



# Content Delivery in Erasure Broadcast Channels With Cache and Feedback

Asma Ghorbel, Mari Kobayashi, Sheng Yang

## ► To cite this version:

Asma Ghorbel, Mari Kobayashi, Sheng Yang. Content Delivery in Erasure Broadcast Channels With Cache and Feedback. IEEE Transactions on Information Theory, 2016, 62 (11), pp.6407-6422. 10.1109/TIT.2016.2605123 . hal-01433728

**HAL Id: hal-01433728**

**<https://centralesupelec.hal.science/hal-01433728>**

Submitted on 28 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Content Delivery in Erasure Broadcast Channels with Cache and Feedback

Asma Ghorbel, Mari Kobayashi, and Sheng Yang

LSS, CentraleSupélec

Gif-sur-Yvette, France

{asma.ghorbel, mari.kobayashi, sheng.yang}@centralesupelec.fr

## Abstract

We study a content delivery problem in a  $K$ -user erasure broadcast channel such that a content providing server wishes to deliver requested files to users, each equipped with a cache of a finite memory. Assuming that the transmitter has state feedback and user caches can be filled during off-peak hours reliably by the decentralized content placement, we characterize the achievable rate region as a function of the memory sizes and the erasure probabilities. The proposed delivery scheme, based on the broadcasting scheme by Wang and Gatzianas et al., exploits the receiver side information established during the placement phase. Our results can be extended to centralized content placement as well as multi-antenna broadcast channels with state feedback.

## I. INTRODUCTION

Today's exponentially growing mobile data traffic is mainly due to video applications such as content-based video streaming. The skewness of the video traffic together with the ever-growing cheap on-board storage memory suggests that the quality of experience can be boosted by caching popular contents at (or close to) the end-users in wireless networks. A number of recent works have studied such concept under different models and assumptions (see [1], [2] and references therein). Most of existing works assume that caching is performed in two phases: *placement phase* to prefetch users' caches under their memory constraints (typically during off-peak hours) prior to the actual demands; *delivery phase* to transmit codewords such that each user, based on the received signal and the contents of its cache, is able to decode the requested file. In this work, we study the delivery phase based on a coded caching model where a server is connected to many users, each equipped with a cache of finite memory [1]. By carefully choosing the sub-files to be distributed across users, coded caching exploits opportunistic multicasting such that a common signal is simultaneously useful for all users even with distinct file requests. A number of extensions of coded caching have been developed (see e.g. [1, Section VIII]). These include the decentralized content placement [5], online coded caching [7], non-uniform popularities [4], [6], more general networks such as device-to-device (D2D) enabled network [3], hierarchical networks [22], heterogeneous networks [10], as well as the performance analysis in different regimes [19], [20]. Further, very recent works have attempted to relax the unrealistic assumption

of a perfect shared link by replacing it by wireless channels (e.g. [8], [9], [18], [25], [26]). If wireless channels are used only to multicast a common signal, naturally the performance of coded caching (delivery phase) is limited by the user in the worst condition of fading channels as observed in [8]. This is due to the information theoretic limit, that is, the multicasting rate is determined by the worst user [24, Chapter 7.2]. However, if the underlying wireless channels enjoy some degrees of freedom to convey simultaneously both private messages and common messages, the delivery phase of coded caching can be further enhanced. In the context of multi-antenna broadcast channel and erasure broadcast channel, the potential gain of coded caching in the presence of channel state feedback has been demonstrated [18], [25], [26]. The key observation behind [18], [26] is that opportunistic multicasting can be performed based on either the receiver side information established during the placement phase or the channel state information acquired via feedback.

In this work, we model the bottleneck link between the server with  $N$  files and  $K$  users equipped with a cache of a finite memory as an erasure broadcast channel (EBC). The simple EBC captures the essential features of wireless channels such as random failure or disconnection of any server-user link that a packet transmission may experience during high-traffic hours, i.e. during the delivery phase. In this work, we consider a memoryless EBC in which erasure is independent across users with probabilities  $\{\delta_k\}$  and each user  $k$  can cache up to  $M_k$  files. Moreover, the server is assumed to acquire the channel states causally via feedback sent by the users. Assuming that users fill the caches randomly and independently according to the decentralized content placement scheme as proposed in [5], we study the achievable rate region of the EBC with cache and state feedback. Our main contribution is the characterization of the rate region in the cache-enabled EBC with state feedback for the case of the decentralized content placement (Theorem 1). The converse proof builds on the genie-aided bounds exploiting two key lemmas, i.e. a generalized form of the entropy inequalities (Lemma 1) as well as the reduced entropy of messages in the presence of receiver side information (Lemma 2). For the achievability, we present a multi-phase delivery scheme extending the algorithm proposed independently by Wang [15] and by Gatzianas et al. [16] to the case with receiver side information and prove that it achieves the optimal rate region for special cases of interest. We provide, as a byproduct of the achievability proof for the symmetric network, an alternative proof for the sum capacity of the EBC with state feedback and without cache. More specifically, we characterize the order- $j$  capacity defined as the maximum transmission rate of a message intended to  $j$  users and express the sum capacity in a convenient manner along the line of [14]. This allows us to characterize the rate region of the symmetric cache-enabled EBC with state feedback easily, since as such all we need is to incorporate the packets generated during the placement phase [18]. However, such proof exploits the specific structure of the rate region of symmetric networks, and unfortunately cannot be applied to a general network setting considered here. Our current work provides a non-trivial extension of [18] to such networks. Furthermore, we show that our results can be extended in a straightforward manner to the centralized content placement [1] as well as the multi-antenna broadcast channel with state feedback. Finally, we provide some numerical examples to quantify the benefit of state feedback, the relative merit of the centralized caching to the decentralized counterpart, as well as the gain due to the optimization of memory sizes, as a function of other system parameters.

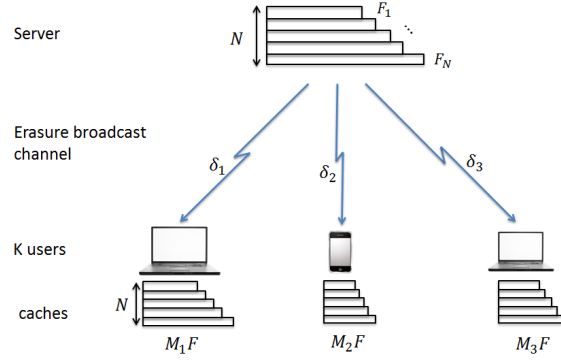


Fig. 1. A cached-enabled erasure broadcast channel with  $K = 3$ .

The rest of the paper is organized as follows. In section II, we describe the system model together with some definitions and then summarize the main results. Section III gives the converse proof of the achievable rate region of the cache-enabled EBC with state feedback. After a high-level description of the well-known algorithm by Wang and Gatzianas et al. in section IV, section V presents our proposed delivery scheme and provides the achievability proof for some special cases of interest. Section VI provides the extensions of the previous results and section VII shows some numerical examples.

Throughout the paper, we use the following notational conventions. The superscript notation  $X^n$  represents a sequence  $(X_1, \dots, X_n)$  of variables.  $X_{\mathcal{J}}$  is used to denote the set of variables  $\{X_i\}_{i \in \mathcal{J}}$ . The entropy of  $X$  is denoted by  $H(X)$ . We let  $[k] = \{1, \dots, k\}$ . We let  $\epsilon_n$  denote a constant which vanishes as  $n \rightarrow \infty$ , i.e.  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ .

## II. SYSTEM MODEL AND MAIN RESULTS

### A. System model and definitions

We consider a cache-enabled network depicted in Fig. 1 where a server is connected to  $K$  users through an erasure broadcast channel (EBC). The server has an access to  $N$  files  $W_1, \dots, W_N$  where the  $i$ -th file  $W_i$  consists of  $F_i$  packets of  $L$  bits each ( $F_i L$  bits). Each user  $k$  has a cache memory  $Z_k$  of  $M_k F$  packets for  $M_k \in [0, N]$ , where  $F \triangleq \frac{1}{N} \sum_{i=1}^N F_i$  is the average size of the files. Under such a setting, consider a discrete time communication system where a packet is sent in each slot over the  $K$ -user EBC. The channel input  $X_k \in \mathbb{F}_q$  belongs to the input alphabet of size  $L \triangleq \log_2 q$  bits. The erasure is assumed to be memoryless and independently distributed across users so that in a given slot we have

$$\Pr(Y_1, Y_2, \dots, Y_K | X) = \prod_{k=1}^K \Pr(Y_k | X) \quad (1)$$

$$\Pr(Y_k | X) = \begin{cases} 1 - \delta_k, & Y_k = X, \\ \delta_k, & Y_k = E \end{cases} \quad (2)$$

where  $Y_k$  denotes the channel output of receiver  $k$ ,  $E$  stands for an erased output,  $\delta_k$  denotes the erasure probability of user  $k$ . We let  $S_l \in \mathcal{S} = 2^{\{1, \dots, K\}}$  denote the state of the channel in slot  $l$  and indicate the set of users who

received correctly the packet. We assume that all the receivers know instantaneously  $S_l$ , and that through feedback the transmitter only knows the past states  $S^{l-1}$  during slot  $l$ .

The caching network is operated in two phases: the placement phase and the delivery phase. In the content placement phase, the server fills the caches of all users,  $Z_1, \dots, Z_K$ , up to the memory constraint. As in most works in the literature, we assume that the placement phase incurs no error and no cost, since it takes place usually during off-peak traffic hours. Once each user  $k$  makes a request  $d_k$ , the server sends the codewords so that each user can decode its requested file as a function of its cache contents and received signals during the delivery phase. We provide a more formal definition below. A  $(M_1, \dots, M_K, F_{d_1}, \dots, F_{d_K}, n)$  caching scheme consists of the following components.

- $N$  message files  $W_1, \dots, W_N$  independently and uniformly distributed over  $\mathcal{W}_1 \times \dots \times \mathcal{W}_N$  with  $\mathcal{W}_i = \mathbb{F}_q^{F_i}$  for all  $i$ .
- $K$  caching functions defined by  $\phi_k : \mathbb{F}_q^{\sum_{i=1}^N F_i} \rightarrow \mathbb{F}_q^{F_{M_k}}$  that map the files  $W_1, \dots, W_N$  into user  $k$ 's cache contents

$$Z_k = \phi_k(W_1, \dots, W_N), \quad k \in [K]. \quad (3)$$

- A sequence of encoding functions which transmit at slot  $l$  a symbol  $X_l = f_l(W_{d_1}, \dots, W_{d_K}, S^{l-1}) \in \mathbb{F}_q$ , based on the requested files and the state feedback up to slot  $l-1$  for  $l = 1, \dots, n$ , where  $W_{d_k}$  denotes the message file requested by user  $k$  for  $d_k \in \{1, \dots, N\}$ .
- $K$  decoding functions defined by  $\psi_k : \mathbb{F}_q^n \times \mathbb{F}_q^{F_{M_k}} \times \mathcal{S}^n \rightarrow \mathbb{F}_q^{F_{d_k}}$ ,  $k \in [K]$ , that decode the file  $\hat{W}_{d_k} = \psi_k(Y_k^n, Z_k, S^n)$  as a function of the received signals  $Y_k^n$ , the cache content  $Z_k$ , as well as the state information  $S^n$ .

A rate tuple  $(R_1, \dots, R_K)$  is said to be achievable if, for every  $\epsilon > 0$ , there exists a  $(M_1, \dots, M_K, F_{d_1}, \dots, F_{d_K}, n)$  caching strategy that satisfies the reliability condition

$$\max_{(d_1, \dots, d_K) \in \{1, \dots, N\}^K} \max_k \Pr(\psi_k(Y_k^n, Z_k, S^n) \neq W_{d_k}) < \epsilon$$

as well as the rate condition

$$R_k < \frac{F_{d_k}}{n} \quad \forall k \in [K]. \quad (4)$$

Throughout the paper, we express for brevity the entropy and the rate in terms of packets in order to avoid the constant factor  $L = \log_2 q$ .

### B. Decentralized content placement

We mainly focus on the decentralized content placement proposed in [5]. Under the memory constraint of  $M_k F$  packets, each user  $k$  independently caches a subset of  $p_k F_i$  packets of file  $i$ , chosen uniformly at random for  $i = 1, \dots, N$ , where  $p_k = \frac{M_k}{N}$ . By letting  $\mathcal{L}_{\mathcal{J}}(W_i)$  denote the sub-file of  $W_i$  stored exclusively by the users in  $\mathcal{J}$ , the cache memory of user  $k$  after the decentralized placement is given by

$$Z_k = \{\mathcal{L}_{\mathcal{J}}(W_i) : \mathcal{J} \subseteq [K], \mathcal{J} \ni k, i = 1, \dots, N\}. \quad (5)$$

The size of each sub-file is given by

$$|\mathcal{L}_{\mathcal{J}}(W_i)| = \prod_{j \in \mathcal{J}} p_j \prod_{j \in [K] \setminus \mathcal{J}} (1 - p_j) F_i + \epsilon_{F_i} \quad (6)$$

as  $F_i \rightarrow \infty$ . It can be easily verified that the memory constraint of each user is fulfilled, namely,

$$|Z_k| = \sum_{i=1}^N \sum_{\mathcal{J}: k \in \mathcal{J}} |\mathcal{L}_{\mathcal{J}}(W_i)| = \sum_{i=1}^N (F_i p_k + \epsilon_{F_i}) = M_k F + \sum_{i=1}^N \epsilon_{F_i} \quad (7)$$

as  $F_i \rightarrow \infty$  for all  $i$ . Throughout the paper, we assume that  $F \rightarrow \infty$  and meanwhile  $\frac{F_i}{F}$  converges to some constant  $\tilde{F}_i > 0$ . Thus, we identify all  $\epsilon_{F_i}$  with a single  $\epsilon_F$ .

