the probability such that a packet is erased by users in $\mathcal{I}$, and is cached by users in $[M]\backslash\mathcal{I}$. In particular, if the system is independent and symmetric ($\epsilon_1=\cdots=\epsilon_M=\epsilon$), then $\sigma(\mathcal{I})=(1-\epsilon)^{M-|\mathcal{I}|}\epsilon^{\mathcal{I}}$.

From the proof of Theorem 1,

$$
\begin{aligned}
\Theta_1(k) = &\sum_{i=1}^{M} \beta_i - \sum_{i=1}^{M} \mu_{i,\varnothing}\beta_i\delta_{i,\varnothing}(k)(1-\epsilon_i) \\
&- \sum_{l=2}^{M} \sum_{\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l}} \mu_{(\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l})} \\
&\times \sum_{j=1}^{l} \beta_{\tau_j}\delta_{i,\mathcal{S}_{\tau_j}}(k)(1-\epsilon_{\tau_j}).
\end{aligned}
$$

From definitions of $w_{i,\mathcal{S}}(k)$, we have $w_{i,\mathcal{S}}(k)\le h_i(k)$ for all $i$ and $\mathcal{S}$. Then, $\mathbb{E}[\delta_{i,\mathcal{S}}(k)]\le 0$. Therefore, remove the terms related to $\mathbb{E}[\delta_{i,\mathcal{S}}(k)]$ when $\mathcal{S}\neq\varnothing$,

$$
\Theta_1(k) \le \sum_{i=1}^{M} \beta_i - \sum_{i=1}^{M} \mu_{i,\varnothing}\beta_i\delta_{i,\varnothing}(k)(1-\epsilon_i).
$$

From [9, Eqn.(60) - (62)],

$$
\begin{aligned}
\big(x_i^+(k+1)\big)^2 - \big(x_i^+(k)\big)^2 &\le -2x_i^+(k)\big(d_i(k)-q_i\big)+1 \\
&\le -2x_i^+(k)\big(1-q_i\big)+1,
\end{aligned}
$$

thus

$$
\Theta_2(k) \le \sum_{i=1}^{M} -2x_i^+(k)\big(1-q_i\big) + M.
$$

Therefore,

$$\sum_{k=1}^{K} \mathbb{E}[\Theta(k)] \leq \sum_{k=1}^{K} \sum_{i=1}^{M} \beta_i(1-\epsilon_i)\mu_{i,\varnothing}\mathbb{E}[w_{i,\varnothing}(k) - h_i(k)]$$

$$- \sum_{i=1}^{M} 2\lambda\mathbb{E}[x_i^+(k)](1-q_i) + K\sum_{i=1}^{M}\beta_i + KM\lambda.$$

By algebra, we have

$$\sum_{k=1}^{K}\sum_{i=1}^{M} \beta_i(1-\epsilon_i)\mu_{i,\varnothing}\mathbb{E}[h_i(k)] + \sum_{k=1}^{K}\sum_{i=1}^{M} 2\lambda\mathbb{E}[x_i^+(k)](1-q_i)$$

$$\leq -\sum_{k=1}^{K}\mathbb{E}[\Theta(k)] + \sum_{k=1}^{K}\sum_{i=1}^{M}\beta_i(1-\epsilon_i)\mu_{i,\varnothing}\mathbb{E}[w_{i,\varnothing}(k)] \quad (37)$$

$$+ K\sum_{i=1}^{M}\beta_i + KM\lambda.$$

From (9), $w_{i,\varnothing}(k) \leq W_i(k)$, where $W_i(k)$ is defined in

$$W_i(k+1) = \begin{cases} 0 & G_i(k) = 1 \\ W_i(k) + 1 & G_i(k) = 0 \end{cases}. \quad (38)$$

Substituting (38) into (37),

$$L_1 + L_2 \leq L_3 + \sum_{k=1}^{K}\sum_{i=1}^{M}\beta_i(1-\epsilon_i)\mu_i\mathbb{E}[W_i(k)] + C. \quad (39)$$

where

$$L_1 = \sum_{k=1}^{K}\sum_{i=1}^{M}\beta_i(1-\epsilon_i)\mu_{i,\varnothing}\mathbb{E}[h_i(k)]$$

$$L_2 = \sum_{k=1}^{K}\sum_{i=1}^{M} 2\lambda\mathbb{E}[x_i^+(k)](1-q_i)$$

$$L_3 = -\sum_{k=1}^{K}\Theta(k), \quad C = K\sum_{i=1}^{M}\beta_i + KM\lambda.$$

