Timely Broadcasting in Erasure Networks: Age-Rate Tradeoffs

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

***Abstract*—The interplay between timeliness of communication 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 lower and upper bounds on the age of information (AoI) as a function of target rate constraints. Simulation results show that (i) coding can be beneficial in reducing AoI even without rate constraints, (ii) AoI increases as the target rate constraint increases.**

I. Introduction

A. Motivation

Sending status updates in a timely manner has significant importance in the Internet of Things (IoT) applications. A new concept *Age of Information (AoI)*, which was introduced in [1], measures the timeliness of information at a remote system. In other words, AoI measures, at the receiving side, how much time has passed since the generation time of the latest received packet.

In the context of broadcast packet erasure channels (BPECs) with feedback, coding is beneficial on AoI in BPECs with 2 users [2]. However, the impact of network coding on AoI is not apriori clear when BPECs have multiple users. In this work, we aim to seek the efficiency both in AoI and rate under coding policies. Furthermore, we reveal tradeoffs between the AoI and rate considering network coding. On 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 incur delay by waiting for the arrival/generation of other packets for the purpose of coding as well as prioritizing their transmission, hence may incur larger AoI.

The following questions will be answered in this paper: (i) Can we devise scheduling policies that schedule different coding actions, as opposed to schemes that schedule users, and show the benefit of coding, in terms of age and rate, over uncoded schemes such as those proposed in [3], [4]? (ii) What is the trends of benefits of coding? Will the benefits increase or decrease with the number of users? (iii) The number of coding actions is extremely large in BPECs with many users. Can we find a good approximation of proposed policies which can reduce the number of coding actions significantly? (iv) Does a tradeoff exist between the AoI and rate?

B. Related Work

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. In [5], 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 [6]–[10]. AoI has also been investigated in network management and scheduling. In particular, [11] 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 [12]. In [13], scheduling algorithms are designed to minimize AoI in wireless broadcast channels. [14] 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 [15]–[17] with and without channel state information.

In coding theory, previous work has mainly studied point to point channels. For example, [18]–[20] consider erasure channels, propose coding schemes, and analyze the resulting average or peak AoI in various setups. More recently, [21] 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, [22] finds an optimal block length for channel coding to minimize the average age and average peak age of information. Under various assumptions on the size of each update and the number of antennas at the transmitter and the receivers, [23] identifies the corresponding age-optimal precoding and transmission scheduling strategies. The effects of quantization and coding on the estimation quality is revealed in [24], [25], which optimizes sampling times such that minimizes the long term average minimum mean square error for each coding scheme. [26] investigates the AoI in a two-user broadcast symbol erasure channel. [23] designs optimal

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

precoding schemes to minimize AoI in a MIMO broadcast channel with multiple users and sources. However, [26] [23] only considered a FIFO channel without packet management. [27] analyzes the AoI in a multicast network with coding, while only coded packet were transmitted, which were encoded with a fixed number of uncoded packets. This is a special case of our setting. In addition, the packet management was not considered in [27].

BPECs and their variants have been considered in previous work such as [28]–[31] and rate-optimal coding algorithms are designed using (network) coding ideas. The key idea is due to [28] 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.

C. Contributions

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, Theorem 4). The bounds are functions of generation rate, erasure probability and minimum required rate. (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.

D. 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 vectors $\underline{x}$ and $\underline{z}$, $\underline{x} \preceq \underline{z}$ represents $x_i \leq z_i$ for all i. For two sets $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{A} \subset \mathcal{B}$ represents that $\mathcal{A}$ is a subset of $\mathcal{B}$; $\mathcal{A} \subsetneq \mathcal{B}$ represents that $\mathcal{A}$ is a proper subset of $\mathcal{B}$;

II. System Model

The system model extends that of [2, 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 θ_i. Let $G_i(k) = 1$ represent that a new packet is generated (for user i) in time slot k. So $\Pr\big(G_i(k) = 1\big) = \theta_i$ for $k = 1, 2, \ldots$. After each transmission, the transmitter receives feedback of age of information of each user. Every broadcast packet is received at user i with probability $1 - \epsilon_i$, $0 \leq \epsilon_i < 1$, and lost with probability ϵ_i. Erasure events at multiple users are either independent or dependent. The *transmission delay* is assumed fixed and equal to one time slot.

If a packet is not recovered at its intended user, it can be cached by other user(s) that have received it. Let $\sigma(\mathcal{I})$ represent the probability that a broadcast packet is erased by users in $\mathcal{I}$, but cached by users in $[M]\backslash\mathcal{I}$. Then,

$$\sum_{i\in\mathcal{I},\ \mathcal{I}\subset[M]} \sigma_{\mathcal{I}} = \epsilon_i, \quad i = 1, 2, \cdots, M. \tag{1}$$

The encoder can track the cached packets using the available feedback. The cached packets can act as side information. We exploit such side information in the code design at the encoder to form more efficient coded packets that are simultaneously useful for multiple users. Such code designs have also appeared in [31]–[34]. 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 [13], [14] (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 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 [35]. 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 τ_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:

$$\mathcal{S}_{\tau_i} \supset \{\cup_{j=1,j\neq i}^{\ell}\tau_j\} \qquad \forall i = 1,\dots\ell. \tag{2}$$

Then XORing packets

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

forms a coded packet x as follows

$$x = \bigoplus_{i=1}^{\ell} p_i \tag{3}$$

which is simultaneously decodable at all users $\{\tau_1, \dots, \tau_\ell\}$. To view condition (2) alternatively, draw a side information graph $\mathcal{G}$ with nodes $V = \{1, \dots, M\}$. Add an edge between 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 (2) corresponds to the subgraph induced by nodes $\{\tau_1, \dots, \tau_\ell\}$ forming a clique of size ℓ. A simple example with a clique of length 3 is given in Figure 1.

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 $3^{\frac{M}{3}}$ [36]) we aim to choose (schedule) one that leads to a coding action with the most benefit in terms of information freshness and rate.

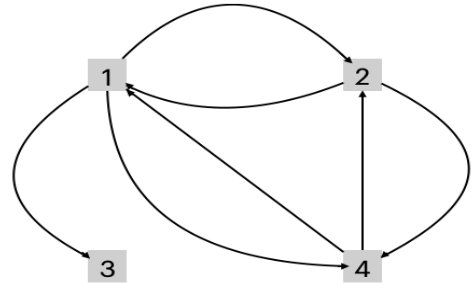

Fig. 1: Consider users $1, 2, 3, 4$. Let packets $p_1 \in Q_{1,\{2,3,4\}}$, $p_2 \in Q_{2,\{14\}}$ and $p_4 \in Q_{4,\{12\}}$, so $1 \leftrightarrow 2 \leftrightarrow 4 \leftrightarrow 1$ is a clique, $x = p_1 \oplus p_2 \oplus p_4$.