To illustrate the placement strategy, let us consider an example of  $K = 3$  users. After the placement phase, each file will be partitioned into 8 sub-files:

$$W_i = \{\mathcal{L}_{\emptyset}(W_i), \mathcal{L}_1(W_i), \mathcal{L}_2(W_i), \mathcal{L}_3(W_i), \mathcal{L}_{12}(W_i), \mathcal{L}_{13}(W_i), \mathcal{L}_{23}(W_i), \mathcal{L}_{123}(W_i)\}. \quad (8)$$

Obviously, the sub-files received by the destination, e.g.  $\mathcal{L}_1(W_1), \mathcal{L}_{12}(W_1), \mathcal{L}_{13}(W_1), \mathcal{L}_{123}(W_1)$  for user 1 requesting  $W_1$ , need not be transmitted in the delivery phase.

### C. Main results

In order to present the main results, we specify two special cases.

**Definition 1.** *The cache-enabled EBC (or the network) is symmetric if the erasure probabilities as well as the memory sizes are the same for all users, i.e.  $\delta_1 = \dots = \delta_K = \delta$ ,  $M_1 = \dots = M_K = M$ ,  $p_1 = \dots = p_K = p$ .*

**Definition 2.** *The rate vector is said to be one-sided fair in the cache-enabled EBC if  $\delta_k \geq \delta_j$  and for  $k \neq j$  implies*

$$\frac{1 - p_k}{p_k} R_k \geq \frac{1 - p_j}{p_j} R_j, \quad \text{and} \quad (9)$$

$$\delta_k R_k \geq \delta_j R_j. \quad (10)$$

For the special case where  $p_k = 0, \forall k \in [K]$ , it is reduced to  $\delta_k R_k \geq \delta_j R_j$  which coincides with the one-sided fairness originally defined in [15].

Focusing on the case of most interest with  $N \geq K$  and  $K$  distinct demands, we present the following main results of this work.

**Theorem 1.** *For  $K \leq 3$ , or for the symmetric network with  $K \geq 3$ , or for the one-sided fair rate vector with  $K > 3$ , the achievable rate region of the cached-enabled EBC with the state feedback under the decentralized content placement is given by*

$$\sum_{k=1}^K \frac{\prod_{j=1}^k (1 - p_{\pi_j})}{1 - \prod_{j=1}^k \delta_{\pi_j}} R_{\pi_k} \leq 1 \quad (11)$$

for any permutation  $\pi$  of  $\{1, \dots, K\}$ .

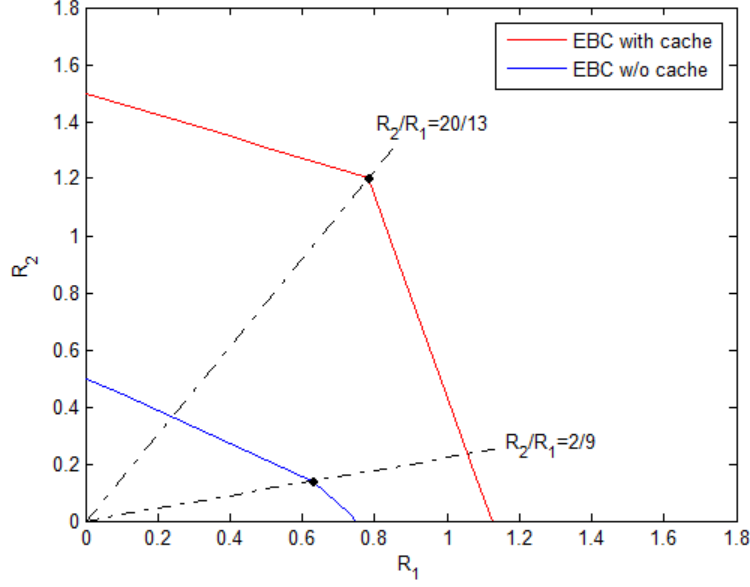


Fig. 2. A two-user rate region with  $(p_1, p_2) = (\frac{1}{3}, \frac{2}{3})$ ,  $(\delta_1, \delta_2) = (\frac{1}{4}, \frac{1}{2})$ .

The above region has a polyhedron structure determined by  $K!$  inequalities in general. It should be remarked that Theorem 1 covers some existing results. For the symmetric network, the above region simplifies to [18]

$$\sum_{k=1}^K \frac{(1-p)^k}{1-\delta^k} R_{\pi_k} \leq 1, \quad \forall \pi. \quad (12)$$

For the case without cache memory, i.e.  $p_k = 0$  for all  $k$ , Theorem 1 boils down to the capacity region of the EBC with state feedback [15], [16] given by

$$\sum_{k=1}^K \frac{1}{1 - \prod_{j=1}^k \delta_{\pi_j}} R_{\pi_k} \leq 1, \quad \forall \pi \quad (13)$$

which is achievable for  $K \leq 3$  or the symmetric network or the one-sided fair rate vector where  $\delta_k \geq \delta_j$  implies  $\delta_k R_k \geq \delta_j R_j$  for any  $k \neq j$ . Comparing (11) and (13), we immediately see that the presence of cache memories decreases the weights in the weighted rate sum and thus enlarges the rate region. In order to gain some further insight, Fig. 2 illustrates a toy example of two users with  $(p_1, p_2) = (\frac{1}{3}, \frac{2}{3})$  and  $(\delta_1, \delta_2) = (\frac{1}{4}, \frac{1}{2})$ . According to Theorem 1, the rate region is given by

$$\begin{aligned} \frac{8}{9} R_1 + \frac{16}{63} R_2 &\leq 1 \\ \frac{16}{63} R_1 + \frac{2}{3} R_2 &\leq 1 \end{aligned} \quad (14)$$

which is characterized by three vertices  $(\frac{9}{8}, 0)$ ,  $(0.78, 1.20)$ , and  $(0, \frac{63}{16})$ . The vertex  $(0.78, 1.20)$ , achieving the sum rate of 1.98, corresponds to the case when the requested files satisfy the ratio  $F_{d_2}/F_{d_1} = 20/13$ . On the other

hand, the region of the EBC without cache is given by

$$\begin{aligned}\frac{4}{3}R_1 + \frac{8}{7}R_2 &\leq 1 \\ \frac{8}{7}R_1 + 2R_2 &\leq 1\end{aligned}\tag{15}$$

which is characterized by three vertices  $(\frac{3}{4}, 0)$ ,  $(0.63, 0.14)$ ,  $(0, \frac{1}{2})$ . The sum capacity of 0.77 is achievable for the ratio  $R_2/R_1 = 2/9$ . The gain due to the cache is highlighted even in this toy example.

Theorem 1 yields the following corollary.

**Corollary 1.** *For  $K \leq 3$ , or for the symmetric network with  $K \geq 3$ , or for the one-sided fair rate vector with  $K > 3$ , the transmission length to deliver requested files to users in the cached-enabled EBC under the decentralized content placement is given by*

$$T_{\text{tot}} = \max_{\pi} \left\{ \sum_{k=1}^K \frac{\prod_{j=1}^k (1 - p_{\pi_j})}{1 - \prod_{j=1}^k \delta_{\pi_j}} F_{d_{\pi_k}} \right\} + \epsilon_F,\tag{16}$$

as  $F \rightarrow \infty$ .

The corollary covers some existing results in the literature. For the symmetric network with files of equal size ( $F_i = F, \forall i$ ), the transmission length simplifies to

$$T_{\text{tot}} = \sum_{k=1}^K \frac{(1-p)^k}{1-\delta^k} F + \epsilon_F,\tag{17}$$

as  $F \rightarrow \infty$  [18]. For the case with files of equal size and without erasure, the transmission length in Corollary 1 normalized by  $F$  coincides with the “rate-memory tradeoff”<sup>1</sup> under the decentralized content placement for asymmetric memory sizes [17] given by

$$\frac{T_{\text{tot}}}{F} = \sum_{k=1}^K \left[ \prod_{j=1}^k (1 - p_j) \right],\tag{18}$$

where the maximum over all permutations is chosen to be identity by assuming  $p_1 \geq \dots \geq p_K$ . If additionally we restrict ourselves to the case with caches of equal size, we recover the rate-memory tradeoff given in [5]

$$\frac{T_{\text{tot}}}{F} = \frac{N}{M} \left( 1 - \frac{M}{N} \right) \left\{ 1 - \left( 1 - \frac{M}{N} \right)^K \right\}.\tag{19}$$

In fact, the above expression readily follows by applying the geometric series to the RHS of (18).

### III. CONVERSE

In this section, we prove the converse of Theorem 1. First we provide two useful lemmas. The first one is a generalized form of the entropy inequality, while the second one is a simple relation of the message entropy in the presence of receiver side information. Although the former has been proved in [18], we restate it for the sake of completeness.

<sup>1</sup>In [5] and all follow-up works, the “rate” is defined as the number of files to deliver over the shared link, which corresponds to our  $T_{\text{tot}}$  here.



**Lemma 1.** [12, Lemma 5] For the erasure broadcast channel, if  $U$  is such that  $X_l \leftrightarrow UY_{\mathcal{J}}^{l-1}S^{l-1} \leftrightarrow (S_{l+1}, \dots, S_n)$ ,  $\forall \mathcal{J}$ ,

$$\frac{1}{1 - \prod_{j \in \mathcal{J}} \delta_j} H(Y_{\mathcal{J}}^n | U, S^n) \leq \frac{1}{1 - \prod_{j \in \mathcal{J}} \delta_j} H(Y_{\mathcal{J}}^n | U, S^n) \quad (20)$$

for any sets  $\mathcal{J}, \mathcal{J}$  such that  $\mathcal{J} \subseteq \mathcal{J} \subseteq \{1, \dots, K\}$ .

*Proof:* We have, for  $\mathcal{J} \subseteq \mathcal{J}$ ,

$$H(Y_{\mathcal{J}}^n | U, S^n) \quad (21)$$

$$= \sum_{l=1}^n H(Y_{\mathcal{J},l} | Y_{\mathcal{J}}^{l-1}, U, S^n) \quad (22)$$

$$= \sum_{l=1}^n H(Y_{\mathcal{J},l} | Y_{\mathcal{J}}^{l-1}, U, S^{l-1}, S_l) \quad (23)$$

$$= \sum_{l=1}^n \Pr\{S_l \cap \mathcal{J} \neq \emptyset\} H(X_l | Y_{\mathcal{J}}^{l-1}, U, S^{l-1}, S_l \cap \mathcal{J} \neq \emptyset) \quad (24)$$

$$= \sum_{l=1}^n (1 - \prod_{i \in \mathcal{J}} \delta_i) H(X_l | Y_{\mathcal{J}}^{l-1}, U, S^{l-1}) \quad (25)$$

$$\leq (1 - \prod_{i \in \mathcal{J}} \delta_i) \sum_{l=1}^n H(X_l | Y_{\mathcal{J}}^{l-1}, U, S^{l-1}) \quad (26)$$

where the first equality is from the chain rule; the second equality is because the current input does not depend on future states conditioned on the past outputs/states and  $U$ ; the third one holds since  $Y_{\mathcal{J},l}$  is deterministic and has entropy 0 when all outputs in  $\mathcal{J}$  are erased ( $S_l \cap \mathcal{J} = \emptyset$ ); the fourth equality is from the independence between  $X_l$  and  $S_l$ ; and we get the last inequality by removing the terms  $Y_{\mathcal{J}}^{l-1}$  in the condition of the entropy. Following the same steps, we have

$$H(Y_{\mathcal{J}}^n | U, S^n) = (1 - \prod_{i \in \mathcal{J}} \delta_i) \sum_{l=1}^n H(X_l | Y_{\mathcal{J}}^{l-1}, U, S^{l-1}), \quad (27)$$

from which and (26), we obtain (20). ■

**Lemma 2.** Under the decentralized content placement [5], the following inequality hold for any  $i$  and  $\mathcal{K} \subseteq [K]$

$$H(W_i | \{Z_k\}_{k \in \mathcal{K}}) \geq \prod_{k \in \mathcal{K}} (1 - p_k) H(W_i).$$

*Proof:* Under the decentralized content placement, we have

$$H(W_i | \{Z_k\}_{k \in \mathcal{K}}) = H(W_i | \{\mathcal{L}_{\mathcal{J}}(W_i)\}_{\mathcal{J} \cap \mathcal{K} \neq \emptyset, l=1, \dots, N}) \quad (28)$$

$$= H(W_i | \{\mathcal{L}_{\mathcal{J}}(W_i)\}_{\mathcal{J} \cap \mathcal{K} \neq \emptyset}) \quad (29)$$

$$= H(\{\mathcal{L}_{\mathcal{J}}(W_i)\}_{\mathcal{J} \cap \mathcal{K} = \emptyset}) \quad (30)$$

$$= \sum_{\mathcal{J}: \mathcal{J} \subseteq [K] \setminus \mathcal{K}} H(\mathcal{L}_{\mathcal{J}}(W_i)) \quad (31)$$

$$\geq \sum_{\mathcal{J}: \mathcal{J} \subseteq [K] \setminus \mathcal{K}} H(\mathcal{L}_{\mathcal{J}}(W_i) | \mathcal{L}_{\mathcal{J}}) \quad (32)$$

where the first equality follows from (5); the second equality follows due to the independence between messages  $W_1, \dots, W_N$ ; the third equality follows by identifying the unknown parts of  $W_i$  given the cache memories of  $\mathcal{K}$  and using the independence of all sub-files; (31) is again from the independence of the sub-files. Note that  $\mathcal{L}_{\mathcal{J}}$  is a random variable indicating which subset of packets of file  $W_i$  are shared by the users in  $\mathcal{J}$ . The size of the random subset  $|\mathcal{L}_{\mathcal{J}}|$  follows thus the binomial distribution  $B(H(W_i), \prod_{j \in \mathcal{J}} p_j \prod_{k \in [K] \setminus \mathcal{J}} (1 - p_k))$ . It is readily shown that  $H(\mathcal{L}_{\mathcal{J}}(W_i) | \mathcal{L}_{\mathcal{J}}) = \mathbb{E}\{|\mathcal{L}_{\mathcal{J}}|\}$ . This implies that

$$H(W_i | \{Z_k\}_{k \in \mathcal{K}}) \geq \sum_{\mathcal{J}: \mathcal{J} \subseteq [K] \setminus \mathcal{K}} \prod_{j \in \mathcal{J}} p_j \prod_{k \in [K] \setminus \mathcal{J}} (1 - p_k) H(W_i) \quad (33)$$

$$= \prod_{k \in \mathcal{K}} (1 - p_k) \sum_{\mathcal{J}: \mathcal{J} \subseteq [K] \setminus \mathcal{K}} \prod_{j \in \mathcal{J}} p_j \prod_{k \in [K] \setminus \mathcal{K} \setminus \mathcal{J}} (1 - p_k) H(W_i) \quad (34)$$

$$= \prod_{k \in \mathcal{K}} (1 - p_k) H(W_i) \quad (35)$$

where the last inequality is obtained from the basic property that we have  $\sum_{\mathcal{J} \subseteq \mathcal{M}} \prod_{j \in \mathcal{J}} p_j \prod_{k \in \mathcal{M} \setminus \mathcal{J}} (1 - p_k) = 1$  for a subset  $\mathcal{M} = [K] \setminus \mathcal{K}$ . ■