Let $K \to \infty$,

$$\lim_{K\to\infty}\frac{L_3}{K} \leq \lim_{K\to\infty}\frac{L(S(1))}{K} = 0.$$

Thus, dividing by $KM$ on both sides of (39), we have

$$\lim_{K\to\infty}\frac{L_1}{K} + \lim_{K\to\infty}\frac{L_2}{K}$$
$$\leq \frac{\sum_{i=1}^{M}\beta_i(1-\epsilon_i)\mu_{i,\varnothing}\sum_{k=1}^{K}\mathbb{E}[W_i(k)]}{KM} + \frac{\sum_{i=1}^{M}\beta_i}{M} + \lambda. \quad (40)$$

Note that $W_i(k)$ is a geometric random variable with parameter $\theta_i$, we have

$$\lim_{K\to\infty}\frac{L_1}{K} + \lim_{K\to\infty}\frac{L_2}{K}$$
$$\leq \frac{1}{M}\sum_{i=1}^{M}\beta_i\left(\frac{(1-\epsilon_i)\mu_{i,\varnothing}}{\theta_i} + 1\right) + \lambda. \quad (41)$$

Since $L_1 > 0$, so (41) is reduced to

$$\sum_{i=1}^{M} 2\lambda(1-q_i)\lim_{K\to\infty}\frac{\sum_{k=1}^{K}\mathbb{E}[x_i^+(k)]}{K}$$
$$\leq \frac{1}{M}\sum_{i=1}^{M}\beta_i\left(\frac{(1-\epsilon_i)\mu_{i,\varnothing}}{\theta_i} + 1\right) + \lambda. \quad (42)$$
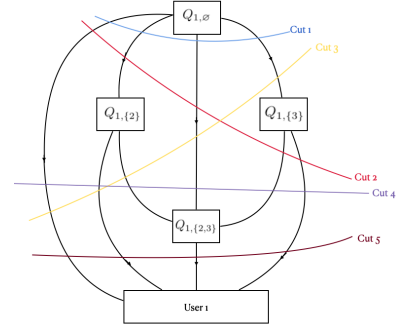


Fig. 9: The flow of user 1 in the virtual network.

Then,

$$\lim_{K\to\infty}\frac{\sum_{k=1}^{K}\mathbb{E}[x_i^+(k)]}{K} < \infty, \quad i = 1, 2, \cdots, M.$$

In fact, strong stability of the process $x_i^+(k)$, i.e.,

$$\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[x_i^+(k)] < \infty,$$

is sufficient to establish that the minimum rate constraint, $r_i^\pi \geq q_i$, is satisfied [9], [30, Theorem 2.8].

Since $L_2 > 0$, so

$$\lim_{K\to\infty}\frac{L_1}{MK} \leq \frac{1}{M}\sum_{i=1}^{M}\beta_i\left(\frac{(1-\epsilon_i)\mu_{i,\varnothing}}{\theta_i} + 1\right) + \lambda. \quad (43)$$

Now, we consider the probabilities of actions, $\{\mu_{i,\varnothing}\}$ and $\{\mu_{\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l}}\}$. Consider $M = 3$. Note that $(q_1, q_2, q_3)$ gives the minimum requirements of rate, then for each $i$. In the graph of virtual network, every user has 5 cuts, see Figure 9. Take user 1 as an example, Cut 1 implies that

$$\mu_{1,\varnothing}(1-\epsilon_1) + \sum_{j\in\{2,3\}}\mu_{1,\varnothing}\sigma([3]\backslash j) + \mu_{1,\varnothing}\sigma(\{1\}) \geq q_1, \quad (44)$$

Cut 2 implies that

$$\mu_{1,\varnothing}(1-\epsilon_1) + \mu_{1,\varnothing}\sigma(\{1,3\}) + \mu_{1,\varnothing}\sigma(\{1\})$$
$$+ \left(\mu_{1,\{3\},3,\{1\}} + \mu_{1,\{3\},3,\{1,2\}}\right)(1-\epsilon_1+\sigma(\{1\})) \geq q_1, \quad (45)$$