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 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 complicated 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) *Uncoded packets which form a packet in $Q_{\mathcal{I},\mathcal{K}}$ are intended for distinct users in $\mathcal{I}$.*

Assumption 1 (1) implies that only if user i receives a coded packet containing an uncoded packet intended for himself, he can decode this uncoded packet. Otherwise, the coded packet would not be transmitted to user i. Under Assumption 1 (2), a coded packet can not contain uncoded packets which are intended for an identical user. 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 can not be larger than M.

Similar to Section II-A1, consider queue sequences $\{Q_{\tau_i,\mathcal{S}_{\tau_i}}\}_{i=1}^{\ell_1}$ and $\{Q_{\mathcal{I}_i,\mathcal{K}_i}\}_{i=1}^{\ell_2}$. Suppose the following condition holds:

$$\begin{aligned} &\mathcal{S}_{\tau_i} \supset \{\tau_j\}_{j=1,j\neq i}^{\ell_1} \cup \{\mathcal{I}_j\}_{j=1}^{\ell_2}, \quad \forall i = 1,\dots\ell_1 \\ &\mathcal{K}_i \supset \{\tau_j\}_{j=1}^{\ell_1} \cup \{\mathcal{I}_j\}_{j=1,j\neq i}^{\ell_2}, \quad \forall i = 1,\dots\ell_2 \\ &\tau_i \cap \mathcal{I}_j \neq \varnothing, \quad \mathcal{I}_i \cap \mathcal{I}_j \neq \varnothing. \end{aligned} \tag{4}$$

The last condition in (4) comes from Assumption 1 (2) directly. Then XORing packets

$$p_i \in Q_{\tau_i,\mathcal{S}_{\tau_i}}, \quad x_i \in Q_{\mathcal{I}_i,\mathcal{K}_i}$$

forms a coded packet x as follows

$$x = \bigoplus_{i=1}^{\ell_1} p_i \bigoplus_{i=1}^{\ell_2} x_i \tag{5}$$

which is simultaneously decodable at all users $\{\tau_1, \ldots, \tau_{\ell_1}\} \cup \{\mathcal{I}_i\}_{i=1}^{\ell_2}$. To view condition (4) 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, j) $\big($respectively, $(\mathcal{I}_i, j)\big)$ if $Q_{i,\mathcal{S}_i}$ $\big($respectively, $Q_{\mathcal{I}_i,\mathcal{K}_i}\big)$ is non-empty for some set $\mathcal{S}_i$ $\big($respectively, $\mathcal{K}_i\big)$ that has j as an element. On this graph, condition (4) 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* defined in [37]. 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 in that case.

Definition 1. *Denote the generation time of the packet received by user i in time slot k as $v_i(k)$. $h_i(k)$ evolves as*

$$h_i(k) = \begin{cases} \min\{h_i(k-1)+1, k - v_i(k)\}, & d_i(k) = 1 \\ h_i(k-1)+1, & d_i(k) = 0 \end{cases} \tag{6}$$

and $h_i(0) = 1$.

We are interested in policies that minimize the following expected weighted sum of AoI (EAoI) at the users:

$$\mathbb{E}\left[\frac{1}{MK}\sum_{k=1}^{K}\sum_{i=1}^{M} \alpha_i h_i^{\pi}(k)\right], \tag{7}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_M$ are weights.

In practice, 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 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 throughput requirement of node i. Without loss of generality, we assume that $(q_1, q_2, \cdots, q_M)$ is in the capacity region. Let $d_i^{\pi}(k) = 1$ if user i decodes a packet of type i in time k, and $d_i^{\pi}(k) = 0$ otherwise. Similar to [4], we define the *long-term rate* of node i when policy π is employed as

$$r_i^{\pi} := \lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\big[d_i^{\pi}(k)\big]. \tag{8}$$

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

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

$$r_i^{\pi} \geq q_i, i = 1, 2, \cdots, M. \tag{9}$$

Ultimately, we seek to schedule coding actions in order to achieve a judicious tradeoff between EAoI and communication rate, as outlined below. Combining (7), (8) and (9), the objective function is given by the following optimization problem:

$$\begin{aligned} \min_{\pi} \quad & \lim_{K\to\infty} \mathbb{E}[J_K^{\pi}] \\ s.t \quad & r_i^{\pi} \geq q_i, i = 1, 2, \cdots, M. \end{aligned} \tag{10}$$

where

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

Recall that coding actions that corresponds to cliques with large sizes are beneficial for rate efficiency, but may not be desired for age efficiency. Note that as M gets large, the number of coding actions increase exponentially. We seek to find the relationship between age and rate efficiency.

III. Scheduling Coding Actions

In this section, we consider coding policies with uncoded caching, i.e., all packets cached are uncoded. We develop and analyze max-weight policies that schedule the coding actions we outlined in the previous section in order to optimize (10). Each coding action refers to 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.

In order to optimize for age, we need to define the AoI of queues $Q_{i,\mathcal{S}}$ in the virtual network of queues and understand their time evolution.

Recall that $Q_{i,\mathcal{S}}$ defines the queue to contain packets for 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 any packet a can not be in two queues, i.e., 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 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.

A. Age of Information

To formulate $h_i(k)$ precisely, we first consider the AoI of the virtual queues.

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.*

Proof. The proof of Lemma 1 is given in Appendix B. $\square$

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 $w_{i,\varnothing}(0) = h_i(0)$. 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 useless 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

$$w_{i,\varnothing}(k+1) = \begin{cases} 0 & G_i(k) = 1 \\ \min\{w_{i,\varnothing}(k) + 1, h_i(k) + 1\} & G_i(k) = 0 \end{cases}. \tag{11}$$

Before obtaining the AoI of $Q_{i,\mathcal{S}}$ as $w_{i,\mathcal{S}}(k)$ with $\mathcal{S} \neq \varnothing$, we first define an indicator $t_{i,\mathcal{S}}(k)$. Let $t_{i,\mathcal{S}}(k) = 1$ if the latest packet in $Q_{i,\mathcal{S}}$ is encoded and transmitted in time slot k, otherwise $t_{i,\mathcal{S}}(k) = 0$. In Figure 1, if the encoder transmits $x = p_1 \oplus p_2 \oplus p_4$ in time slot k, then $t_{1,\{2,3,4\}}(k) = 1$, $t_{2,\{1,4\}}(k) = 1$ and $t_{4,\{1,2\}}(k) = 1$. The relationship between $d_i(k)$ and $t_{i,\mathcal{S}}$ is

$$d_i^\pi(k) = E_i \sum_{\mathcal{S} \subset [M] \setminus i} t_{i,\mathcal{S}}^\pi(k), \tag{12}$$