We apply genie-aided bounds to create a degraded erasure broadcast channel by providing the messages, the channel outputs, as well as the receiver side information (contents of cache memories) to the enhanced receivers. Without loss of generality, we focus on the case without permutation and the demand  $(d_1, \dots, d_K) = (1, \dots, K)$ .

$$n \prod_{j=1}^k (1 - p_j) R_k = \prod_{j=1}^k (1 - p_j) H(W_k) \quad (36)$$

$$\leq H(W_k | Z^k S^n) \quad (37)$$

$$\leq I(W_k; Y_{[k]}^n | Z^k S^n) + n\epsilon'_{n,k} \quad (38)$$

$$\leq I(W_k; Y_{[k]}^n, W^{k-1} | Z^k S^n) + n\epsilon'_{n,k} \quad (39)$$

$$= I(W_k; Y_{[k]}^n | W^{k-1} Z^k S^n) + n\epsilon'_{n,k} \quad (40)$$

where the second inequality is by applying Lemma 2 and noting that  $S^n$  is independent of others; (38) is from Fano's inequality; the last equality is from  $I(W_k; W^{k-1} | Z^k S^n) = 0$  since the caches  $Z^k$  only store disjoint pieces of individual files by the decentralized content placement [5]. Putting all the rate constraints together, and defining

$\epsilon_{n,k} \triangleq \epsilon'_{n,k} / \prod_{j=1}^k (1 - p_j)$ , we have

$$n(1 - p_1)(R_1 - \epsilon_{n,1}) \leq H(Y_1^n | Z_1 S^n) - H(Y_1^n | W_1 Z_1 S^n)$$

$\vdots$

$$n \prod_{j=1}^K (1 - p_j)(R_K - \epsilon_{n,K}) \leq H(Y_{[K]}^n | W^{K-1} Z^K S^n) - H(Y_{[K]}^n | W^K Z^K S^n). \quad (41)$$

We now sum up the above inequalities with different weights, and apply  $K - 1$  times Lemma 1, namely, for  $k = 1, \dots, K - 1$ ,

$$\frac{H(Y_{[k+1]}^n | W^k Z^{k+1} S^n)}{1 - \prod_{j \in [k+1]} \delta_j} \leq \frac{H(Y_{[k+1]}^n | W^k Z^k S^n)}{1 - \prod_{j \in [k+1]} \delta_j} \quad (42)$$

$$\leq \frac{H(Y_{[k]}^n | W^k Z^k S^n)}{1 - \prod_{j \in [k]} \delta_j}, \quad (43)$$

where the first inequality follows because removing conditioning increases entropy. Finally, we have

$$\begin{aligned} \sum_{k=1}^K \frac{\prod_{j \in [k]} (1 - p_j)}{1 - \prod_{j \in [k]} \delta_j} (R_k - \epsilon_n) \\ \leq \frac{H(Y_1^n | Z_1 S^n)}{n(1 - \delta_1)} - \frac{H(Y_{[K]}^n | W^K Z^K S^n)}{n(1 - \prod_{j \in [K]} \delta_j)} \end{aligned} \quad (44)$$

$$\leq \frac{H(Y_1^n)}{n(1 - \delta_1)} \leq 1 \quad (45)$$

which establishes the converse proof.

#### IV. BROADCASTING WITHOUT RECEIVER SIDE INFORMATION

In this section, we first revisit the algorithm proposed in [15], [16] achieving the capacity region of the EBC with state feedback for some cases of interest, as an important building block of our proposed scheme. Then, we provide an alternative achievability proof for the symmetric channel with uniform erasure probabilities across users.

##### A. Revisiting the algorithm by Wang and Gatzianas et al.

We recall the capacity region of the EBC with state feedback as below.

**Theorem 2** ([15], [16]). *For  $K \leq 3$ , or for the symmetric channel with  $K \geq 3$ , or for the one-sided fair rate vector<sup>2</sup> with  $K > 3$ , the capacity region of the erasure broadcast channel with state feedback is given by*

$$\sum_{k=1}^K \frac{1}{1 - \prod_{j=1}^k \delta_{\pi_j}} R_{\pi_k} \leq 1, \quad \forall \pi. \quad (46)$$

We provide a high-level description of the broadcasting scheme [15], [16] which is optimal under the special cases as specified in the above theorem. We recall that the number of private packets  $\{F_k\}$  is assumed to be arbitrarily large so that the length of each phase becomes deterministic. Thus, we drop the  $\epsilon_F$  term wherever confusion is not

<sup>2</sup> $\delta_k \geq \delta_j$  implies  $\delta_k R_k \geq \delta_j R_j$  for any  $k \neq j$ .

probable. The broadcasting algorithm has two main roles: 1) broadcast new information packets and 2) multicast side information or overheard packets based on state feedback. Therefore, we can call phase 1 *broadcasting phase* and phases 2 to  $K$  *multicasting phase*. Phase  $j$  consists of  $\binom{K}{j}$  sub-phases in each of which the transmitter sends packets intended to a subset of users  $\mathcal{J}$  for  $|\mathcal{J}| = j$ . Similarly to the receiver side information obtained after the placement phase, we let  $\mathcal{L}_{\mathcal{J}}(V_{\mathcal{K}})$  denote the part of packet  $V_{\mathcal{K}}$  received by users in  $\mathcal{J}$  and erased at users in  $[K] \setminus \mathcal{J}$ .

Here is a high-level description of the broadcasting algorithm:

- 1) Broadcasting phase (phase 1): send each message  $V_k = W_k$  of  $F_k$  packets sequentially for  $k = 1, \dots, K$ . This phase generates overheard symbols  $\{\mathcal{L}_{\mathcal{J}}(V_k)\}$  to be transmitted via linear combination in multicasting phase, where  $\mathcal{J} \subseteq [K] \setminus k$  for all  $k$ .
- 2) Multicasting phase (phases 2 –  $K$ ): for a subset  $\mathcal{J}$  of users, generate  $V_{\mathcal{J}}$  as a linear combination of overheard packets such that

$$V_{\mathcal{J}} = \mathcal{F}_{\mathcal{J}}(\{\mathcal{L}_{\mathcal{J} \setminus \mathcal{J}'}(V_{\mathcal{J}'})\}_{\mathcal{J}' : \mathcal{J}' \subset \mathcal{J} \subset [K]}), \quad (47)$$

where  $\mathcal{F}_{\mathcal{J}}$  denotes a linear function. Send  $V_{\mathcal{J}}$  sequentially for all  $\mathcal{J} \subseteq [K]$  of the cardinality  $|\mathcal{J}| = 2, \dots, K$ .

The achievability result of Theorem 2 implies the following corollary.

**Corollary 2.** *For  $K \leq 3$ , or for the symmetric channel with  $K > 3$ , or for the one-sided fair rate vector with  $K > 3$ , the total transmission length to convey  $W_1, \dots, W_K$  to users  $1, \dots, K$ , respectively, is given by*

$$T_{\text{tot}} = \sum_{k=1}^K \frac{F_{\pi_k}}{1 - \prod_{j=1}^k \delta_{\pi_j}} + \epsilon_F.$$

The proof is omitted because the proof in section V-B covers the case without user memories.

TABLE I  
NOTATIONS FOR THE ERASURE BROADCAST CHANNEL.

$R_k$	Message rate for user $k$
$t_{\mathcal{J}}$	Length of sub-phase $\mathcal{J}$
$t_{\mathcal{J}}^{\{k\}}$	Length needed by user $k$ for sub-phase $\mathcal{J}$
$V_{\mathcal{K}}$	Packets intended to users in $\mathcal{K}$
$\mathcal{L}_{\mathcal{J}}(V_{\mathcal{K}})$	Part of packets $V_{\mathcal{K}}$ received by users in $\mathcal{J}$ and erased at users in $[K] \setminus \mathcal{J}$
$N_{\mathcal{J} \rightarrow \mathcal{J}}^{\{k\}}$	Number of packets useful for user $k$ generated in sub-phase $\mathcal{J}$ and to be sent in sub-phase $\mathcal{J}$

In order to calculate the total transmission length of the algorithm, we need to introduce further some notations and parameters (Table I) which are explained as follows.

- A packet intended to  $\mathcal{J}$  is consumed for a given user  $k \in \mathcal{J}$ , if this user or at least one user in  $[K] \setminus \mathcal{J}$  receives it. The probability of such event is equal to  $1 - \prod_{j \in [K] \setminus \mathcal{J} \cup \{k\}} \delta_j$ .
- A packet intended to  $\mathcal{J}$  becomes a packet intended to  $\mathcal{J}$  and useful for user  $k \in \mathcal{J} \subset \mathcal{J} \subseteq [K]$ , if erased at user  $k$  and all users in  $[K] \setminus \mathcal{J}$  but received by  $\mathcal{J} \setminus \mathcal{J}$ . The number of packets useful for user  $k$  generated in

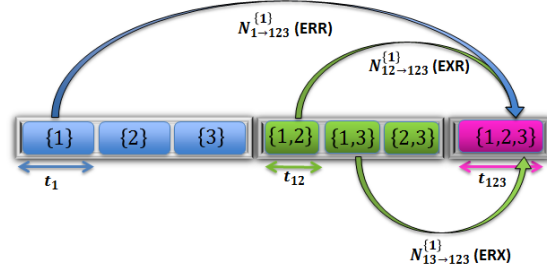


Fig. 3. Phase organization for  $K = 3$  and packet evolution viewed by user 1.

sub-phase  $\mathcal{J}$  and to be sent in sub-phase  $\mathcal{J}$ , denoted by  $N_{\mathcal{J} \rightarrow \mathcal{J}}^{\{k\}}$ , is then given by

$$N_{\mathcal{J} \rightarrow \mathcal{J}}^{\{k\}} = t_{\mathcal{J}}^{\{k\}} \prod_{j' \in [K] \setminus \mathcal{J} \cup \{k\}} \delta_{j'} \prod_{j \in \mathcal{J} \setminus \mathcal{J}} (1 - \delta_j) \quad (48)$$

where  $t_{\mathcal{J}}^{\{k\}}$  denotes the length of sub-phase  $\mathcal{J}$  viewed by user  $k$  to be defined shortly. We can also express  $N_{\mathcal{J} \rightarrow \mathcal{J}}^{\{k\}}$  as

$$N_{\mathcal{J} \rightarrow \mathcal{J}}^{\{k\}} = \sum_{\mathcal{J}' \subseteq \mathcal{J} \setminus k} |\mathcal{L}_{\mathcal{J} \setminus \mathcal{J}' \cup \mathcal{J}'}(V_{\mathcal{J}}^{\{k\}})|, \quad (49)$$

where we let  $V_{\mathcal{J}}^{\{k\}}$  denotes the part of  $V_{\mathcal{J}}$  required for user  $k$ .

- The duration  $t_{\mathcal{J}}$  of sub-phase  $\mathcal{J}$  is given by

$$t_{\mathcal{J}} = \max_{k \in \mathcal{J}} t_{\mathcal{J}}^{\{k\}}, \quad (50)$$

where

$$t_{\mathcal{J}}^{\{k\}} = \frac{\sum_{k \in \mathcal{J} \subseteq \mathcal{J}} N_{\mathcal{J} \rightarrow \mathcal{J}}^{\{k\}}}{1 - \prod_{j \in [K] \setminus \mathcal{J} \cup \{k\}} \delta_j}. \quad (51)$$

The total transmission length is given by summing up all sub-phases, i.e.  $T_{\text{tot}} = \sum_{\mathcal{J} \subseteq [K]} t_{\mathcal{J}}$ .

Fig. 3 illustrates the phase organization for  $K = 3$  and the packet evolution viewed by user 1. The packets intended to  $\{1, 2, 3\}$  are created from both phases 1 and 2. More precisely, sub-phase  $\{1\}$  creates  $\mathcal{L}_{23}(V_1)$  to be sent in phase 3 if erased at user 1 and received by others (ERR). The number of such packets is  $N_{1 \rightarrow 123}^{\{1\}}$ . Sub-phase  $\{1, 2\}$  creates  $\mathcal{L}_3(V_{12}), \mathcal{L}_{23}(V_{12})$  if erased at user 1 but received by user 3 (EXR), while sub-phase  $\{1, 3\}$  creates  $\mathcal{L}_2(V_{13}), \mathcal{L}_{23}(V_{13})$  if erased at user 1 and received by user 2 (ERX). The total number of packets intended to  $\{1, 2, 3\}$  generated in phase 2 and required by user 1 is  $N_{12 \rightarrow 123}^{\{1\}} + N_{13 \rightarrow 123}^{\{1\}}$ .