Cut 3 implies that

$$\mu_{1,\varnothing}(1-\epsilon_1) + \mu_{1,\varnothing}\sigma(\{1,2\}) + \mu_{1,\varnothing}\sigma(\{1\})$$
$$+ \left(\mu_{1,\{2\},2,\{1\}} + \mu_{1,\{2\},2,\{1,3\}}\right)(1-\epsilon_1+\sigma(\{1\})) \geq q_1, \quad (46)$$

Cut 4 implies that

$$\mu_{1,\varnothing}(1-\epsilon_1) + \mu_{1,\varnothing}\sigma(\{1\})$$
$$+ \left(\mu_{1,\{3\},3,\{1\}} + \mu_{1,\{3\},3,\{1,2\}}\right)(1-\epsilon_1+\sigma(\{1\})) \quad (47)$$
$$+ \left(\mu_{1,\{2\},2,\{1\}} + \mu_{1,\{2\},2,\{1,3\}}\right)(1-\epsilon_1+\sigma(\{1\})) \geq q_1,$$

Cut 5 implies that

$$\mu_{1,\varnothing}(1-\epsilon_1) + \left(\mu_{1,\{2,3\},2,\{1\}} + \mu_{1,\{2,3\},3,\{1\}}\right.$$
$$\left.+ \mu_{1,\{2,3\},2,\{1,3\}} + \mu_{1,\{2,3\},3,\{1,2\}}\right)(1-\epsilon_1)$$
$$+ \left(\mu_{1,\{3\},3,\{1\}} + \mu_{1,\{3\},3,\{1,2\}}\right)(1-\epsilon_1) \quad (48)$$
$$+ \left(\mu_{1,\{2\},2,\{1\}} + \mu_{1,\{2\},2,\{1,3\}}\right)(1-\epsilon_1) \geq q_1.$$

Similar process for users 2 and 3, we get (49) $\sim$ (52),

$$\mu_{i,\varnothing}(1-\epsilon_i) + \sum_{j\in[3]\setminus i}\mu_{i,\varnothing}\sigma([3]\setminus j) + \mu_{i,\varnothing}\sigma(\{i\}) \geq q_i, \tag{49}$$

$$\begin{aligned}
&\mu_{i,\varnothing}(1-\epsilon_i) + \mu_{i,\varnothing}\sigma([3]\setminus j) + \mu_{i,\varnothing}\sigma(\{i\}) \\
&+ \big(\mu_{i,\{k\},k,\{i\}} + \mu_{i,\{k\},k,[3]\setminus k}\big)(1-\epsilon_i+\sigma(\{i\})) \geq q_i \\
&j,k \neq i,\ j,k \in [3],\ j \neq k,
\end{aligned} \tag{50}$$

$$\begin{aligned}
&\mu_{i,\varnothing}(1-\epsilon_i) + \mu_{i,\varnothing}\sigma(\{i\}) \\
&+ \sum_{j\in[3]\setminus i}\big(\mu_{i,\{j\},j,\{i\}} + \mu_{i,\{j\},j,[3]\setminus j}\big)(1-\epsilon_i+\sigma(\{i\})) \geq q_i,
\end{aligned} \tag{51}$$

$$\begin{aligned}
&\mu_{i,\varnothing}(1-\epsilon_i) + \sum_{j\in[3]\setminus i}\big(\mu_{i,[3]\setminus i,j,\{i\}} + \mu_{i,[3]\setminus i,j,[3]\setminus j}\big)(1-\epsilon_i) \\
&+ \sum_{j\in[3]\setminus i}\big(\mu_{i,\{j\},j,\{i\}} + \mu_{i,\{j\},j,[3]\setminus j}\big)(1-\epsilon_i) \geq q_i.
\end{aligned} \tag{52}$$