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,

$$\mathbb{E}[d_i^\pi(k)] = (1 - \epsilon_i) \sum_{\mathcal{S} \subset [M] \setminus i} \mathbb{E}[t_{i,\mathcal{S}}^\pi(k)]. \tag{13}$$

Similarly, if $\mathcal{S} \neq \varnothing$, 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 $Q_{i,\mathcal{S}}(0) = h_i(0)$. Thus, $w_{i,\mathcal{S}}(k)$ increases by 1 unless $Q_{i,\mathcal{S}}$ recovers a fresher packet. If packet $p \in Q_{i,\mathcal{S}}$ is recovered by other users in $\mathcal{I}$, $\mathcal{I} \cap \mathcal{S} = \varnothing$, so $p \in Q_{i,\mathcal{I} \cup \mathcal{S}}$ and $p \notin Q_{i,\mathcal{S}}$. Recall that $d_i(k) = 1$ represents that user i recovers a packet in time k (in Section II-B). Similarly, let $d_{\mathcal{S}}(k) = 1$ if users in $\mathcal{S}$ receive a (coded) packet, otherwise $d_{\mathcal{S}}(k) = 0$. Let $\mathcal{S}' \subset \mathcal{S}$. The recursion of $w_{i,\mathcal{S}}(k)$ is

$$w_{i,\mathcal{S}}(k+1) = \begin{cases} \min\{w_{i,\varnothing}(k) + 1, h_i(k) + 1\} & \mathcal{P}_{i,\mathcal{S}}^{(1)}(k) \\ \min\{w_{i,\mathcal{S}'}(k) + 1, h_i(k) + 1\} & \mathcal{P}_{i,\mathcal{S}}^{(2)}(k) \\ \min\{w_{i,\mathcal{S}}(k) + 1, h_i(k) + 1\} & \text{otherwise} \end{cases}. \tag{14}$$

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 (11) and (14), the recursion of $h_i(k)$ is

$$h_i(k+1) = \begin{cases} w_{i,\mathcal{S}}(k) + 1 & t_{i,\mathcal{S}}(k) = 1, d_i(k) = 1 \\ h_i(k) + 1 & \text{otherwise} \end{cases}. \tag{15}$$

B. Age-based Max-Weight Policy

It is well established that coding actions can enhance the communication rate of broadcast channels [32]. In this work, we seek the tradeoff between the AoI and communication rate. To schedule packets efficiently, similar to [2], [4], [14], we propose Age-Rate Max-Weight (ARM) policies to minimize EAoI defined in (7).

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

$$\delta_{i,\mathcal{S}}(k) = h_i(k) - w_{i,\mathcal{S}}(k). \tag{16}$$

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

To capture the impact of communication rate, let $x_i(k)$ be the throughput debt associated with node i at the beginning of slot k [4]. The throughput debt evolves as

$$x_i(k+1) = kq_i - \sum_{\tau=1}^{k} d_i^\pi(\tau) \tag{17}$$

The value of kq_i is the minimum average number of packets that node i should have received by slot $k+1$ and $\sum_{\tau=0}^{k} d_i^\pi(\tau)$ is the total number of packets actually recovered in the same interval. In fact, strong stability of the process $x_i^+(k)$, i.e.,

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

is sufficient to establish that the minimum rate constraint, $r_i^\pi \geq q_i$, is satisfied [4], [38, Theorem 2.8].

We define the encode'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

$$L\big(S(k)\big) = \sum_{i=1}^{M} \beta_i h_i(k) + \lambda \sum_{i=1}^{M} \big(x_i^+(k)\big)^2 \tag{18}$$

where $\beta_i, \lambda > 0$. Here, quadratic function for $x_i(k)$ is to maximize communication rate [3], [4], [14], and linear function for $h_i(k)$ is to simplify the derivation. The one-slot Lyapunov Drift is defined as

$$\Theta(k) = \mathbb{E}\Big[L\big(S(k+1)\big) - L\big(S(k)\big) | S(k)\Big]. \tag{19}$$

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

$$f_i(k) = \Big(\big(x_i(k) + q_i\big)^+\Big)^2 - \Big(\big(x_i(k) + q_i - 1\big)^+\Big)^2. \tag{20}$$

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

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

Remark 1. *In the case only AoI plays role in decision making ($q_i=0$ for all i, hence it is easy to show that $f_i(k)=0$ for all i,k), from Lemma 1, only the latest packets matters. We can assume that the buffer size of every queue is 1. Thus, the stability region is that $\{\theta_i\le 1, i=1,2,\cdots,M\}$.*

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. □

Let $\beta_i=0$ for all i in Definition 2, then applying AMW lead to the capacity region, denoted by $C^{uncoded}$.

Theorem 2. *If $(q_1,q_2,\cdots,q_M)\in C^{uncoded}$, then the strong stability of the process $x_i^+(k)$ satisfies. EAoI has the following upper bound:*

$$UB=\frac{1}{M}\sum_{i=1}^{M}\big(\frac{\alpha_i}{\theta_i}+\frac{\alpha_i}{q_i(1-\epsilon_i)}\big)+\lambda \tag{21}$$

Proof. The proof of Theorem 2 is given in Appendix D. □

IV. Coding Policies with Coded Caching

Now we consider coded caching, where coded packets can be cached in caches. Caching coded packets can bring more coding opportunity in the system. However, the number of coding actions increases exponentially due to 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 $\{a_t\}_{t\ge1}$, $\{b_t\}_{t\ge1}$ and $\{c_t\}_{t\ge1}$, respectively. From Assumption 1 (2), the length of coded packets is at most 3. From Assumption 1 (1), if coded packets have length 3, then they must have forms as $a_{t_1}\oplus b_{t_2}\oplus c_{t_3}$. Again, from Assumption 1 (1), coded packets which are cached must have length 2.

Let $\mathcal{X}=\big\{\{1,2\},\{1,3\},\{2,3\}\big\}$, $\mathcal{I}\in\mathcal{X}$ and $u=[3]\backslash\mathcal{I}$. Queues $Q_{\mathcal{I},u}$ are the queues where 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.
- $A(k)=Q_{\tau_i,\mathcal{S}_{\tau_i}}\oplus Q_{\mathcal{I},u}$: 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.