### B. Achievability in the symmetric channel

We focus now on the special case of the symmetric channel with uniform erasure probabilities, i.e.  $\delta_k = \delta$  for all  $k$ . In this case, the capacity region of the EBC with state feedback in (46) simplifies to

$$\sum_{k=1}^K \frac{1}{1 - \delta^k} R_{\pi_k} \leq 1, \quad \forall \pi. \quad (52)$$

It readily follows that the capacity region yields the symmetric capacity, i.e.  $R_1 = \dots = R_K = R_{\text{sym}}(K)$ , given by

$$R_{\text{sym}}(K) = \frac{1}{\sum_{k=1}^K \frac{1}{1-\delta^k}}. \quad (53)$$

In the following, we provide an alternative proof of the achievability of the symmetric capacity. Notice that other vertices of the capacity region can be characterized similarly as proved in subsection V-C. Our proof follows the footsteps of [14] and uses the notion of order- $j$  packets. Let us define message set  $\{W_{\mathcal{J}}\}$  independently and uniformly distributed over  $\{\mathcal{W}_{\mathcal{J}}\}$  for all  $\mathcal{J} \subseteq [K]$ . For  $\mathcal{J}$  with the cardinality  $j = |\mathcal{J}|$ , the message set  $\{W_{\mathcal{J}}\}$  are called order- $j$  messages. We define  $R_{\mathcal{J}}$  an achievable rate of the message  $W_{\mathcal{J}}$  and define the sum rate of order- $j$  messages as

$$R^j(K) \triangleq \sum_{\mathcal{J}: |\mathcal{J}|=j} R_{\mathcal{J}} = \binom{K}{j} R_{\mathcal{J}}. \quad (54)$$

The supremum of  $R^j(K)$  is called the sum capacity of order- $j$  messages. We characterize the sum capacity of order- $j$  messages, in the erasure broadcast channel with state feedback in the following theorem.

**Theorem 3.** *In the  $K$ -user erasure broadcast channel with state feedback, the sum capacity of order- $j$  packets is upper bounded by*

$$R^j(K) \leq \frac{\binom{K}{j}}{\sum_{k=1}^{K-j+1} \frac{\binom{K-k}{j-1}}{1-\delta^k}}, \quad j = 1, \dots, K. \quad (55)$$

*The algorithms in [15], [16] achieve the RHS with equality.*

*Proof:* We first provide the converse proof. Similarly to section III, we build on genie-aided bounds together with Lemma 1. Let us assume that the transmitter wishes to convey the message  $W_{\mathcal{J}}$  to a subset of users  $\mathcal{J} \subseteq \{1, \dots, K\}$ , and receiver  $k$  wishes to decode all messages  $\tilde{W}_k \triangleq \{W_{\mathcal{J}}\}_{\mathcal{J}: \mathcal{J} \ni k}$  for  $j = 1, \dots, K$ . In order to create a degraded broadcast channel, we assume that receiver  $k$  provides the message set  $\tilde{W}_k$  and the channel output  $Y_k^n$  to receivers  $k+1$  to  $K$  for  $k = 1, \dots, K-1$ . Under this setting and using Fano's inequality, we have for receiver 1 :

$$n \left( \sum_{1 \in \mathcal{J} \subseteq [K]} R_{\mathcal{J}} - \epsilon_{n,1} \right) \leq H(Y_1^n | S^n) - H(Y_1^n | \tilde{W}_1 S^n). \quad (56)$$

For receiver  $k = 2, \dots, K$ , we have:

$$n \left( \sum_{k \in \mathcal{J} \subseteq \{k, \dots, K\}} R_{\mathcal{J}} - \epsilon_{n,k} \right) \leq H(Y_1^n \dots Y_k^n | \tilde{W}^{k-1} S^n) - H(Y_1^n \dots Y_k^n | \tilde{W}^k S^n), \quad (57)$$

where we used  $\tilde{W}_k \setminus \tilde{W}^{k-1} = \{W_{\mathcal{J}}\}_{\mathcal{J}: \mathcal{J} \setminus \{k, \dots, K\}}$  in the LHS. Summing up the above inequalities and applying Lemma 1  $K-1$  times, we readily obtain:

$$\sum_{k=1}^K \frac{\sum_{k \in \mathcal{J} \subseteq \{k, \dots, K\}} (R_{\mathcal{J}} - \epsilon_{n,k})}{1 - \delta^k} \leq \frac{H(Y_1^n | S^n)}{n(1 - \delta)} \quad (58)$$

$$\leq 1. \quad (59)$$

TABLE II  
NOTATIONS FOR THE SYMMETRIC CHANNEL.

$W_{\mathcal{J}}$	Message intended to users in $\mathcal{J}$
$R_{\mathcal{J}}$	Rate of $W_{\mathcal{J}}$
$R^j(K)$	Sum rate of order- $j$ messages
$t_j = t_j^1$	Length of any sub-phase in phase $j$
$t_j^i$	Length of any sub-phase $j$ when starting from phase $i$
$N_{i \rightarrow j} = N_{i \rightarrow j}^1$	Number of packets created in sub-phase $\mathcal{J}$ and to be sent in sub-phase $\mathcal{J}$ for any $\mathcal{J} \subset \mathcal{J}$ of cardinality $i < j$
$N_{i \rightarrow j}^{i'}$	$N_{i \rightarrow j}$ when starting from phase $i'$ for $i \leq i' \leq j$

We further impose the symmetric rate condition such that  $R_{\mathcal{J}} = R_{\mathcal{J}'}$  for any  $\mathcal{J} \neq \mathcal{J}'$  with the same cardinality. By focusing on  $\mathcal{J}$  of the same cardinality  $j$  in (58) and noticing that there are  $\binom{K-k}{j-1}$  such subset,  $R_{\mathcal{J}}$  is upper bounded by

$$R_{\mathcal{J}} \leq \frac{1}{\sum_{k=1}^{K-j+1} \frac{\binom{K-k}{j-1}}{1-\delta^k}}, \quad \forall \mathcal{J}, |\mathcal{J}| = j. \quad (60)$$

This establishes the converse part.

In order to prove the achievability of  $R^i(K)$  in Theorem 3, we apply the broadcasting algorithm of [15], [16] from phase  $i > 1$  by sending  $N_i$  packets to each subset  $\mathcal{J} \subseteq [K]$  with  $|\mathcal{J}| = i$ . First, we redefine some parameters by taking into account the symmetry across users as summarized in Table II. Due to the symmetry, we drop the user index  $k$  in  $t_{\mathcal{J}}^{\{k\}}$ ,  $N_{j \rightarrow \mathcal{J}}^{\{k\}}$  and replace them by  $t_j$ ,  $N_{i \rightarrow j}$ , respectively for  $\mathcal{J} \subset \mathcal{J} \subseteq [K]$  with  $|\mathcal{J}| = i$ ,  $|\mathcal{J}| = j$ . Now, we introduce variants of these notations to reflect the fact that the algorithm starts from phase  $i > 1$ , rather than from phase 1. The length of any sub-phase in phase  $j$  when starting the algorithm from phase  $i$ , denoted by  $t_j^i$ , is given by

$$t_j^i = \frac{1}{1 - \delta^{K-j+1}} \sum_{l=i}^{j-1} \binom{j-1}{l-1} N_{l \rightarrow j}^i, \quad j > i, \quad (61)$$

where

$$N_{l \rightarrow j}^i = t_l^i \delta^{K-j+1} (1 - \delta)^{j-l} \quad (62)$$

denotes the number of order- $j$  packets generated during a given sub-phase in phase  $i$ , again starting from phase  $i$ .

For  $j = i$ , we have

$$t_i^i = \frac{N_i}{1 - \delta^{K-i+1}}. \quad (63)$$

By counting the total number of order- $i$  packets and the transmission length from phase  $i$  to phase  $K$ , the sum rate of order- $i$  messages achieved by the algorithm [15], [16] is given by

$$\tilde{R}^i(K) = \frac{\binom{K}{i} N_i}{\sum_{j=i}^K \binom{K}{j} t_j^i}, \quad \forall i. \quad (64)$$

It remains to prove that  $\tilde{R}^i(K)$  coincides with the RHS expression of (55). We notice that the transmission length from phase  $j$  to  $K$  can be expressed in the following different way, i.e.

$$\sum_{j=i}^K \binom{K}{j} t_j^i = \sum_{j=i}^K U_j^i, \quad (65)$$

where we let

$$U_j^i = \sum_{l=i}^j \binom{j-1}{l-1} t_l^i, \quad \forall j \geq i. \quad (66)$$

By following similar steps as [16, Appendix C], we obtain the recursive equations given by

$$U_j^i = \frac{1}{1 - \delta^{K-j+1}} \sum_{l=1}^{j-i} \binom{j-1}{l} (-1)^{l+1} (1 - \delta^{K-j+l+1}) U_{j-l}^i \quad (67)$$

for  $j > i$ . Since we have  $U_i^i = t_i^i = \frac{N_i}{1 - \delta^{K-i+1}}$  and using the equality  $\binom{j-1}{c} \binom{j-c-1}{i-1} = \binom{j-1}{j-i} \binom{j-i}{c}$  and the binomial theorem  $\sum_{k=0}^n \binom{n}{k} x^k y^{n-k} = (x+y)^n$ , it readily follows that we have

$$U_j^i = \frac{N_i}{1 - \delta^{K-j+1}} \binom{j-1}{j-i}, \quad j \geq i. \quad (68)$$

By plugging the last expression into (64) using (65), we have

$$\tilde{R}^i(K) = \frac{\binom{K}{i} N_i}{\sum_{j=i}^K \frac{N_i}{1 - \delta^{K-j+1}} \binom{j-1}{j-i}} \quad (69)$$

$$= \frac{\binom{K}{i}}{\sum_{k=1}^{K-i+1} \frac{\binom{K-k}{i-1}}{1 - \delta^k}} \quad (70)$$

which coincides the RHS of (55) for  $i = 1, \dots, K$ . This establishes the achievability proof.  $\blacksquare$

As a corollary of Theorem 3, we provide an alternative expression for the sum capacity.

**Corollary 3.** *The sum capacity of the  $K$ -user symmetric broadcast erasure channel with state feedback can be expressed as a function of  $R^2(K), \dots, R^K(K)$  by*

$$R^1(K) = \frac{KN_1}{\frac{KN_1}{1 - \delta^K} + \sum_{i=2}^K \frac{\binom{K}{i} N_{1 \rightarrow i}}{R^i(K)}}, \quad (71)$$

where  $\frac{KN_1}{1 - \delta^K}$  is the duration of phase 1,  $\binom{K}{j} N_{1 \rightarrow j}$  corresponds to the total number of order- $j$  packets generated in phase 1.

*Proof:* By letting  $f$  denote the RHS of (71), we wish to prove the equality  $f = R^1(K) = \frac{KN_1}{\sum_{k=1}^K \frac{1}{1 - \delta^k}}$  by proving  $f = \tilde{R}^1(K)$ . If it is true, from the achievability proof of Theorem 3 that proves  $\tilde{R}^i = R^i$  for all  $i$ , the proof is complete. In the RHS of (71), we replace  $R^i$  by the expression  $\tilde{R}^i$  in (64) by letting  $N_{1 \rightarrow i} = N_i$  for  $i \geq 2$ . Then, we have

$$f = \frac{KN_1}{\frac{KN_1}{1 - \delta^K} + \sum_{i=2}^K \sum_{j=i}^K \binom{K}{j} t_j^i} \quad (72)$$

$$= \frac{KN_1}{\frac{KN_1}{1 - \delta^K} + \sum_{j=2}^K \binom{K}{j} \sum_{i=2}^j t_j^i}. \quad (73)$$



Comparing the desired equality  $f = \tilde{R}^1(K) = \frac{KN_1}{\sum_{j=1}^K \binom{K}{j} t_j^1}$  with the above expression and noticing that  $\frac{KN_1}{1-\delta^K} = Kt_1^1$ , we immediately see that it remains to prove the following equality.

$$t_j^1 = \sum_{i=2}^j t_j^i \quad \forall j \geq 2. \quad (74)$$

We prove this relation recursively. For  $j = 2$ , the above equality follows from (61) and (63).

$$t_2^1 = \frac{N_{1 \rightarrow 2}}{1 - \delta^{K-1}} = t_2^2. \quad (75)$$

Now suppose that (74) holds for  $l = 2, \dots, j-1$  and we prove it for  $j$ . From (61) we have

$$t_j^1 = \frac{1}{1 - \delta^{K-j+1}} \sum_{l=1}^{j-1} \binom{j-1}{l-1} N_{l \rightarrow j}^1 \quad (76)$$

$$= \frac{1}{1 - \delta^{K-j+1}} \left[ N_{1 \rightarrow j} + \sum_{l=2}^{j-1} \binom{j-1}{l-1} t_l^1 \delta^{K-j+1} (1 - \delta)^{j-l} \right] \quad (77)$$

$$= \frac{1}{1 - \delta^{K-j+1}} \left[ N_{1 \rightarrow j} + \sum_{l=2}^{j-1} \binom{j-1}{l-1} \sum_{i=2}^l t_l^i \delta^{K-j+1} (1 - \delta)^{j-l} \right] \quad (78)$$

$$= \frac{1}{1 - \delta^{K-j+1}} \left[ N_{1 \rightarrow j} + \sum_{l=2}^{j-1} \binom{j-1}{l-1} \sum_{i=2}^l N_{l \rightarrow j}^i \right] \quad (79)$$

$$= \frac{1}{1 - \delta^{K-j+1}} \left[ N_{1 \rightarrow j} + \sum_{i=2}^{j-1} \sum_{l=i}^{j-1} \binom{j-1}{l-1} N_{l \rightarrow j}^i \right] \quad (80)$$

$$= t_j^j + \sum_{i=2}^{j-1} t_j^i, \quad (81)$$

where (77) follows from (62); (78) follows from our hypothesis (74); (79) follows from (62); (80) is due to the equality  $\sum_{l=2}^{j-1} \sum_{i=2}^l = \sum_{i=2}^{j-1} \sum_{l=i}^{j-1}$ ; the last equality is due to (61). Therefore, the desired equality holds also for  $j$ . This completes the proof of Corollary 3.  $\blacksquare$

## V. ACHIEVABILITY

We provide the achievability proof of Theorem 1 for the case of one-sided fair rate vector as well as the symmetric network. The proof for the case of  $K = 3$  is omitted, since it is a straightforward extension of [15, Section V].