Thus, $\{\mu_{i,\varnothing}\}$ and $\{\mu_{\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l}}\}$ satisfying (36), (49) $\sim$ (52) give the inequality (43). Note that $\sum_{i=1}^3 q_i < 1$. So let $\mu_{i,\varnothing} = q_i$, we can choose proper $\{\mu_{\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l}}\}$ such that (36), (49) $\sim$ (52) are satisfied. Substituting $\mu_{i,\varnothing} = q_i$ into (43), we get

$$\begin{aligned}
&\frac{1}{3}\sum_{i=1}^3(1-\epsilon_i)q_i \cdot \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^K \beta_i\mathbb{E}[h_i(k)] \\
&\leq \frac{1}{3}\sum_{i=1}^3 \beta_i\Big(\frac{(1-\epsilon_i)q_i}{\theta_i}+1\Big) + \lambda.
\end{aligned}$$

In particular, set $\beta_i = \frac{\alpha_i}{(1-\epsilon_i)q_i}$, which yields to

$$J(\underline{q}) \leq \frac{1}{3}\sum_{i=1}^3 \Big(\frac{\alpha_i}{\theta_i}+\frac{\alpha_i}{q_i(1-\epsilon_i)}\Big) + \lambda.$$

## APPENDIX E
## PROOF OF LEMMA 2

**Step 1**. We first consider the case where $x_1$ and $x_2$ are transmitted directly. If $x_1$ is delivered to users $1,2$, then from Definition 1, the AoI reduction of the users $1,2$ is

$$\begin{aligned}
&\alpha_1 \min\{k-u(t_1), h_1(k-1)+1\} \\
&+\alpha_2 \min\{k-v(t_2), h_2(k-1)+1\},
\end{aligned}$$

which is larger than the AoI reduction when $x_2$ is delivered,

$$\begin{aligned}
&\alpha_1 \min\{k-u(t_3), h_1(k-1)+1\} \\
&+\alpha_2 \min\{k-v(t_4), h_2(k-1)+1\}.
\end{aligned}$$

This is because $\alpha_1 u(t_1)+\alpha_2 v(t_2) > \alpha_1 u(t_3)+\alpha_2 v(t_4)$. Then, transmitting $x_2$ can not be worse than transmitting $x_1$.

**Step 2**. If the encoder transmits a coded packet which is formed by $x_1 \oplus p$ or $x_2 \oplus p$, where $p \in Q_{3,\mathcal{S}_3}$. Note that $x_1 \oplus p$ and $x_2 \oplus p$ provide the same AoI reduction to user 3. We only consider the case where the two coded packets are fully decoded at users $1,2$.

**(1)** If the coded packet is delivered to users $1,2$, By similar analysis from **Step 1**, $x_2\oplus p$ provides larger AoI reduction for

users $1,2$. Then, encoding $x_2$ can not be worse than encoding $x_1$ in terms of AoI.

**(2)** In the future time slot, suppose that $x_2$ has been recovered, and $x_1$ can be encoded with another packet $p' \in Q_{3,\mathcal{S}_3'}$. If $x_1 \oplus p'$ is delivered to users $1,2$, by Definition 1, $x_1 \oplus p'$ can not provide AoI reduction for users $1,2$. Transmitting $p'$ directly can provide the same AoI reduction for user 3.

Therefore, from **(1)** and **(2)**, encoding $x_2$ can not be worse than encoding $x_1$.

## APPENDIX F
## PROOF OF THEOREM 3

Let $t_{\mathcal{I},u}(k) = 1$ represents the latest packet in $Q_{\mathcal{I},u}$ is encoded and transmitted in time slot $k$, otherwise $t_{\mathcal{I},u}(k) = 0$. So (30) can be extended as

$$d_i(\tau) = E_i\Big(\sum_{\mathcal{S}\subset[M]\setminus i}t_{i,\mathcal{S}}(\tau) + \sum_{\mathcal{I}:i\in\mathcal{I}}t_{\mathcal{I},u}(\tau)\Big). \tag{53}$$

From the definition of $t_{i,\mathcal{S}}(k)$ and $t_{\mathcal{I},u}(k)$, we have

$$t_{i,\mathcal{S}}(k), t_{\mathcal{I},u}(k) \in \{0,1\}$$

and

$$\sum_{\mathcal{S}\subset[M]\setminus i}t_{i,\mathcal{S}}(k) + \sum_{\mathcal{I}:i\in\mathcal{I}}t_{\mathcal{I},u}(k) \leq 1$$