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

Lemma 2. *If $\alpha_1u(t_1)+\alpha_2v(t_2)>\alpha_1u(t_3)+\alpha_2v(t_4)$, 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.,

$$x_0=\min\{x\in Q_{\mathcal{I},u}|\alpha_iu_i+\alpha_jv_j\}.$$

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

$$w_{\mathcal{I},u}(k)=(w_{\mathcal{I},u}^{(i)}(k),w_{\mathcal{I},u}^{(j)}(k)) \tag{22}$$

where

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

and $w_{\mathcal{I},l}(0)=(h_i(0),h_j(0))$. Similar with (16), we can define *age-gain of $Q_{\mathcal{I},j}$* as

$$\delta_{\mathcal{I},l}(k)=h_{\mathcal{I}}(k)-w_{\mathcal{I},l}(k)\triangleq(\delta_{\mathcal{I},l}^{(i)}(k),\delta_{\mathcal{I},l}^{(j)}(k)) \tag{23}$$

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

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 (12) can be extended as

$$d_i(\tau)=E_i\Big(\sum_{\mathcal{S}\subset[M]\backslash i}t_{i,\mathcal{S}}(\tau)+\sum_{\mathcal{I}:i\in\mathcal{I}}t_{\mathcal{I},u}^{(i)}(\tau)\Big). \tag{24}$$

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 $L\big(\vec{h}(k),\vec{Y}(k)\big)$ is defined as

$$L\big(b(k)\big)=\sum_{i=1}^{M}\beta_ih_i(k)+\lambda\sum_{i=1}^{M}\big(x_i^+(k)\big)^2, \tag{25}$$

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

$$\Theta(k)=\mathbb{E}\Big[L\big(b(k+1)\big)-L\big((b(k)\big)|b(k)\Big]. \tag{26}$$

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

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

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 $\beta_i = 0$ for all i, then applying AMW in Definition 3 lead to the capacity region, denoted by C^{coded}.

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* (21).

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

V. Lower Bound

From [2], [39], the lower bound of AoI depends on the channel capacity. Similar to [39, Section III], we derive two lower bounds on the achievable age performance. The first lower bound, denoted by LB_1, is derived by assuming that there is always a fresh (regular) packet to be delivered. The second lower bound, denoted by LB_2, is derived by assuming that all (regular) packets are delivered instantaneously upon their arrivals.

Theorem 5. *For any communication policy π,*

(1) *Let R_i^π be the communication rate of user i under π,*

$$LB_1^\pi = \frac{M}{2\sum_{i=1}^M R_i^\pi/\alpha_i} + \sum_{i=1}^M \frac{\alpha_i}{2M};$$

(2) $LB_2 = \frac{1}{M}\sum_{i=1}^M \frac{\alpha_i}{\theta_i}$.

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

To find an upper bound of R_i^π, we consider the capacity outer bound because the channel capacity is sophisticated due to these coding actions. From [32, Section III], denote $\hat{C}$ as the channel capacity outer bound. Let π is a permutation of $[M]$ such that $\pi(M-i+1) = \hat{\pi}(i)$, where $\hat{\pi}$ is the permutation defined in [32, 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 [32, Lemma 3, Lemma 4], we can obtain the outer bound $\hat{C}$

$$\hat{C} = \cap_\pi \hat{C}_\pi \tag{27}$$

where

$$\hat{C}_\pi = \Big\{(R_1, R_2, \cdots, R_M)|\sum_{i\in[M]} \frac{R_{\pi(i)}}{1-\hat{\epsilon}_{\pi(i)}} \le 1\Big\}. \tag{28}$$

Consider symmetric and independent channels, i.e., $\epsilon_i = \epsilon$ for all $i \in [M]$. From (28), the capacity outer bound $\hat{\mathcal{C}}$ is given by

$$\hat{\mathcal{C}} = \Big\{(R_1, \cdots, R_M)|0 \le R_i \le \frac{1}{\sum_{j=1}^M 1/(1-\epsilon^j)}\Big\}. \tag{29}$$

From (29), LB_1^π 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 $\epsilon(M) = \frac{1}{\sum_{j=1}^M 1/(1-\epsilon^j)}$.

VI. Numerical Results and Discussion

In this section, we answer the questions proposed in Section I-A. Let $\epsilon_i = \epsilon$, $q_i = q$, and $\theta_i = \theta$ for $i \in [M]$.

A. Benefits of Coding

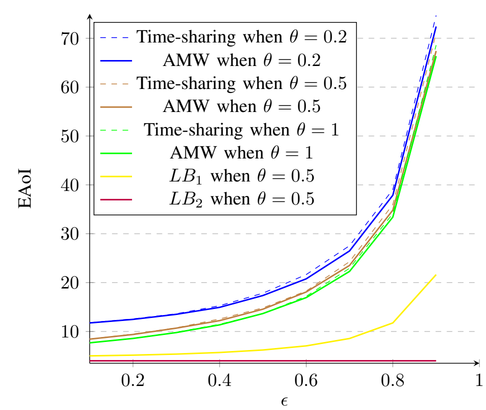

Fig. 2: EAoI as a function of ϵ and θ when $M = 6$, the upper bound, and the lower bounds.

We first consider benefits of coding actions. The ARM policy in Definition 2 and the Time-sharing policy in Appendix A are compared in Figure 2 - Figure 4. 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 $\{\theta \le 1\}$. Set $M = 6$, $\lambda = 0$ and $\beta_i = \min\{i, \max\{0, i-3\}\}$.

Figure 2 plots the EAoI under ARM policy and time-sharing policy, as well as the two lower bounds in Theorem 5, which shows coding is indeed beneficial when the erasure probability ϵ is relative large (≥ 0.6), or the arrival rate θ is relative small (≤ 0.5). When θ is fixed, EAoI increases with ϵ.

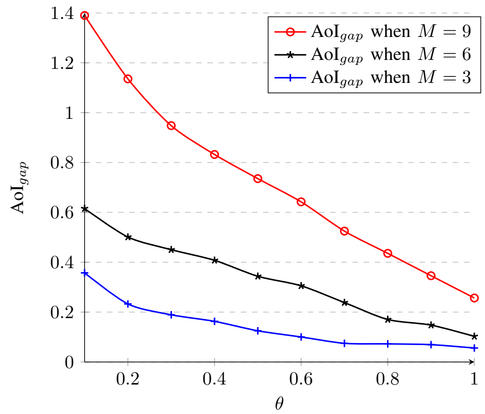

Fig. 3: AoI$_{gap}$ v.s θ when $\epsilon = 0.6$.