### A. Proposed delivery scheme for $K > 3$

We describe the proposed delivery scheme for the case of  $K > 3$  assuming that user  $k$  requests file  $W_k$  of size  $F_k$  packets for  $k = 1, \dots, K$  without loss of generality. Compared to the algorithm [15], [16] revisited previously, our scheme must convey packets created during the placement phase as well as all previous phases in each phase. Here is a high-level description of our proposed delivery scheme.

- 1) Placement phase (phase 0): fill the caches  $Z_1, \dots, Z_K$  according to the decentralized content placement (see subsection II-B). This phase creates “overheard” packets  $\{\mathcal{L}_{\mathcal{J} \setminus k}(W_k)\}$  for  $\mathcal{J} \subset [K]$  and all  $k$  to be delivered during phases 1 to  $K$ .

- 2) Broadcasting phase (phase 1): the transmitter sends  $V_1, \dots, V_K$  sequentially until at least one user receives it, where  $V_k = \mathcal{L}_\emptyset(W_k)$  corresponds to the order-1 packets.
- 3) Multicasting phase (phases 2- $K$ ): for a subset  $\mathcal{J}$  of users, generate  $V_{\mathcal{J}}$  as a linear combination of overheard packets during the placement phase as well as during phases 1 to  $j - 1$ . Send  $V_{\mathcal{J}}$  sequentially for  $\mathcal{J} \subseteq [K]$ ,

$$V_{\mathcal{J}} = \mathcal{F}_{\mathcal{J}} \left( \{ \mathcal{L}_{\mathcal{J} \setminus \mathcal{J} \cup \mathcal{J}'}(V_{\mathcal{J}}) \}_{\mathcal{J}' : \mathcal{J}' \subset \mathcal{J} \subset \mathcal{J}, \mathcal{L}_{\mathcal{J} \setminus \{k\}}(W_k)} \right). \quad (82)$$

The proposed delivery scheme achieves the optimal rate region only in two special cases. We provide the proof separately in upcoming subsections.

### B. Proof of Theorem 1 for the case of one-sided fair rate vector

We assume without loss of generality  $\delta_1 \geq \dots \geq \delta_K$ ,  $\delta_1 R_1 \geq \dots \geq \delta_K R_K$ , and  $\frac{1-p_1}{p_1} R_1 \geq \dots \geq \frac{1-p_2}{p_2} R_K$ . Under this setting, we wish to prove the achievability of the following equality.

$$\sum_{k=1}^K \frac{\prod_{j=1}^k (1-p_j)}{1 - \prod_{j=1}^k \delta_j} R_k = 1. \quad (83)$$

By replacing  $R_k = \frac{F_{d_k}}{T_{\text{tot}}}$  and further assuming  $d_k = k$  for all  $k$  without loss of generality, the above equality is equivalent to

$$T_{\text{tot}} = \sum_{k=1}^K \frac{\prod_{j=1}^k (1-p_j)}{1 - \prod_{j=1}^k \delta_j} F_k. \quad (84)$$

The rest of the subsection is dedicated to the proof of the total transmission length (84). We start by rewriting  $t_{\mathcal{J}}^{\{k\}}$  in (51) by incorporating the packets generated during the placement phase. Namely we have for  $k \in \mathcal{J} \subseteq [K]$

$$t_{\mathcal{J}}^{\{k\}} = \frac{\sum_{\mathcal{J}: k \in \mathcal{J} \subset \mathcal{J}} N_{\mathcal{J} \rightarrow \mathcal{J}}^{\{k\}} + |\mathcal{L}_{\mathcal{J} \setminus \{k\}}(W_k)|}{1 - \prod_{j \in [K] \setminus \mathcal{J} \cup \{k\}} \delta_j}. \quad (85)$$

We recall that the length of sub-phase  $\mathcal{J}$  is given by  $t_{\mathcal{J}} = \max_{k \in \mathcal{J}} t_{\mathcal{J}}^{\{k\}}$ . Our proof consists of four steps.

a) *Step 1* : We express  $t_{\mathcal{J}}^{\{k\}}$  as a function of key parameters  $\{\delta_k\}, \{p_k\}, \{F_k\}$  in two different ways. By following similar steps as in [16, Appendix C], the aggregate length of sub-phases  $\mathcal{J} \subseteq \mathcal{J}$  required by user  $k$  for a fixed  $\mathcal{J} \subseteq [K]$  is given by

$$\sum_{\mathcal{J}: k \in \mathcal{J} \subseteq \mathcal{J}} t_{\mathcal{J}}^{\{k\}} = \frac{\prod_{j \in [K] \setminus \mathcal{J} \cup \{k\}} (1-p_j)}{1 - \prod_{j \in [K] \setminus \mathcal{J} \cup \{k\}} \delta_j} F_k. \quad (86)$$

We have an alternative expression for  $t_{\mathcal{J}}^{\{k\}}$  which is useful as will be seen shortly. The length of sub-phase  $\mathcal{J}$  needed by user  $k$  such that  $k \in \mathcal{J} \subseteq [K]$  is equal to

$$t_{\mathcal{J}}^{\{k\}} = \sum_{\mathcal{H}: \mathcal{H} \subseteq \mathcal{J} \setminus \{k\}} (-1)^{|\mathcal{H}|} \frac{\prod_{j \in [K] \setminus \mathcal{J} \cup \{k\} \cup \mathcal{H}} (1-p_j)}{1 - \prod_{j \in [K] \setminus \mathcal{J} \cup \{k\} \cup \mathcal{H}} \delta_j} F_k. \quad (87)$$

The proof is provided in Appendix A.

b) *Step 2* : The length of sub-phase  $\mathcal{J}$  is determined by the worst user which requires the maximum length, i.e.  $\arg \max_{k \in \mathcal{J}} t_{\mathcal{J}}^{\{k\}}$ . For the special case of one-sided fair rate vector, by using (87) it is possible to prove that the worst user is given by

$$\arg \max_{k \in \mathcal{J}} t_{\mathcal{J}}^{\{k\}} = \min\{\mathcal{J}\}, \forall \mathcal{J} \subseteq [K], \quad (88)$$

where  $\min\{\mathcal{J}\}$  is the smallest index in the set of users  $\mathcal{J}$  that corresponds to the user with the largest erasure probability. The proof is provided in Appendix B. This means that the user permutation (which determines the sub-phase length) is preserved in all sub-phases for the one-sided fair rate vector.

c) *Step 3* : By combining the two previous steps, the total transmission length can be derived as follows.

$$T_{\text{tot}} = \sum_{\mathcal{J}: \mathcal{J} \subseteq [K]} \max_{k \in \mathcal{J}} t_{\mathcal{J}}^{\{k\}} \quad (89)$$

$$= \sum_{\mathcal{J}: \mathcal{J} \subseteq [K]} t_{\mathcal{J}}^{\{\min \mathcal{J}\}} \quad (90)$$

$$= \sum_{k=1}^K \sum_{\mathcal{J}: k \in \mathcal{J} \subseteq \{k, \dots, K\}} t_{\mathcal{J}}^{\{k\}} \quad (91)$$

$$= \sum_{k=1}^K F_k \frac{\prod_{j=1}^k (1 - p_j)}{1 - \prod_{j=1}^k \delta_j}, \quad (92)$$

where (90) is obtained from (88); the last equality follows from (86). Then, we obtain the desired equality (84).

d) *Step 4* : The final step is to prove that under the one-sided fair rate vector (83) implies all the other  $K! - 1$  inequalities of the rate region (11). This is proved in Appendix C. Hence, the achievability proof for the one-sided rate vector is completed.

### C. Proof of Theorem 1 for the symmetric network

First we recall the rate region of the symmetric network with uniform channel statistics and memory sizes given in (12),

$$\sum_{k=1}^K \frac{(1-p)^k}{1-\delta^k} R_{\pi_k} \leq 1, \quad \forall \pi. \quad (93)$$

Exploiting the polyhedron structure and following the same footsteps as [14, Section VI], we can prove that the vertices of the above rate region are characterized as:

$$R_k = \begin{cases} R_{\text{sym}}(|\mathcal{K}|), & k \in \mathcal{K} \\ 0, & k \notin \mathcal{K} \end{cases} \quad (94)$$

for  $\mathcal{K} \subseteq [K]$ , where the symmetric rate  $R_{\text{sym}}(K)$  is given by

$$R_{\text{sym}}(K) = \frac{1}{\sum_{k=1}^K \frac{(1-p)^k}{1-\delta^k}}. \quad (95)$$

This means that when only  $|\mathcal{K}|$  users are active in the system, each of these users achieves the same symmetric rate as the reduced system of dimension  $|\mathcal{K}|$ . Then, it suffices to prove the achievability of the symmetric rate for a given

dimension  $K$ . As explained in subsection V-A, the placement phase generates “overheard packets”  $\{\mathcal{L}_{\mathcal{J} \setminus k}(W_k)\}$  for  $\mathcal{J} \subseteq [K]$  and all  $k$ . We let  $N_{0 \rightarrow j} = |\mathcal{L}_{\mathcal{J} \setminus j}(W_k)|$  denote the number of order- $j$  packets created during the placement phase. Then, we can express the sum rate of the cached-enabled EBC by incorporating the packets generated from the placement phase into (71) as follows,

$$KR_{\text{sym}}(K) = \frac{KF}{\frac{KN_{0 \rightarrow 1}}{\beta_1} + \sum_{j=2}^K \frac{\binom{K}{j}(N_{0 \rightarrow j} + N_{1 \rightarrow j})}{R^j(K)}}. \quad (96)$$

By repeating the same steps as the proof of Corollary 3, it readily follows that the above expression boils down to

$$\frac{K}{\sum_{k=1}^K \frac{(1-p)^k}{1-\delta^k}}. \text{ This establishes the achievability proof for the symmetric network.}$$

## VI. EXTENSIONS

In this section, we provide rather straightforward extensions of our previous results to other scenarios such as the centralized content placement and the multi-antenna broadcast channel with the state feedback.

### A. Centralized content placement

So far, we have focused on the decentralized content placement. We shall show in this subsection that the rate region under the decentralized content placement can be easily modified to the case of the centralized content placement proposed in [1]. We restrict ourselves to the symmetric memory size  $M_k = M$  such that  $M \in \{0, N/K, 2N/K, \dots, N\}$  so that the parameter  $b = \frac{MK}{N}$  is an integer. Each file is split into  $\binom{K}{b}$  disjoint equal size sub-files. Each sub-file is cached at a subset of users  $\mathcal{J}$ ,  $\forall \mathcal{J} \subseteq [K]$  with cardinality  $|\mathcal{J}| = b$ . Namely, the size of any sub-file of file  $i$  is given by

$$|\mathcal{L}_{\mathcal{J}}(W_i)| = \frac{1}{\binom{K}{b}} F_i, \quad (97)$$

which satisfies the memory constraint for user  $k$

$$|Z_k| = \sum_{i=1}^N \sum_{\mathcal{J}: k \in \mathcal{J}; |\mathcal{J}|=b} |\mathcal{L}_{\mathcal{J}}(W_i)| = \sum_{i=1}^N \binom{K-1}{b-1} \frac{F_i}{\binom{K}{b}} = \sum_{i=1}^N \frac{b}{K} F_i = MF. \quad (98)$$

In analogy to Lemma 2 for the decentralized content placement, we can characterize the message entropy given the receiver side information.

**Lemma 3.** *For the centralized content placement [1], the following equalities hold for any  $i$  and  $\mathcal{K} \subseteq [K]$*

$$H(W_i | \{Z_k\}_{k \in \mathcal{K}}) = \frac{\binom{K-|\mathcal{K}|}{b}}{\binom{K}{b}} H(W_i).$$

*Proof:* Under the centralized content placement

$$H(W_i | \{Z_k\}_{k \in \mathcal{K}}) = \sum_{\mathcal{J} \subseteq [K] \setminus \mathcal{K}} H(\mathcal{L}_{\mathcal{J}}(W_i)) \quad (99)$$

$$= \sum_{\mathcal{J} \subseteq [K] \setminus \mathcal{K}; |\mathcal{J}|=b} H(\mathcal{L}_{\mathcal{J}}(W_i)) \quad (100)$$

$$= \sum_{\mathcal{J} \subseteq [K] \setminus \mathcal{K}; |\mathcal{J}|=b} \frac{1}{\binom{K}{b}} H(W_i) \quad (101)$$

$$= \frac{\binom{K-|\mathcal{K}|}{b}}{\binom{K}{b}} H(W_i), \quad (102)$$

where the first equality follows by repeating the same steps from (28) to (31); (100) and (101) follows from the definition of the centralized content placement (97). ■

Then, we present the rate region of the cache-enabled EBC under the centralized content placement.

**Theorem 4.** *For the symmetric network, the rate region of the cached-enabled EBC with the state feedback under the centralized content placement is given by*

$$\sum_{k=1}^{K-b} \frac{\binom{K-k}{b} / \binom{K}{b}}{1 - \delta^k} R_{\pi_k} \leq 1 \quad (103)$$

for any permutation  $\pi$  of  $\{1, \dots, K\}$ .

*Proof:* Following the same steps as in section III and replacing Lemma 2 with Lemma 3, the converse proof follows immediately.