Let $\Theta(k) = \Theta_1(k) + \lambda\Theta_2(k)$, where

$$\begin{aligned}
\Theta_1(k) &= \mathbb{E}\Big[\sum_{i=1}^M \beta_i h_i(k+1) - \sum_{i=1}^M \beta_i h_i(k)|b(k)\Big] \\
\Theta_2(k) &= \mathbb{E}\Big[\sum_{i=1}^M \big(x_i^+(k+1)\big)^2 - \sum_{i=1}^M \big(x_i^+(k)\big)^2|b(k)\Big].
\end{aligned}$$

Similar with the proof of Theorem 1,

$$\begin{aligned}
\Theta_1(k) = &-\sum_{i=1}^M \Big(\alpha_i \sum_{\mathcal{S}}t_{i,\mathcal{S}}(k)\mathbb{E}\big[d_i(k)\big]\delta_{i,\mathcal{S}}(k) \\
&+ \sum_{\mathcal{I}:i\in\mathcal{I}}t_{\mathcal{I},u}(k)\mathbb{E}\big[d_i(k)\big]\delta_{\mathcal{I},u}^{(i)}(k)\Big) + \sum_{i=1}^M \alpha_i.
\end{aligned} \tag{54}$$

Define $\tilde{\Theta}_2(k) = \mathbb{E}\Big[\sum_{i=1}^M \big(x_i^+(k+1)\big)^2|b(k)\Big]$. Then, by similar proof of Theorem 1, we can get the desired results.

## APPENDIX G
## PROOF OF THEOREM 5

For a large time horizon $K$ and look at the packets intended for user $i$. Let $N_i(K)$ denote the number of recovered packets up to and including time slot $K$. Now consider the $m^{th}$ and $(m+1)^{th}$ recovered packets and denote the delivery time of them at user $i$ by $T_i(m)$ and $T_i(m+1)$, respectively. The inter-delivery time

$$I_i(m) = T_i(m+1) - T_i(m) \tag{55}$$

is the time between these two consecutive deliveries. Upon arrival of the $m^{th}$ recovered packet for user $i$, the age of information of user $i$ drops to the value $D_i(m)$ which represents how much delay the packet has experienced in the system. Let $L_i$ be the number of remaining time slots after the last packet recovery for user $i$. Now define $\Gamma_i(m)$ as the sum of age functions $h_i(k)$, where $k$ is in the interval $[T_i(m), T_i(m+1))$:

$$\Gamma_i = \sum_{k=T_i(m)}^{T_i(m)+I_i(m)-1} h_i(k) \tag{56}$$

$$= \frac{1}{2}I_i^2(m) - \frac{1}{2}I_i(m) + D_i(m-1)I_i(m). \tag{57}$$

It follows that in the limit of large $K$, we have

$$J^\pi(M) = \lim_{K \to \infty} \left[ \frac{1}{M} \sum_{i=1}^{M} \alpha_i \frac{1}{K} \sum_{m=1}^{N_i(K)} \Gamma_i(m) \right]. \tag{58}$$

Using this formulation, we next lower bound EAoI. Let $R_i$ denote the capacity outer bound of user $i$, defined in (61) and (62). Note that in the limit of large $K$, $\frac{N_i(K)}{K}$ is the rate/throughput of user $i$ under policy $\pi$, denoted by $R_i^\pi$.

Then, we prove the first part of Theorem 5. Consider any scheduling policy and a large time-horizon $K$. The EAoI can be re-written in terms of $\Gamma_i(m)$:

$$J_K^\pi = \frac{1}{M} \sum_{i=1}^{M} \frac{\alpha_i}{K} \left( \sum_{m=1}^{N_i(K)} \Gamma_i(m) + \frac{1}{2}L_i^2 + D_i(N_i(K))L_i - \frac{1}{2}L_i \right).$$

Since $D_i(m) \geq 1$ for all $1 \leq m \leq N_i(K)$, we can lower bound (58) by substituting $D_i(m-1) = 1$. Using similar steps as [11, Eqns. (12) - (17)],

$$\lim_{K \to \infty} \mathbb{E}[J_K^\pi] \geq \lim_{K \to \infty} \mathbb{E}\left[ \frac{1}{2M} \sum_{i=1}^{M} \frac{\alpha_i K}{N_i(K)} + \frac{\sum_{i=1}^{M} \alpha_i}{2M} \right]. \tag{59}$$