Then, we define AoI$_{gap}$ as the gap between the EAoI under the ARM and time-sharing policies. The relationship between AoI$_{gap}$ and θ (ϵ) is provided in Figure 3 (Figure 4). The curves are not smooth because of the randomness of the system. In Figure 3, fix $\epsilon = 0.6$, AoI$_{gap}$ (benefits of coding) decreases with the arrival rate θ. This is because the (expected) number of newly incoming packets increases when θ gets large, which can weaken the impact of coding actions. In Figure 4, fix $\theta = 0.2$, AoI$_{gap}$ increases with ϵ. This is because lost packets (due to erasure probability) can be cached and caches can provide more coding opportunities. AoI$_{gap}$ increases slowly when is small, and sharply when is large. In addition, from Figure 3 and Figure 4, the benefits of coding increase with M.

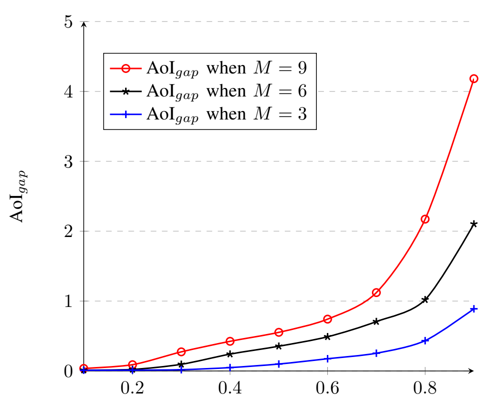

Fig. 4: AoI$_{gap}$ as a function of ϵ when $\theta = 0.2$.

B. Impact of Maximal Clique Size

The impact of maximal clique size is captured in Figure 5. Let buffer size of (virtual) queues be 1, $M = 6$, $\lambda = 0$, $\epsilon = 0.6$ and $\beta_i = \min\{i, \max\{0, i-3\}\}$. ARM in Definition 2 with maximal clique size $= 2, 3, 4$ are applied in Figure 5. From Figure 5, we can see that the ARM policy with maximal clique size of 2 is a good approximation. Then, we can reduce the number of coding actions sharply by setting the maximal clique size be 2.

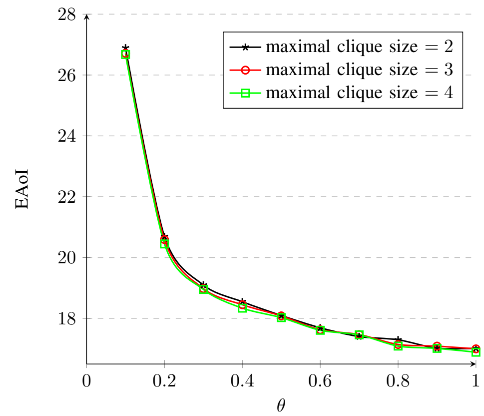

Fig. 5: EAoI under ARM policies under the.

C. Tradeoff between Age and Rate

We investigate the tradeoff between the AoI and rate in this subsection. Take $M = 3$ as an example, let $\beta_i = 3$, $\epsilon = 0.6$ in the ARM policy (Definition 2). The channel capacity under the ARM policy can be obtained numerically (by letting $\beta_i = 0$), which is around 0.44.

In Figure 6, we first investigate the relationship between q and EAoI (the red star curve). Let $\theta = 0.14$, $\lambda = 10$, $q \in [0, 0.1368]$. EAoI increases with q, which implies that if the minimum required throughput becomes larger, the system has to sacrifice EAoI to satisfy the rate constraints. Then, the relationship between λ and EAoI is investigated (the black circle curve). Let $\theta = 0.44/3$, $q = 0.14$, $\lambda \in [0, 100]$. EAoI increases with λ. In other words, if the rate constraints become more important, then EAoI increases.

The EAoI and rate under the time-sharing policy, the ARM with uncoded and coded caching (defined in Definition 3) is revealed in Figure 7. Set $\theta = q$, $\lambda = 1$. Among these policies, the AMW with coded caching (respectively, time-sharing policy) provides the largest (respectively, smallest) channel capacity. Approaching the channel capacity, EAoI reach the minimum.

VII. Conclusion

We investigated the benefit of coding in memoryless broadcast channel with M users. A novel framework of network AoI based on information graph was proposed. We proposed an age-rate MW policy under rate constraints. 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.

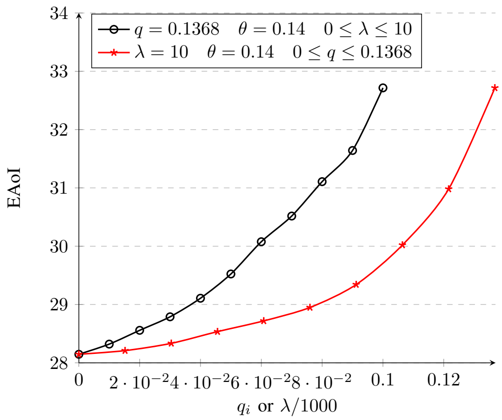

Fig. 6: EAoI vs. q and λ when $M = 3$, $\epsilon = 0.6$

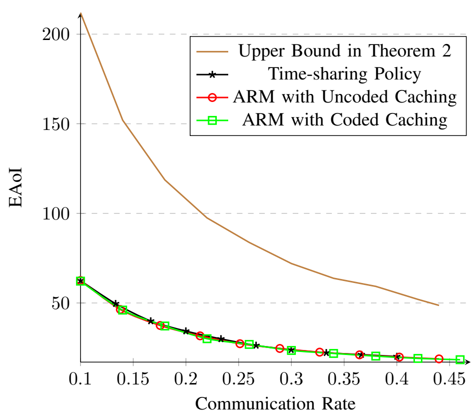

Fig. 7: EAoI v.s rate (when $M = 3$) under time-sharing, ARM with uncoded and coded caching polices when $\epsilon = 0.6$.

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

[3] 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.

[4] 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.

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

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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, 2018.

[12] 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, 2018.

[13] 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.

[14] 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.

[15] 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.

[16] 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)*, 2018, pp. 1 – 8.

[17] ——, "Distributed Scheduling Algorithms for Optimizing Information Freshness in Wireless Networks," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1 – 5.

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

[19] 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.

[20] 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.

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

[22] 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.

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

[24] A. Arafa, K. Banawan, K. G. Seddik and V. Poor, "Timely Estimation Using Coded Quantized Samples," arXiv: 2004.12982, 2020.

[25] ——, "Samppe, Quantize and Encode: Timely Estimation over Noisy Channels," arXiv: 2007.10200, 2020.