For achievability, as explained in subsection V-C, it is sufficient to consider the case of symmetric rate for a given dimension. By focusing without loss of generality on the dimension  $K$ , we fix the number of packets per user to be  $F$  and prove that our proposed scheme can deliver requested files to users within the total transmission length given by

$$T_{\text{tot}} = F \sum_{k=1}^{K-b} \frac{\binom{K-k}{b} / \binom{K}{b}}{1 - \delta^k} + \epsilon_F, \quad (104)$$

as  $F \rightarrow \infty$ . We proceed our proposed delivery scheme from phase  $b+1$  by sending packets of order  $b+1$ . More precisely, in phase  $b+1$  we generate and send the packets intended to  $\mathcal{J}$  by the following linear combination

$$V_{\mathcal{J}} = \mathcal{F}_{\mathcal{J}}(\mathcal{L}_{\mathcal{J} \setminus k}(W_k)), \quad (105)$$

for  $\mathcal{J} \subseteq [K]$  with  $|\mathcal{J}| = b+1$ . In subsequent phases  $b+2$  to  $K$ , we repeat

$$V_{\mathcal{J}} = \mathcal{F}_{\mathcal{J}}(\{\mathcal{L}_{\mathcal{J} \setminus \mathcal{J}' \cup \mathcal{J}'}(V_{\mathcal{J}})\}_{\mathcal{J}' \setminus \mathcal{J}; \mathcal{J}' \subset \mathcal{J} \subset \mathcal{J}}) \quad (106)$$

for  $\mathcal{J} \subseteq [K]$  with  $|\mathcal{J}| = b+2, \dots, K$ . In order to calculate the total transmission length required by our delivery algorithm, we follow the same footsteps as in subsection V-C and exploit Theorem 3 on the sum capacity of order- $i$  messages that we recall here for the sake of clarity.

$$R^i(K) = \frac{\binom{K}{i}}{\sum_{k=1}^{K-i+1} \frac{\binom{K-k}{i-1}}{1 - \delta^k}}. \quad (107)$$

Noticing that there are  $\binom{K}{b+1}$  sub-phases in phase  $b+1$  and in each sub-phase we send a linear combination whose size is  $\frac{F}{\binom{K}{b}}$ , the total transmission length is given by

$$T_{\text{tot}} = \frac{\binom{K}{b+1}/\binom{K}{b}}{R^{b+1}} F \quad (108)$$

$$= \sum_{k=1}^{K-b} \frac{\binom{K-k}{b}/\binom{K}{b}}{1 - \delta^k} F, \quad (109)$$

where the last equality follows by plugging the expression  $R^{b+1}$ .  $\blacksquare$

For the case without erasure, Theorem 4, in particular, the expression of the transmission length in (104), becomes the rate-memory tradeoff under the centralized content placement [1] given by

$$\frac{T_{\text{tot}}}{F} = K (1 - M/N) \frac{1}{1 + KM/N}. \quad (110)$$

### B. MISO-BC

We consider the multi-input single-output broadcast channel (MISO-BC) between a  $N_t$ -antennas transmitter and  $K$  single-antenna receivers. The channel state  $S_l$  in slot  $l$  is given by the  $N_t \times K$  matrix and we restrict ourselves to the i.i.d. channels across time and users. Here, we are interested in the capacity scaling in the high signal-to-noise ratio (SNR) regime and define the degree of freedom (DoF) of user  $k$  as

$$\text{DoF}_k = \lim_{\text{snr} \rightarrow \infty} \frac{R_k}{\log_2 \text{snr}}.$$

We define the sum DoF of order- $j$  messages given by

$$\text{DoF}^j = \lim_{\text{snr} \rightarrow \infty} \sum_{j: |j|=j} \frac{R_j}{\log_2 \text{snr}}. \quad (111)$$

First we recall the mains results on the MISO-BC with state feedback by Maddah-Ali and Tse [14]. In [14, Theorem 3], the DoF region of the MISO-BC with state feedback has been characterized as

$$\sum_{k=1}^K \frac{\text{DoF}_{\pi_k}}{k} \leq 1, \quad \forall \pi. \quad (112)$$

The sum DoF of order- $j$  messages has been characterized in [14, Theorem 2] and is given by

$$\text{DoF}^j = \frac{\binom{K}{j}}{\sum_{k=1}^{K-j+1} \frac{\binom{K-k}{j-1}}{k}}. \quad (113)$$

It is worth comparing the DoF region of the MISO-BC in (112) and the capacity region of the EBC in (52). In fact, as remarked in [12], both regions have exactly the same structure and can be unified through a parameter  $\alpha_k = k$  for the MISO-BC and  $\alpha_k = 1 - \delta^k$  for the EBC. The same holds for the sum DoF of order- $j$  messages in the MISO-BC in (113) and the sum capacity of order- $j$  packets in the EBC characterized in Theorem 3. By exploiting this duality and replacing  $1 - \delta^k$  with  $k$  in the rate region of the symmetric EBC (12), we can easily characterize the DoF region of the cache-enabled MISO-BC with state feedback. Namely, under the decentralized content placement, the DoF region is given by

$$\sum_{k=1}^K \frac{(1-p)^k}{k} \text{DoF}_{\pi_k} \leq 1, \quad \forall \pi \quad (114)$$

for  $N_t \geq K$ , while under the centralized content placement, the DoF region is given by

$$\sum_{k=1}^{K-b} \frac{\binom{K-k}{b} / \binom{K}{b}}{k} \text{DoF}_{\pi_k} \leq 1, \quad \forall \pi \quad (115)$$

for  $N_t \geq K - b$ . The converse follows exactly in the same manner except that we use the entropy inequality for the MISO-BC given in [12, Lemma 4] by replacing the entropy by the differential entropy and again  $1 - \delta^k$  by  $k$ . The achievability can be proved by modifying the scheme in [14] to the case of receiver side information along the line of [13].

As a final remark, for the case of the centralized content placement, our DoF region in (115) yields the following transmission length

$$T_{\text{tot}} = \sum_{k=1}^{K-b} \frac{\binom{K-k}{b} / \binom{K}{b}}{k} F, \quad (116)$$

which coincides with [26, Corollary 2b].

## VII. NUMERICAL EXAMPLES

In this section, we provide some numerical examples to show the performance of our proposed delivery scheme. Fig. 4 illustrates the tradeoff between the erasure probability and the memory size for the symmetric network with  $K = 3$  for the case of the decentralized content placement. Each curve corresponds to a different symmetric rate  $R_{\text{sym}}(3) = \frac{1}{\sum_{k=1}^3 \frac{(1-p)^k}{1-\delta^k}}$ . The arrow shows the increasing symmetric rate from  $1/3$ , corresponding to case with no memory and no erasure, to infinity. The memory size increases the rate performance even in the presence of erasure and the benefit of caching is significant for smaller erasure probabilities as expected from the analytical expression.

Fig. 5 compares the transmission length  $T_{\text{tot}}$ , normalized by the file size  $F$ , achieved by our delivery scheme with feedback and the scheme without feedback for the case of the decentralized content placement. We consider the system with  $N = 100$ ,  $K = 10$  and the erasure probabilities of  $\delta = 0$  (perfect link), 0.2, and 0.6. We observe that state feedback can be useful especially when the memory size is small and the erasure probability is large. In fact, it can be easily shown that the rate region of the cached-enabled EBC without feedback under the decentralized content placement is given by

$$\sum_{k=1}^K \frac{\left(1 - \frac{M}{N}\right)^k}{1 - \delta} R_{\pi_k} \leq 1 \quad (117)$$

where the denominator in the LHS reflects the fact that each packet must be received by all  $K$  users. This yields the transmission length given by

$$T_{\text{tot-noFB}} = \frac{\sum_{k=1}^K \left(1 - \frac{M}{N}\right)^k}{1 - \delta} F + \epsilon_F. \quad (118)$$

Under the centralized content placement, the rate region of the cached-enabled EBC without feedback is given by

$$\sum_{k=1}^{K-b} \frac{\binom{K-k}{b} / \binom{K}{b}}{1 - \delta} R_{\pi_k} \leq 1 \quad (119)$$

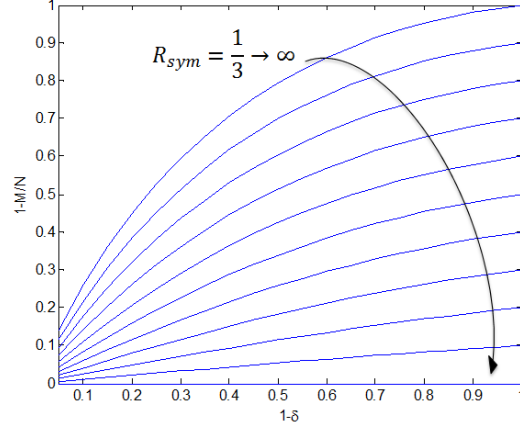


Fig. 4. The tradeoff between the memory and the erasure for  $K = 3$ .

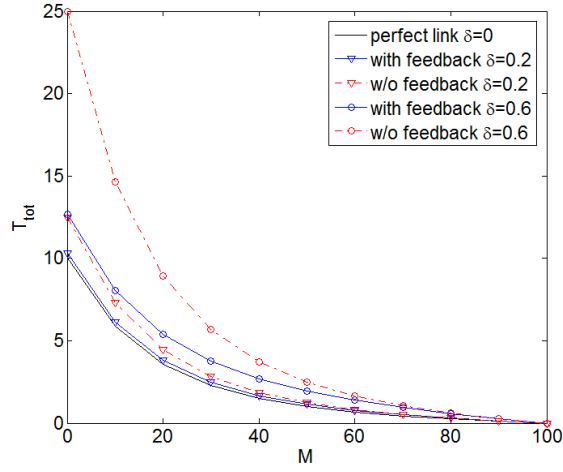


Fig. 5. The transmission length  $T_{\text{tot}}$  as a function of memory size  $M$  for  $N = 100, K = 10$ .

yielding

$$T_{\text{tot-noFB}} = \frac{K(1 - M/N) \frac{1}{1+KM/N}}{1 - \delta} F + \epsilon_F. \quad (120)$$

Without state feedback, the transmission length in (118), (120) corresponds to the transmission length over the perfect link expanded by a factor  $\frac{1}{1-\delta} > 1$ , because each packet must be received by all users. The merit of feedback becomes significant if the packets of lower-order dominate the order- $K$  packets. The case of small  $p = \frac{M}{N}$  and large erasure probability corresponds to such a situation.

Fig. 6 plots the normalized transmission length  $T_{\text{tot}}/F$  versus the memory size  $M$  in the symmetric network with  $N = 100, K = 10$ . We compare the performance with and without feedback under the decentralized and the centralized caching for  $\delta = 0$  and  $\delta = 0.6$ . size. The relative merit of the centralized content placement compared to the decentralized the counterpart can be observed.

Fig. 7 plots the normalized transmission length  $T_{\text{tot}}/F$  versus *average* memory size  $M$  in the asymmetric



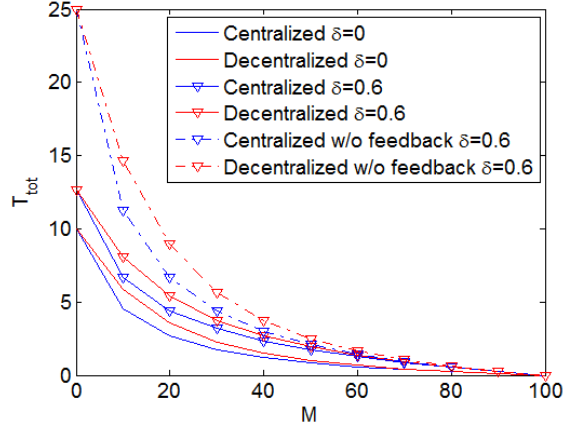


Fig. 6. The transmission length  $T_{\text{tot}}$  as a function of memory size  $M$  for  $N = 100, K = 10$ .

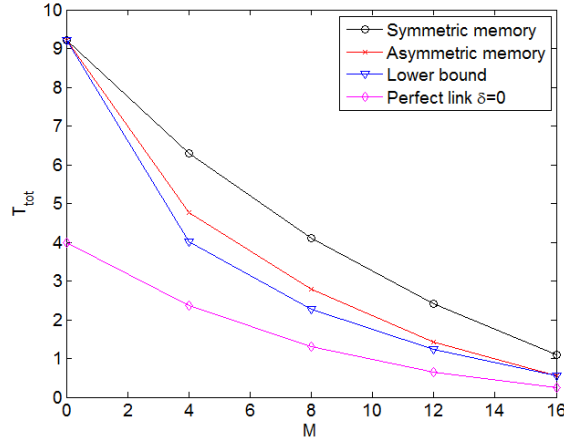


Fig. 7.  $\delta_i = \frac{i}{5}$ ,  $N = 20$ ,  $K = 4$  and  $F_i = 1$ .

network with  $N = 20$  and  $K = 4$  under the decentralized content placement. We let erasure probabilities  $\delta_k = \frac{k}{5}$  for  $k = 1, \dots, 4$  and consider files of equal size. We compare “symmetric memory” ( $M_k = M, \forall k$ ), “asymmetric memory” obtained by optimizing over all possible sets of  $\{M_k\}$  using our delivery scheme, as well as “lower bound” obtained by optimizing over all possible of  $\{M_k\}$  based on (16). This result shows the advantage (in terms of delivery time) of optimally allocating cache sizes across users, whenever possible, according to the condition of the delivery channels.

## VIII. CONCLUSION

In this paper, we investigated the content delivery problem in the erasure broadcast channel (EBC) with state feedback, assuming that the content placement phase is performed with existing methods proposed in the literature. Our main contribution was the characterization of the optimal rate region of the channel under these conditions, based on a scheme that optimally exploits the receiver side information acquired during the placement phase. This appears as a non-trivial extension of the work by Wang and Gatzianas et al. [16], [17] which have characterized the capacity region of the EBC with state feedback for some cases of interest. We provided an intuitive interpretation

of the algorithm proposed in these works and revealed an explicit connection between the capacity in the symmetric EBC and the degree of freedom (DoF) in the MISO-BC. More specifically, we showed that there exists a duality in terms of the order- $j$  multicast capacity/DoF. Such a connection was fully exploited to generalize our results to the cache-enabled MISO-BC. Our work demonstrated the benefits of coded caching combined with state feedback in the presence of random erasure. An interesting future direction is to include some more practical constraints, such as the popularity profile of contents and the non-asymptotic file size, into the current system model.

## APPENDIX

In the appendix, we repeatedly use the following weight expression.

$$w_{\mathcal{J}} = \frac{\prod_{j \in \mathcal{J}} (1 - p_j)}{1 - \prod_{j \in \mathcal{J}} \delta_j} = \frac{\bar{p}_{\mathcal{J}}}{1 - \delta_{\mathcal{J}}} \quad (121)$$

where we let  $\bar{p}_j = 1 - p_j$  and use a short-hand notation  $\delta_{\mathcal{J}} = \prod_{j \in \mathcal{J}} \delta_j$ .