Now note that by the Cauchy-Schwarz inequality, we have

$$\lim_{K \to \infty} \mathbb{E}\left[ \sum_{i=1}^{M} \frac{N_i(K)}{\alpha_i K} \right] \mathbb{E}\left[ \sum_{i=1}^{M} \frac{\alpha_i K}{N_i(K)} \right] \geq M^2,$$

and thus

$$\lim_{K \to \infty} \mathbb{E}\left[ \sum_{i=1}^{M} \frac{\alpha_i K}{N_i(K)} \right] \geq M^2 / \sum_{i=1}^{M} \frac{R_i}{\alpha_i}. \tag{60}$$

Inserting this back into (59), we obtain

$$J^\pi(M) \geq \frac{M}{2 \sum_{i=1}^{M} \frac{R_i}{\alpha_i}} + \frac{\sum_{i=1}^{M} \alpha_i}{2M}.$$

Next, we prove the second part of Theorem 5. Suppose that all packets are recovered instantaneously with one time-unit delay. A lower bound to EAoI in this scenario constitutes a lower bound to EAoI in our setup. Let $X_i(m)$ denote the inter arrival time between $m^{th}$ and $(m+1)^{th}$ packets. $\{X_i(m)\}_m$

is a geometric i.i.d sequence. Under the assumption of instantaneous recovery, $I_i(m) = X_i(m)$. It hence follows from (57) that

$$\Gamma_i(m) = \frac{1}{2}X_i^2(m) + \frac{1}{2}X_i(m).$$

Thus, similar with [31], the time-average AoI of user $i$ is

$$\mathbb{E}[h_i] = \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} h_i(k)$$

$$= \frac{\mathbb{E}[X_i^2(m)]}{2\mathbb{E}[X_i(m)]} + \frac{1}{2} = \frac{2 - \theta_i}{2\theta_i} + \frac{1}{2}$$

and

$$J^\pi(M) \geq \frac{1}{M} \sum_{i=1}^{M} \frac{\alpha_i}{\theta_i}.$$

APPENDIX H
PROOF OF COROLLARY 1

The proof consists of minimizing (22) over all rate tuples $(r_1^\pi, \ldots, r_M^\pi) \in \hat{\mathcal{C}}$ where $\hat{\mathcal{C}}$ is an outer bound on the capacity region [24, Section III].

From [24, Section III], denote $\hat{C}$ as the channel capacity outer bound. Let $\pi$ is a permutation of $[M]$ such that $\pi(M-i+1) = \hat{\pi}(i)$, where $\hat{\pi}$ is the permutation defined in [24, Definition 1]. Recall that $\sigma(\mathcal{I})$ is the probability that an erasure occurs for all users in $\mathcal{I}$. Denote $\hat{\epsilon}_{\pi(i)} = \sigma(\cup_{j=1}^{i} \{\pi(j)\})$. From [24, Lemma 3, Lemma 4], we can obtain the outer bound $\hat{C}$

$$\hat{C} = \cap_\pi \hat{C}_\pi \tag{61}$$

where

$$\hat{C}_\pi = \left\{ (R_1, R_2, \cdots, R_M) | \sum_{i \in [M]} \frac{R_{\pi(i)}}{1 - \hat{\epsilon}_{\pi(i)}} \leq 1 \right\}. \tag{62}$$

Consider symmetric and independent channels, i.e., $\epsilon_i = \epsilon$ for all $i \in [M]$. From (62), the capacity outer bound $\hat{\mathcal{C}}$ is given by

$$\hat{\mathcal{C}} = \left\{ (R_1, \cdots, R_M) | 0 \leq R_i \leq \frac{1}{\sum_{j=1}^{M} 1/(1 - \epsilon^j)} \right\}. \tag{63}$$

From (63), $LB_1^\pi$ can be re-written as

$$LB_1^\pi = \frac{M}{2\epsilon(M) \sum_{i=1}^{M} 1/\alpha_i} + \sum_{i=1}^{M} \frac{\alpha_i}{2M}.$$

where $\epsilon(M) = \frac{1}{\sum_{j=1}^{M} 1/(1-\epsilon^j)}$.