[26] S. Feng and J. Yang, "Adaptive Coding for Information Freshness in a Two-user Broadcast Erasure Channel," arXiv: 1905.00521, 2019.

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

[28] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback - Capacity and algorithms," in *2009 Workshop on Network Coding, Theory, and Applications*, 2009, pp. 54 – 61.

[29] 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)*, 2010, pp. 1347 – 1354.

[30] 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)*, 2012, pp. 108 – 113.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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

[37] 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.

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

[39] 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 ESAoI 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

$$L\big(S(k)\big)=\sum_{i=1}^{M}\beta_i h_i(k)+\lambda\sum_{i=1}^{M}\big(x_i^+(k)\big)^2, \tag{30}$$

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

$$\Theta(k)=\mathbb{E}\Big[L\big(S(k+1)\big)-L\big(S(k)\big)|S(k)\Big]. \tag{31}$$

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

Definition 4. *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

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

We first consider $\Theta_1(k)$. From (15),

$$\begin{aligned}h_i(k+1)=t_{i,\varnothing}(k)d_i(k)\big(w_{i,\varnothing}(k)+1\big)\\ +\big(1-t_{i,\varnothing}(k)d_i(k)\big)\big(h_i(k)+1\big).\end{aligned} \tag{32}$$

Similar with (12), we have $d_i(k)=H_i t_{i,\varnothing}(k)$. Using (36), we can re-write the Lyapunov Drift as follows:

$$\begin{aligned}\Theta_1(k)=&\sum_{i=1}^{M}\beta_i\mathbb{E}\Big[H_i t_{i,\varnothing}(k)\big(w_{i,\varnothing}(k)-h_i(k)\big)+1|S(k)\Big]\\ =&-\sum_{i=1}^{M}\beta_i\mathbb{E}\big[H_i t_{i,\varnothing}(k)\delta_{i,\varnothing}(k)|S(k)\big]+\sum\beta_i\\ =&-\sum_{i=1}^{M}\beta_i(1-\epsilon_i)\mathbb{E}\big[t_{i,\varnothing}(k)|S(k)\big]\delta_{i,\varnothing}(k)+\sum\beta_i.\end{aligned}$$

Therefore,

$$\Theta_1(\vec{h}(k))=\sum_{i=1}^{M}\beta_i-\sum_{i=1}^{M}1_{\{A(k)=Q_{i,\varnothing}\}}\beta_i\delta_{i,\varnothing}(k)(1-\epsilon_i). \tag{33}$$

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

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

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

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

From (33) and (39), minimizing $\Theta(k)$ is equivalent to maximize

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

□

APPENDIX B
PROOF OF LEMMA 1

For any $\mathcal{S}\subset[M]\backslash i$, we first consider $\mathcal{S}=\varnothing$ (Step 1), and 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_τ, $\tau\in\{1,2\}$, 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. Another coded packet encoded by uncoded packets from $j_1',j_2',\cdots,j_m'$ provides the same AoI reduction (as c') for users $j_1',j_2',\cdots,j_m'$.

Therefore, 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,

$$\sum_{\mathcal{S}_i\subset[M]\backslash i}t_{i,\mathcal{S}_i}(k)\leq 1. \tag{35}$$

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

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

We first consider $\Theta_1(k)$. Using (12), (15) and (35),

$$\begin{aligned}h_i(k+1)=&\sum_{\mathcal{S}\subset[M]\backslash i}t_{i,\mathcal{S}}(k)E_i\big(w_{i,\mathcal{S}}(k)+1\big)\\&+\big(1-\sum_{\mathcal{S}\subset[M]\backslash i}t_{i,\mathcal{S}}(k)E_i\big)\big(h_i(k)+1\big).\end{aligned} \tag{36}$$

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

$$\begin{aligned}\Theta_1(k)=&\sum_{i=1}^{M}\beta_i\mathbb{E}\left[\sum_{\mathcal{S}_i\subset[M]\backslash i}E_it_{i,\mathcal{S}_i}(k)\big(w_{i,\mathcal{S}_i}(k)-h_i(k)\big)+1|S(k)\right]\\=&-\sum_{i=1}^{M}\beta_i\mathbb{E}\big[\sum_{\mathcal{S}_i\subset[M]\backslash i}E_it_{i,\mathcal{S}_i}(k)\delta_{i,\mathcal{S}_i}(k)|S(k)\big]+\sum\beta_i\\=&-\sum_{i=1}^{M}\beta_i(1-\epsilon_i)\sum_{\mathcal{S}_i\subset[M]\backslash i}\mathbb{E}\big[t_{i,\mathcal{S}_i}(k)|S(k)\big]\delta_{i,\mathcal{S}_i}(k)+\sum\beta_i.\end{aligned}$$

Therefore,

$$\begin{aligned}\Theta_1(\vec{h}(k))=&\sum_{i=1}^{M}\beta_i-\sum_{i=1}^{M}1_{\{A(k)=Q_{i,\varnothing}\}}\beta_i\delta_{i,\varnothing}(k)(1-\epsilon_i)\\&-\sum_{l=2}^{M}\sum_{\tau_1,\tau_2,\cdots,\tau_l}1_{\{A(k)=\oplus_{u\in[l]}Q_{\tau_j,\mathcal{S}_{\tau_u}}\}}\\&\times\sum_{u=1}^{l}\beta_{\tau_u}\delta_{i,\mathcal{S}_{\tau_u}}(k)(1-\epsilon_{\tau_u}).\end{aligned} \tag{37}$$

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

$$\mathbb{E}[(x_i^+(k+1))^2|S(k)] = \mathbb{E}\Big[\Big(\big(x_i(k)+q_i-d_i(k)\big)^+\Big)^2|S(k)\Big]$$
$$= \begin{cases} \big((x_i(k)+q_i)^+\big)^2 & \text{if } \sum_{\mathcal{S}_i\subset[M]\setminus i} t_{i,\mathcal{S}_i}(k)=0 \\ \big((x_i(k)+q_i-1)^+\big)^2(1-\epsilon_i) & \\ +\big((x_i(k)+q_i)^+\big)^2\epsilon_i & \text{if } \sum_{\mathcal{S}_i\subset[M]\setminus i} t_{i,\mathcal{S}_i}(k)=1 \end{cases}. \tag{38}$$

Then, we consider the following two cases:

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

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

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

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

Therefore,

$$\tilde{\Theta}_2(k)=\sum_{j=1}^M\Big(\big(x_j(k)+q_j\big)^+\Big)^2-\sum_j^M 1_{\{A(k)=Q_{j,\varnothing}\}}f_j(k) - \sum_{l=2}^M\sum_{\tau_1,\tau_2,\cdots,\tau_l}1_{\{A(k)=\oplus_{u\in[l]}Q_{\tau_u,\mathcal{S}_{\tau_u}}\}}\sum_{u=1}^l f_{\tau_u}(k) \tag{39}$$