### A. Length of sub-phase

In this section, we prove (87) given by

$$t_{\mathcal{J}}^{\{k\}} = \sum_{\mathcal{H}: \mathcal{H} \subseteq \mathcal{J} \setminus \{k\}} (-1)^{|\mathcal{H}|} \frac{\prod_{j \in [K] \setminus \mathcal{J} \cup \{k\} \cup \mathcal{H}} (1 - p_j)}{1 - \prod_{j \in [K] \setminus \mathcal{J} \cup \{k\} \cup \mathcal{H}} \delta_j} F_k. \quad (122)$$

To this end, we first introduce a new variable  $g_{\mathcal{J}}^{\{k\}} = \frac{t_{\mathcal{J}}^{\{k\}}}{F_k}$  for  $k \in \mathcal{J} \subseteq [K]$ . Using (86) we obtain

$$\sum_{\mathcal{J}: k \in \mathcal{J} \subseteq \mathcal{J}} g_{\mathcal{J}}^{\{k\}} = w_{[K] \setminus \mathcal{J} \cup \{k\}}. \quad (123)$$

We first need to prove the following lemma

**Lemma 4.** *For any nonempty set  $[K]$  and  $\mathcal{J} \subseteq [K]$ . It holds*

$$\sum_{\mathcal{J}: \mathcal{J} \subseteq \mathcal{J}} \sum_{\mathcal{H}: \mathcal{H} \subseteq \mathcal{J}} (-1)^{|\mathcal{H}|} w_{[K] \setminus \mathcal{J} \cup \mathcal{H}} = w_{[K] \setminus \mathcal{J}} \quad (124)$$

*Proof:*

$$\sum_{\mathcal{J}:\mathcal{J}\subseteq\mathcal{J}} \sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus\mathcal{J}\cup\mathcal{H}} = \sum_{\mathcal{J}:\mathcal{J}\subseteq\mathcal{J}} \sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus(\mathcal{J}\setminus\mathcal{H})} \quad (125)$$

$$= \sum_{\mathcal{J}:\mathcal{J}\subseteq\mathcal{J}} \sum_{\mathcal{H}':\mathcal{H}'\subseteq\mathcal{J}} (-1)^{|\mathcal{J}\setminus\mathcal{H}'|} w_{[K]\setminus\mathcal{H}'} \quad (126)$$

$$= \sum_{\mathcal{H}':\mathcal{H}'\subseteq\mathcal{J}} \sum_{\mathcal{J}:\mathcal{H}'\subseteq\mathcal{J}\subseteq\mathcal{J}} (-1)^{|\mathcal{J}\setminus\mathcal{H}'|} w_{[K]\setminus\mathcal{H}'} \quad (127)$$

$$= \sum_{\mathcal{H}':\mathcal{H}'\subseteq\mathcal{J}} w_{[K]\setminus\mathcal{H}'} \sum_{\mathcal{J}:\mathcal{H}'\subseteq\mathcal{J}\subseteq\mathcal{J}} (-1)^{|\mathcal{J}\setminus\mathcal{H}'|} \quad (128)$$

$$= \sum_{\mathcal{H}':\mathcal{H}'\subseteq\mathcal{J}} w_{[K]\setminus\mathcal{H}'} \sum_{\mathcal{J}':\mathcal{J}'\subseteq\mathcal{J}\setminus\mathcal{H}'} (-1)^{|\mathcal{J}'|} \quad (129)$$

$$= w_{[K]\setminus\mathcal{J}} + \sum_{\mathcal{H}':\mathcal{H}'\subset\mathcal{J}} w_{[K]\setminus\mathcal{H}'} \sum_{\mathcal{J}':\mathcal{J}'\subseteq\mathcal{J}\setminus\mathcal{H}'} (-1)^{|\mathcal{J}'|} \quad (130)$$

$$= w_{[K]\setminus\mathcal{J}}. \quad (131)$$

We set  $\mathcal{H}' = \mathcal{J} \setminus \mathcal{H}$  and  $\mathcal{J}' = \mathcal{J} \setminus \mathcal{H}'$  to obtain (126) and (129), respectively. The last equality follows from  $\sum_{\mathcal{J}:\mathcal{J}\subseteq\mathcal{J}} (-1)^{|\mathcal{J}|} = 0$  for all  $\mathcal{J} \neq \emptyset$ .  $\blacksquare$

We prove (87) by induction on  $|\mathcal{J}|$ . For  $\mathcal{J} = \{i\}$  we have  $\sum_{\mathcal{J}:i\in\mathcal{J}\subseteq\mathcal{J}} g_{\mathcal{J}}^{\{i\}} = g_{\mathcal{J}}^{\{i\}}$  and  $\sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}\setminus\{i\}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus\mathcal{J}\cup\{i\}\cup\mathcal{H}} = w_{[K]\setminus\mathcal{J}\cup\{i\}}$ . By apply (123) for  $\mathcal{J} = \{i\}$ , we obtain the proof for  $|\mathcal{J}| = 1$ .

Now suppose (87) holds for any  $\mathcal{J} \subseteq [K]$  such that  $|\mathcal{J}| < |\mathcal{J}|$  and we prove in the following that it holds for  $\mathcal{J}$  too. We have

$$\sum_{\mathcal{J}:i\in\mathcal{J}\subseteq\mathcal{J}} g_{\mathcal{J}}^{\{i\}} = w_{[K]\setminus\mathcal{J}\cup\{i\}} \quad (132)$$

$$= g_{\mathcal{J}}^{\{i\}} + \sum_{\mathcal{J}:i\in\mathcal{J}\subset\mathcal{J}} g_{\mathcal{J}}^{\{i\}}. \quad (133)$$

Thus, we obtain

$$g_{\mathcal{J}}^{\{i\}} = w_{[K]\setminus\mathcal{J}\cup\{i\}} - \sum_{\mathcal{J}:i\in\mathcal{J}\subset\mathcal{J}} g_{\mathcal{J}}^{\{i\}} \quad (134)$$

$$= w_{[K]\setminus\mathcal{J}\cup\{i\}} - \sum_{\mathcal{J}:i\in\mathcal{J}\subset\mathcal{J}} \sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}\setminus\{i\}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus\mathcal{J}\cup\{i\}\cup\mathcal{H}} \quad (135)$$

$$= w_{[K]\setminus\mathcal{J}\cup\{i\}} - \sum_{\mathcal{J}:i\in\mathcal{J}\subseteq\mathcal{J}} \sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}\setminus\{i\}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus\mathcal{J}\cup\{i\}\cup\mathcal{H}} + \sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}\setminus\{i\}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus\mathcal{J}\cup\{i\}\cup\mathcal{H}} \quad (136)$$

$$= w_{[K]\setminus\mathcal{J}\cup\{i\}} - \sum_{\mathcal{J}:\mathcal{J}\subseteq\mathcal{J}\setminus\{i\}} \sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus\mathcal{J}\cup\mathcal{H}} + \sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}\setminus\{i\}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus\mathcal{J}\cup\{i\}\cup\mathcal{H}} \quad (137)$$

$$= w_{[K]\setminus\mathcal{J}\cup\{i\}} - w_{[K]\setminus(\mathcal{J}\setminus\{i\})} + \sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}\setminus\{i\}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus\mathcal{J}\cup\{i\}\cup\mathcal{H}} \quad (138)$$

$$= \sum_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{J}\setminus\{i\}} (-1)^{|\mathcal{J}^c|} w_{[K]\setminus\mathcal{J}\cup\{i\}\cup\mathcal{H}}, \quad (139)$$

where (138) is from Lemma 4.

### B. Existence of the permutation

In this section, we prove that the worst user under the one-sided fair rate vector is determined by (88), namely

$$\arg \max_{k \in \mathcal{J}} t_{\mathcal{J}}^{\{k\}} = \min\{\mathcal{J}\} \quad , \forall \mathcal{J} \subseteq [K]. \quad (140)$$

We set  $m = \min(\mathcal{J})$  for any subset  $\mathcal{J} \subseteq [K]$  such that  $|\mathcal{J}| \geq 2$ . Proving (88) is equivalent to prove

$$R_m g_{\mathcal{J}}^{\{m\}} \geq R_i g_{\mathcal{J}}^{\{i\}} \quad \forall i \in \mathcal{J}. \quad (141)$$

Recall that from our one-sided rate vector assumption we have for  $i \in \mathcal{J}$ ,  $\delta_m \geq \delta_i$ ;  $\delta_m R_m \geq \delta_i R_i$  and  $\frac{\bar{p}_m}{p_m} R_m \geq \frac{\bar{p}_i}{p_i} R_i$ . Plugging (6) and (48) into (85), we obtain

$$g_{\mathcal{J}}^{\{i\}} = \frac{1}{1 - \delta_{[K] \setminus \mathcal{J} \cup \{i\}}} \left[ \sum_{\mathcal{J}: i \in \mathcal{J} \subset \mathcal{J}} g_{\mathcal{J}}^{\{i\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}} + p_{\mathcal{J} \setminus \{i\}} \bar{p}_{[K] \setminus \mathcal{J} \cup \{i\}} \right], \quad (142)$$

and

$$g_{\mathcal{J}}^{\{m\}} = \frac{1}{1 - \delta_{[K] \setminus \mathcal{J} \cup \{m\}}} \left[ \sum_{\mathcal{J}: m \in \mathcal{J} \subset \mathcal{J}} g_{\mathcal{J}}^{\{m\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}} + p_{\mathcal{J} \setminus \{m\}} \bar{p}_{[K] \setminus \mathcal{J} \cup \{m\}} \right]. \quad (143)$$

We prove by induction on  $|\mathcal{J}|$  that  $R_m g_{\mathcal{J}}^{\{m\}} \geq R_i g_{\mathcal{J}}^{\{i\}}$ : For  $|\mathcal{J}| = 2$ ,  $\mathcal{J} = \{m, i\}$  hence (142) and (143) imply the following

$$g_{\mathcal{J}}^{\{i\}} = \frac{1}{1 - \delta_{[K] \setminus \mathcal{J} \cup \{i\}}} \left[ g_i^{\{i\}} \bar{\delta}_m \delta_{[K] \setminus \mathcal{J} \cup \{i\}} + p_m \bar{p}_{[K] \setminus \mathcal{J} \cup \{i\}} \right], \quad (144)$$

and

$$g_{\mathcal{J}}^{\{m\}} = \frac{1}{1 - \delta_{[K] \setminus \mathcal{J} \cup \{m\}}} \left[ g_m^{\{m\}} \bar{\delta}_i \delta_{[K] \setminus \mathcal{J} \cup \{m\}} + p_i \bar{p}_{[K] \setminus \mathcal{J} \cup \{m\}} \right]. \quad (145)$$

Since  $\delta_m \geq \delta_i$ , it holds  $\frac{1}{1 - \delta_{[K] \setminus \mathcal{J} \cup \{m\}}} \geq \frac{1}{1 - \delta_{[K] \setminus \mathcal{J} \cup \{i\}}}$  and  $\bar{\delta}_i \geq \bar{\delta}_m$ . Since  $\frac{\bar{p}_m}{p_m} R_m \geq \frac{\bar{p}_i}{p_i} R_i$ , then it holds  $p_i \bar{p}_{[K] \setminus \mathcal{J} \cup \{m\}} R_m \geq p_m \bar{p}_{[K] \setminus \mathcal{J} \cup \{i\}} R_i$ . In addition we have from (87):  $g_m^{\{m\}} = g_i^{\{i\}} = \frac{\bar{p}_{[K]}}{1 - \delta_{[K]}}$  and  $\delta_m R_m \geq \delta_i R_i$ , thus we obtain  $R_m g_{\mathcal{J}}^{\{m\}} \geq R_i g_{\mathcal{J}}^{\{i\}}$  for  $|\mathcal{J}| = 2$ .

Suppose that (141) holds for any  $\mathcal{J} \subseteq [K]$  such that  $|\mathcal{J}| < |\mathcal{J}|$  and we prove that it holds also for  $\mathcal{J}$  in the following.