From (37) and (39), minimizing $\Theta(k)$ is equivalent to maximize

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

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,

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

Then,

$$\sum_{j=1}^M\mu_{i,\varnothing}+\sum_{l=2}^M\sum_{\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l}}\mu_{(\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l})}=1. \tag{40}$$

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,

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

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 $\mathcal{S}\ne\varnothing$,

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

From [4, Eqn.(60) - (62)],

$$\big(x_i^+(k+1)\big)^2-\big(x_i^+(k)\big)^2 \le -2x_i^+(k)\big(d_i(k)-q_i\big)+1 \le -2x_i^+(k)\big(1-q_i\big)+1,$$

thus

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

Therefore,

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

By algebra, we have

$$\sum_{k=1}^K\sum_{i=1}^M\beta_i(1-\epsilon_i)\mu_{i,\varnothing}\mathbb{E}[h_i(k)]+\sum_{k=1}^K\sum_{i=1}^M 2\lambda\mathbb{E}[x_i^+(k)](1-q_i) \le-\sum_{k=1}^K\mathbb{E}[\Theta(k)]+\sum_{k=1}^K\sum_{i=1}^M\beta_i(1-\epsilon_i)\mu_{i,\varnothing}\mathbb{E}[w_{i,\varnothing}(k)] + K\sum_{i=1}^M\beta_i+KM\lambda. \tag{41}$$

From (11), $w_{i,\varnothing}(k)\le W_i(k)$, where $W_i(k)$ is defined in

$$W_i(k+1)=\begin{cases}0 & G_i(k)=1\\ W_i(k)+1 & G_i(k)=0\end{cases}. \tag{42}$$

Substituting (42) into (41),

$$L_1+L_2\le L_3+\sum_{k=1}^K\sum_{i=1}^M\beta_i(1-\epsilon_i)\mu_i\mathbb{E}[W_i(k)]+C. \tag{43}$$

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 $K \to \infty$,

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

Thus, dividing by KM on both sides of (43), we have

$$\begin{aligned}&\lim_{K\to\infty}\frac{L_1}{K} + \lim_{K\to\infty}\frac{L_2}{K}\\ &\le \frac{\sum_{i=1}^{M}\beta_i(1-\epsilon_i)\mu_{i,\varnothing}\sum_{k=1}^{K}\mathbb{E}[W_i(k)]}{KM} + \frac{\sum_{i=1}^{M}\beta_i}{M} + \lambda.\end{aligned} \tag{44}$$

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

$$\begin{aligned}&\lim_{K\to\infty}\frac{L_1}{K} + \lim_{K\to\infty}\frac{L_2}{K}\\ \le&\frac{1}{M}\sum_{i=1}^{M}\beta_i\big(\frac{(1-\epsilon_i)\mu_{i,\varnothing}}{\theta_i}+1\big)+\lambda.\end{aligned} \tag{45}$$

Since $L_1 > 0$, so (45) is reduced to

$$\begin{aligned}&\sum_{i=1}^{M}2\lambda(1-q_i)\lim_{K\to\infty}\frac{\sum_{k=1}^{K}\mathbb{E}[x_i^+(k)]}{K}\\ &\le \frac{1}{M}\sum_{i=1}^{M}\beta_i\big(\frac{(1-\epsilon_i)\mu_{i,\varnothing}}{\theta_i}+1\big)+\lambda.\end{aligned} \tag{46}$$

Then,

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

Since $L_2 > 0$, so

$$\lim_{K\to\infty}\frac{L_1}{MK} \le \frac{1}{M}\sum_{i=1}^{M}\beta_i\big(\frac{(1-\epsilon_i)\mu_{i,\varnothing}}{\theta_i}+1\big)+\lambda. \tag{47}$$

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 8. Take user 1 as an example, Cut 1 implies that

$$\mu_{1,\varnothing}(1-\epsilon_1) + \sum_{j\in\{2,3\}}\mu_{1,\varnothing}\sigma([3]\backslash j) + \mu_{1,\varnothing}\sigma(\{1\}) \ge q_1, \tag{48}$$

Cut 2 implies that

$$\begin{aligned}&\mu_{1,\varnothing}(1-\epsilon_1) + \mu_{1,\varnothing}\sigma(\{1,3\}) + \mu_{1,\varnothing}\sigma(\{1\})\\ &+ \big(\mu_{1,\{3\},3,\{1\}} + \mu_{1,\{3\},3,\{1,2\}}\big)\big(1-\epsilon_1+\sigma(\{1\})\big) \ge q_1,\end{aligned} \tag{49}$$

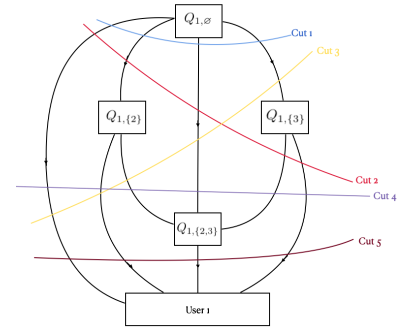

Fig. 8: The flow of user 1 in the virtual network.

Cut 3 implies that

$$\begin{aligned}&\mu_{1,\varnothing}(1-\epsilon_1) + \mu_{1,\varnothing}\sigma(\{1,2\}) + \mu_{1,\varnothing}\sigma(\{1\})\\ &+ \big(\mu_{1,\{2\},2,\{1\}} + \mu_{1,\{2\},2,\{1,3\}}\big)\big(1-\epsilon_1+\sigma(\{1\})\big) \ge q_1,\end{aligned} \tag{50}$$

Cut 4 implies that

$$\begin{aligned}&\mu_{1,\varnothing}(1-\epsilon_1) + \mu_{1,\varnothing}\sigma(\{1\})\\ &+ \big(\mu_{1,\{3\},3,\{1\}} + \mu_{1,\{3\},3,\{1,2\}}\big)\big(1-\epsilon_1+\sigma(\{1\})\big)\\ &+ \big(\mu_{1,\{2\},2,\{1\}} + \mu_{1,\{2\},2,\{1,3\}}\big)\big(1-\epsilon_1+\sigma(\{1\})\big) \ge q_1,\end{aligned} \tag{51}$$