Since  $\delta_m \geq \delta_i$ , it holds  $\frac{1}{1 - \delta_{[K] \setminus \mathcal{J} \cup \{m\}}} \geq \frac{1}{1 - \delta_{[K] \setminus \mathcal{J} \cup \{i\}}}$ . Since  $\frac{\bar{p}_m}{p_m} R_m \geq \frac{\bar{p}_i}{p_i} R_i$ , it holds  $p_{\mathcal{J} \setminus \{m\}} \bar{p}_{[K] \setminus \mathcal{J} \cup \{m\}} R_m \geq p_{\mathcal{J} \setminus \{i\}} \bar{p}_{[K] \setminus \mathcal{J} \cup \{i\}} R_i$ . By observing (143) and (142), it remains to prove that

$$R_m \sum_{\mathcal{J}: m \in \mathcal{J} \subset \mathcal{J}} g_{\mathcal{J}}^{\{m\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}} \geq R_i \sum_{\mathcal{J}: i \in \mathcal{J} \subset \mathcal{J}} g_{\mathcal{J}}^{\{i\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}}. \quad (146)$$

We have for user  $m$

$$\sum_{\mathcal{J}: m \in \mathcal{J} \subset \mathcal{J}} g_{\mathcal{J}}^{\{m\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}} = \sum_{\mathcal{J}: \{m, i\} \subseteq \mathcal{J} \subset \mathcal{J}} g_{\mathcal{J}}^{\{m\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}} + \sum_{m \in \mathcal{J} \subset \mathcal{J} \setminus \{i\}} g_{\mathcal{J}}^{\{m\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}} \quad (147)$$

$$= \sum_{\mathcal{J}: \{m, i\} \subseteq \mathcal{J} \subset \mathcal{J}} g_{\mathcal{J}}^{\{m\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}} + \sum_{\mathcal{J}: \mathcal{J} \subset \mathcal{J} \setminus \{i, m\}} g_{\mathcal{J} \cup \{m\}}^{\{m\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J} \cup \{m\}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}}, \quad (148)$$

and similarly for user  $i$

$$\sum_{j:i \in \mathcal{J} \subset \mathcal{J}} g_j^{\{i\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}} = \sum_{\mathcal{J}:\{m,i\} \subseteq \mathcal{J} \subset \mathcal{J}} g_j^{\{i\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}} + \sum_{\mathcal{J}:i \in \mathcal{J} \subset \mathcal{J} \setminus \{m\}} g_j^{\{i\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}} \quad (149)$$

$$= \sum_{\mathcal{J}:\{m,i\} \subseteq \mathcal{J} \subset \mathcal{J}} g_j^{\{i\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}} + \sum_{\mathcal{J}:\mathcal{J} \subset \mathcal{J} \setminus \{m,i\}} g_{\mathcal{J} \cup \{i\}}^{\{i\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}}. \quad (150)$$

For any  $\mathcal{J}$  satisfying  $\{m, i\} \subseteq \mathcal{J} \subset \mathcal{J}$  we have  $|\mathcal{J}| < |\mathcal{J}|$ ,  $\min(\mathcal{J}) = m$  and  $i \in \mathcal{J}$  so by the hypothesis we have  $g_j^{\{m\}} R_m \geq g_j^{\{i\}} R_i$ . In addition we have  $\delta_m \geq \delta_i$  thus

$$\sum_{\mathcal{J}:\{m,i\} \subseteq \mathcal{J} \subset \mathcal{J}} g_j^{\{m\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}} R_m \geq \sum_{\mathcal{J}:\{m,i\} \subseteq \mathcal{J} \subset \mathcal{J}} g_j^{\{i\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}} R_i. \quad (151)$$

For any  $\mathcal{J}$  satisfying  $\mathcal{J} \subset \mathcal{J} \setminus \{m, i\}$  we have from (87)  $g_{\mathcal{J} \cup \{m\}}^{\{m\}} = g_{\mathcal{J} \cup \{i\}}^{\{i\}}$ . In addition we have  $\bar{\delta}_i \geq \bar{\delta}_m$  and  $R_m \delta_m \geq R_i \delta_i$ , then  $\bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}} R_m \geq \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}} R_i$ . As a result we obtain

$$R_m \sum_{\mathcal{J} \subset \mathcal{J} \setminus \{i,m\}} g_{\mathcal{J} \cup \{m\}}^{\{m\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{m\}} \geq R_i \sum_{\mathcal{J} \subset \mathcal{J} \setminus \{m,i\}} g_{\mathcal{J} \cup \{i\}}^{\{i\}} \bar{\delta}_{\mathcal{J} \setminus \mathcal{J}} \delta_{[K] \setminus \mathcal{J} \cup \{i\}}. \quad (152)$$

Hence the proof is completed.

### C. The outer-bound under the one-sided fair rate vector

Suppose that there exists  $\pi_1$  such that  $\sum_{j=1}^K R_{\pi_1(j)} w_{\pi_1(1) \dots \pi_1(j)} \leq 1$  and that  $\pi_1(i) \leq \pi_1(i+1)$  holds for some  $i \in [K-1]$ . We prove that for any permutation  $\pi_2$  that satisfies  $\pi_2(i+1) = \pi_1(i) = k$ ,  $\pi_2(i) = \pi_1(i+1) = k'$  and  $\pi_1(j) = \pi_2(j) \forall j \in [K] \setminus \{i, i+1\}$ , it holds  $\sum_{j=1}^K R_{\pi_2(j)} w_{\pi_2(1) \dots \pi_2(j)} \leq 1$ . It suffices to show that

$$w_{\pi_1(1) \dots \pi_1(i)} R_{\pi_1(i)} + w_{\pi_1(1) \dots \pi_1(i+1)} R_{\pi_1(i+1)} \geq w_{\pi_2(1) \dots \pi_2(i)} R_{\pi_2(i)} + w_{\pi_2(1) \dots \pi_2(i+1)} R_{\pi_2(i+1)}$$

equivalent to

$$(w_{\pi_1(1) \dots \pi_1(i)} - w_{\pi_2(1) \dots \pi_2(i+1)}) R_{\pi_1(i)} \geq (w_{\pi_2(1) \dots \pi_2(i)} - w_{\pi_1(1) \dots \pi_1(i+1)}) R_{\pi_1(i+1)}$$

equivalent to

$$(w_{\mathcal{J}k} - w_{\mathcal{J}kk'}) R_k \geq (w_{\mathcal{J}k'} - w_{\mathcal{J}kk'}) R_{k'}, \quad (153)$$

where  $\mathcal{J} = \pi_1(1) \dots \pi_1(i-1)$ . By replacing the weight by its expression (121) we obtain

$$w_{\mathcal{J}k} - w_{\mathcal{J}kk'} = \frac{\bar{p}_{\mathcal{J}k}}{1 - \delta_{\mathcal{J}k}} - \frac{\bar{p}_{\mathcal{J}kk'}}{1 - \delta_{\mathcal{J}kk'}} \quad (154)$$

$$= \bar{p}_{\mathcal{J}k} \left[ \frac{1}{1 - \delta_{\mathcal{J}k}} - \frac{1}{1 - \delta_{\mathcal{J}kk'}} + \frac{p_{k'}}{1 - \delta_{\mathcal{J}kk'}} \right] \quad (155)$$

$$= \bar{p}_{\mathcal{J}k} \left[ \frac{(1 - \delta_{\mathcal{J}kk'}) - (1 - \delta_{\mathcal{J}k})}{(1 - \delta_{\mathcal{J}k})(1 - \delta_{\mathcal{J}kk'})} + \frac{p_{k'}}{1 - \delta_{\mathcal{J}kk'}} \right] \quad (156)$$

$$= \frac{\bar{p}_{\mathcal{J}k}}{1 - \delta_{\mathcal{J}kk'}} \left[ \frac{\delta_{\mathcal{J}k}(1 - \delta_{k'})}{(1 - \delta_{\mathcal{J}k})} + p_{k'} \right], \quad (157)$$

and similarly

$$w_{\mathcal{J}k'} - w_{\mathcal{J}kk'} = \frac{\bar{p}_{\mathcal{J}k'}}{1 - \delta_{\mathcal{J}k'}} \left[ \frac{\delta_{\mathcal{J}k'}(1 - \delta_k)}{(1 - \delta_{\mathcal{J}k'})} + p_k \right]. \quad (158)$$

Thus, (153) is equivalent to

$$\frac{\delta_{\mathcal{J}}(1 - \delta_{k'})}{(1 - \delta_{\mathcal{J}k})} \bar{p}_k \delta_k R_k - \frac{\delta_{\mathcal{J}}(1 - \delta_k)}{(1 - \delta_{\mathcal{J}k'})} \bar{p}_{k'} \delta_{k'} R_{k'} + (\bar{p}_k p_{k'} R_k - \bar{p}_{k'} p_k R_{k'}) \geq 0. \quad (159)$$

Since  $k \leq k'$  then  $\delta_k \geq \delta_{k'}$ , so it is sufficient to prove that

$$\frac{\delta_{\mathcal{J}}(1 - \delta_k)}{(1 - \delta_{\mathcal{J}k'})} \underbrace{[\bar{p}_k \delta_k R_k - \bar{p}_{k'} \delta_{k'} R_{k'}]}_A + \underbrace{(\bar{p}_k p_{k'} R_k - \bar{p}_{k'} p_k R_{k'})}_B \geq 0. \quad (160)$$

This is satisfied if  $A \geq 0$  and  $B \geq 0$ . The condition B holds thanks to the definition of one-sided fair rate vector, and it is equivalent to

$$\frac{R_{k'}}{R_k} \leq \frac{\bar{p}_k p_{k'}}{\bar{p}_{k'} p_k} \triangleq \theta. \quad (161)$$

We will examine condition A by considering the case  $p_{k'} \geq p_k$  and  $p_k \geq p_{k'}$  separately.

- Case  $\theta > 1$

In this case we have  $p_k < p_{k'}$ , or  $\bar{p}_k > \bar{p}_{k'}$ . Condition A reduces to:

$$\delta_k R_k - \delta_{k'} R_{k'} \geq 0.$$

- Case  $\theta < 1$

In this case we have  $p_k > p_{k'}$  or  $\bar{p}_k < \bar{p}_{k'}$ . Then we have

$$\frac{R_{k'}}{R_k} \leq \frac{\bar{p}_k p_{k'}}{\bar{p}_{k'} p_k} \leq \frac{\bar{p}_k \delta_k}{\bar{p}_{k'} \delta_{k'}} \leq \frac{\delta_k}{\delta_{k'}}.$$

This means that B implies A so that the desired inequality holds once B holds. Since A is inactive, we can then consider a looser bounds

$$\delta_k R_k - \delta_{k'} R_{k'} \geq 0,$$

which holds by the definition of one-sided fair rate vector.

Thus we obtain the result. Starting by  $\pi_1$  as the identity we can obtain all the remaining  $K! - 1$  permutations.

## REFERENCES

- [1] M. Maddah-Ali and U. Niesen, "Fundamental Limits of Caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, 2014.
- [2] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, G. Caire, "FemtoCaching: Wireless Video Content Delivery Through Distributed Caching Helpers", *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, 2013.
- [3] M. Ji, G. Caire, A. Molisch, "Fundamental Limits of Distributed Caching in D2D Wireless Networks", arXiv:1304.5856, 2013.
- [4] M. Ji, A. Tulino, J. Llorca, and G. Caire, "Order-Optimal Rate of Caching and Coded Multicasting with Random Demands", arXiv:1502.03124, 2015.
- [5] M. Maddah-Ali and U. Niesen, "Decentralized Coded Caching Attains Order-Optimal Memory-Rate Tradeoff", *IEEE/ACM Trans. on Networking*, vol. 23, no. 4, pp. 1029–1040, 2015.
- [6] M. Maddah-Ali and U. Niesen, "Coded Caching with Nonuniform Demands", in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM)*, Toronto, Canada, 2014, <http://arxiv.org/abs/1308.0178v3>, 2015.

- [7] R. Pedarsani, M. Maddah-Ali, and U. Niesen, "Online Coded Caching", <http://arxiv.org/abs/1311.3646>, 2013.
- [8] W. Huang, S. Wang, L. Ding, F. Yang, and W. Zhang, "The Performance Analysis of Coded Cache in Wireless Fading Channel", arXiv:1504.01452v1, 2015.
- [9] R. Timo and M. Wigger, "Joint Cache-Channel Coding over Erasure Broadcast Channels", arXiv:1505.01016, 2015.
- [10] J. Hachem, N. Karamchandani, and S. Diggavi, "Effect of Number of Users in Multi-Level Coded Caching", in *Proceedings of the IEEE International Symposium on Information Theory (ISIT'2015)*, Hong-Kong, China, 2015.
- [11] J. Zhang, X. Lin, C. C. Wang, and X. Wang, "Coded Caching for Files with Distinct File Sizes", in *Proceedings of the IEEE International Symposium on Information Theory (ISIT'2015)*, Hong-Kong, China, 2015.
- [12] S. Yang and M. Kobayashi, "Secrecy Communications in  $K$ -User Multi-Antenna Broadcast Channel with State Feedback", in *Proceedings of the IEEE International Symposium on Information Theory (ISIT'2015)*, Hong-Kong, China, 2015.
- [13] P. Piantanida, M. Kobayashi, and G. Caire, "Analog Index Coding Over Block-Fading MISO Broadcast Channels with Feedback", in *Proceedings of the IEEE Information Theory Workshop (ITW)*, 2013, Seville, Spain, 2013.
- [14] M. A. Maddah-Ali and D. N. C. Tse, "Completely Stale Transmitter Channel State Information is Still Very Useful," *IEEE Trans. Inf. Theory* vol. 58, no. 7, pp. 4418–4431, July 2012.
- [15] C. C. Wang, "On the Capacity of 1-to-Broadcast Packet Erasure Channels with Channel Output feedback", *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 931–956, February 2012.
- [16] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser Broadcast Erasure Channel With Feedback-Capacity and Algorithms", *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5779–5804, September 2013.
- [17] S. Wang, W. Li, X. Tian, and H. Liu, "Coded Caching with Heterogenous Cache Sizes", arXiv preprint arXiv:1504.01123v3, 2015.
- [18] A. Ghorbel, M. Kobayashi, and S. Yang, "Cache-Enabled Broadcast Packet Erasure Channels with State Feedback", in *Proceedings of the 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, IL, USA, 2015.
- [19] N. Ajaykrishnan, N. Prem, S. Prabhakaran, M. V., and R. Vaze, "Critical Database Size for Effective Caching", arXiv preprint arXiv:1501.02549, 2015.
- [20] S. Karthikeyan, M. Ji, A. Tulino, J. Llorca and A. Dimakis "Finite Length Analysis of Caching-Aided Coded Multicasting", in *Proceedings of the 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, IL, USA, 2014
- [21] M. Ji, G. Caire, and A. Molisch. "Fundamental Limits of Distributed Caching in D2D Wireless Networks", in *Proceedings of the IEEE Information Theory Workshop (ITW)*, 2013, Seville, Spain, 2013.
- [22] N. Karamchandani, U. Niesen, M. Maddah-Ali, and S. Diggavi "Hierarchical Coded Caching.", in *Proceedings of the IEEE International Symposium on Information Theory (ISIT'2014)*, HI, USA, 2014.
- [23] J. Hachem, N. Karamchandani, and S. Diggavi, "Coded Caching for Heterogeneous Wireless Networks with Multi-Level Access" arXiv preprint arXiv:1404.6560v2, 2015.
- [24] A. El Gamal and Y. H. Kim, "Network Information Theory", Cambridge university press, 2011.
- [25] J. Zhang, F. Engelmann, and P. Elia "Coded Caching for Reducing CSIT-Feedback in Wireless Communications", in *Proceedings of the 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, IL, USA, 2015.
- [26] J. Zhang, and P. Elia, "Fundamental Limits of Cache-Aided Wireless BC: Interplay of Coded-Caching and CSIT Feedback", arXiv preprint arXiv:1511.03961, 2015.