Cut 5 implies that

$$\begin{aligned}&\mu_{1,\varnothing}(1-\epsilon_1) + \big(\mu_{1,\{2,3\},2,\{1\}} + \mu_{1,\{2,3\},3,\{1\}}\\ &+ \mu_{1,\{2,3\},2,\{1,3\}} + \mu_{1,\{2,3\},3,\{1,2\}}\big)(1-\epsilon_1)\\ &+ \big(\mu_{1,\{3\},3,\{1\}} + \mu_{1,\{3\},3,\{1,2\}}\big)(1-\epsilon_1)\\ &+ \big(\mu_{1,\{2\},2,\{1\}} + \mu_{1,\{2\},2,\{1,3\}}\big)(1-\epsilon_1) \ge q_1.\end{aligned} \tag{52}$$

Similar process for users 2 and 3, we get (53) $\sim$ (56),

$$\mu_{i,\varnothing}(1-\epsilon_i) + \sum_{j\in[3]\backslash i}\mu_{i,\varnothing}\sigma([3]\backslash j) + \mu_{i,\varnothing}\sigma(\{i\}) \ge q_i, \tag{53}$$

$$\begin{aligned}&\mu_{i,\varnothing}(1-\epsilon_i) + \mu_{i,\varnothing}\sigma([3]\backslash j) + \mu_{i,\varnothing}\sigma(\{i\})\\ &+ \big(\mu_{i,\{k\},k,\{i\}} + \mu_{i,\{k\},k,[3]\backslash k}\big)\big(1-\epsilon_i+\sigma(\{i\})\big) \ge q_i\\ &j,k \ne i,\ j,k\in[3],\ j\ne k,\end{aligned} \tag{54}$$

$$\begin{aligned}&\mu_{i,\varnothing}(1-\epsilon_i) + \mu_{i,\varnothing}\sigma(\{i\})\\ &+ \sum_{j\in[3]\backslash i}\big(\mu_{i,\{j\},j,\{i\}} + \mu_{i,\{j\},j,[3]\backslash j}\big)\big(1-\epsilon_i+\sigma(\{i\})\big) \ge q_i,\end{aligned} \tag{55}$$

$$\begin{aligned}&\mu_{i,\varnothing}(1-\epsilon_i) + \sum_{j\in[3]\backslash i}\big(\mu_{i,[3]\backslash i,j,\{i\}} + \mu_{i,[3]\backslash i,j,[3]\backslash j}\big)(1-\epsilon_i)\\ &+ \sum_{j\in[3]\backslash i}\big(\mu_{i,\{j\},j,\{i\}} + \mu_{i,\{j\},j,[3]\backslash j}\big)(1-\epsilon_i) \ge q_i.\end{aligned} \tag{56}$$

Thus, $\{\mu_{i,\varnothing}\}$ and $\{\mu_{\tau_1,\mathcal{S}_{\tau_1},\cdots,\tau_l,\mathcal{S}_{\tau_l}}\}$ satisfying (40), (53) $\sim$ (56) give the inequality (47). 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 (40), (53) $\sim$ (56) are satisfied. Substituting $\mu_{i,\varnothing} = q_i$ into (47), we get

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

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

$$\text{EAoI} \leq \frac{1}{3}\sum_{i=1}^{3}\big(\frac{\alpha_i}{\theta_i} + \frac{\alpha_i}{q_i(1-\epsilon_i)}\big) + \lambda.$$

APPENDIX E
PROOF OF LEMMA 2

Step 1. If x_1 is delivered to users $1,2$, then from Definition 1, the AoI reduction of the users is

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

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

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

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

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

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

and

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

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

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

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

Similar with the proof of Theorem 1,

$$\begin{aligned}\Theta_1(k) = &-\sum_{i=1}^{M}\Big(\alpha_i\sum_{\mathcal{S}} t_{i,\mathcal{S}}(k)\mathbb{E}\big[d_i(k)\big]\delta_{i,\mathcal{S}}(k) \\ &+ \sum_{\mathcal{I}:i\in\mathcal{I}} t_{\mathcal{I},u}^{(i)}(k)\mathbb{E}\big[d_i(k)\big]\delta_{\mathcal{I},u}^{(i)}(k)\Big) + \sum_{i=1}^{M}\alpha_i.\end{aligned} \tag{57}$$

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

$$I_i(m) = T_i(m+1) - T_i(m) \tag{58}$$

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 $[T_i(m), T_i(m+1))$:

$$\Gamma_i = \sum_{k=T_i(m)}^{T_i(m)+I_i(m)-1} h_i(k) \tag{59}$$

$$= \frac{1}{2}I_i^2(m) - \frac{1}{2}I_i(m) + D_i(m-1)I_i(m). \tag{60}$$

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

$$J^{\pi}(M) = \lim_{K\to\infty}\Big[\frac{1}{M}\sum_{i=1}^{M}\alpha_i\frac{1}{K}\sum_{m=1}^{N_i(K)}\Gamma_i(m)\Big]. \tag{61}$$

Using this formulation, we next lower bound EAoI. Let R_i denote the capacity outer bound of user i, defined in (27) and (28). Note that in the limit of large K, $\frac{N_i(K)}{K}$ is the rate/throughput of user i under policy π, denoted by R_i^{π}.

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 $\Gamma_i(m)$:

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

Since $D_i(m) \geq 1$ for all $1 \leq m \leq N_i(K)$, we can lower bound (61) by substituting $D_i(m-1) = 1$. Using similar steps as [14, Eqns. (12) - (17)],

$$\lim_{K\to\infty}\mathbb{E}\big[J_K^{\pi}\big] \geq \lim_{K\to\infty}\mathbb{E}\big[\frac{1}{2M}\sum_{i=1}^{M}\frac{\alpha_i K}{N_i(K)} + \frac{\sum_{i=1}^{M}\alpha_i}{2M}\big]. \tag{62}$$

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

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

and thus

$$\lim_{K\to\infty} \mathbb{E}\big[\sum_{i=1}^{M} \frac{\alpha_i K}{N_i(K)}\big] \geq M^2 / \sum_{i=1}^{M} \frac{R_i}{\alpha_i}. \tag{63}$$

Inserting this back into (62), 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. $\big\{X_i(m)\big\}_m$ is a geometric i.i.d sequence. Under the assumption of instantaneous recovery, $I_i(m) = X_i(m)$. It hence follows from (60) that

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

Thus, similar with [39], the time-average AoI of user i is

$$\begin{aligned}\mathbb{E}[h_i] &= \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} h_i(k) \\ &= \frac{\mathbb{E}[X_i^2(m)]}{2\mathbb{E}[X_i(m)]} + \frac{1}{2} = \frac{2-\theta_i}{2\theta_i} + \frac{1}{2}\end{aligned}$$

and

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