Cache-Aided Data Delivery over Erasure Broadcast Channels

Mohammad Mohammadi Amiri and Deniz Gündüz
Electrical and Electronic Engineering Department, Imperial College London
Email: {m.mohammadi-amiri15, d.gunduz}@imperial.ac.uk

***Abstract*—A cache-aided erasure broadcast channel is studied. The receivers are divided into two sets: the *weak* and *strong* receivers, where the receivers in the same set all have the same erasure probability. The weak receivers, in order to compensate for the high erasure probability, are equipped with cache memories of equal size, while the receivers in the strong set have no caches. Data can be pre-delivered to weak receivers' caches over the off-peak traffic period before the receivers reveal their demands. A joint caching and channel coding scheme is proposed such that all the receivers, even the receivers without any cache memories, benefit from the presence of caches across the network. The trade-off between the cache size and the achievable rate is studied, and it is shown that the proposed scheme significantly improves the achievable trade-off upon the state-of-the-art.**

I. Introduction

Wireless content caching at the network edge is a promising technique to flatten data traffic over the backhaul network by shifting it from the peak to off-peak traffic periods [1]. Video-on-demand services for mobile users would particularly benefit from content caching due to a highly skewed popularity distribution across files. Contents that are likely to be requested by the users can be pre-placed into caches at the network edge during a period of low network traffic, known as the *placement phase*. The *delivery phase* is performed during a peak traffic period, when the receivers reveal their demands, and the cached contents can be used to reduce the load over the backhaul links.

An information-theoretic approach to the caching problem has been initiated by Maddah-Ali and Niesen in [1], [2]. The authors consider the delivery of requested contents from a library of same-size files over a noiseless shared channel, where the receivers are equipped with cache memories of equal capacity. They characterize a trade-off between the cache capacity and the minimum rate required during the delivery phase to serve all the receivers for all demand combinations; and show that, coding can significantly reduce the required delivery rate. Several improved coded caching schemes for the above setting have been introduced since then [3]–[9].

In contrast to the above setting [1], a noisy channel is considered for the delivery phase in [10]–[14]. Here, we follow the model considered in [11], [12], and assume that the delivery phase takes place over a memoryless packet erasure broadcast channel, and the receivers are grouped into two disjoint sets of *weak* and *strong* receivers. The receivers in each set all have the same erasure probability, that is, all the weak receivers have statistically worse channels compared to all the strong receivers. To compensate for the worse channel quality, each weak receiver is equipped with a cache memory of equal size. Assuming equal-rate files in the library, our goal is to characterize the trade-off between the size of the caches provided to the weak receivers and the rate of the files in the library, such that any demand combination can be reliably satisfied over the erasure broadcast channel. We propose a novel joint caching and channel coding scheme, and show that, for a given cache capacity, the proposed scheme achieves a higher rate than that is achieved in [12].

The rest of this paper is organised as follows. We introduce the system model in Section II. In Section III, our main result is summarized and compared with the state-of-the-art. In Section IV, we illustrate the proposed scheme through an example. The proposed scheme is briefly described in Section V. We conclude the paper in Section VI.

II. System Model and Preliminaries

We consider a server with a library of N popular files $\{W_1, ..., W_N\}$. Each file W_i is distributed uniformly over the set[1] $\left[\lfloor 2^{nR} \rfloor\right]$, $\forall i \in [N]$, where R denotes the rate of each file, n is the number of channel uses during the delivery phase, and $\lfloor X \rfloor$ is the greatest integer less than or equal to X. There are K receivers, $\mathrm{Rx}_1, \dots, \mathrm{Rx}_K$, requesting files from the server, where $K \leq N$, and d_k represents the file requested by Rx_k, where $d_k \in [N]$, $\forall k \in [K]$. The server needs to deliver file W_{d_k} to Rx_k, $\forall k \in [K]$, simultaneously.

Following [12], the channel from the server to the receivers is modeled as a memoryless packet erasure broadcast channel. For each channel use, the server transmits an F-bit packet over the channel, and receiver Rx_k receives the transmitted packet correctly with probability $1 - \delta_k$, or the erasure symbol with probability δ_k, for $k \in [K]$.

There are two sets of receivers, and all the receivers in the same set have the same erasure probability. Without loss of generality, the first K_w receivers, $\mathrm{Rx}_1, \dots, \mathrm{Rx}_{K_w}$, have erasure probability δ_w, and the next $K_s = K - K_w$ receivers, $\mathrm{Rx}_{K_w+1}, \dots, \mathrm{Rx}_K$, have erasure probability δ_s, such that $\delta_s < \delta_w$. The first and the second set of receivers are called the *weak* and *strong* receivers, respectively. Since $\delta_s < \delta_w$, the channel quality of any strong receiver is statistically better than that of any weak receiver. To compensate, each weak receiver is equipped with a local cache memory of capacity nM bits.

[1] For two integers $K_1 < K_2$, we will denote the set $\{K_1, K_1 + 1, ..., K_2\}$ by $[K_1 : K_2]$. The set $[1 : K_2]$ will shortly be denoted by $[K_2]$.

Transmission is performed in two phases. It starts with the placement phase which takes place during the off-peak traffic period, and the caches of the weak receivers are filled by the server without knowing their future demands. Thus, only the weak receivers take part in the placement phase, and the contents of the cache of receiver Rx_k, for $k \in [K_w]$, at the end of this phase is denoted by Z_k. It is to be noted that, since the placement phase is performed over a low-congestion period, it is assumed that no erasure occurs during this phase. The delivery phase follows once the demands of the receivers are revealed to the server, which transmits a length-n codeword, corresponding to n channel uses, over the packet erasure broadcast channel. The weak receivers reconstruct their requested file using both the received channel output and their cache contents, while the strong receivers rely only on the channel output.

Definition 1. *An error occurs if $\hat{W}_k \neq W_{d_k}$ for any $k \in [K]$, where $\hat{W}_k$ is the reconstructed file at Rx_k, and the probability of error is defined as*

$$P_e \triangleq \max_{(d_1,\dots,d_K)\in[N]^K} \Pr\left\{ \bigcup_{k=1}^{K} \left\{ \hat{W}_k \neq W_{d_k} \right\} \right\}. \quad (1)$$

Definition 2. *For given cache capacity of nM bits at the weak receivers, a memory-rate pair (M, R) is said to be achievable in a cache-aided packet erasure broadcast channel, if for any arbitrary $\varepsilon > 0$, there exists an n large enough, such that $P_e < \varepsilon$, i.e., each receiver can decode its desired file with arbitrarily low probability of error. For given M, the capacity is defined as*

$$C(M) \triangleq \sup\{R : (M,R) \text{ is achievable}\}. \quad (2)$$

We note that the above capacity remains an open problem even when the delivery channel is an error-free shared bit pipe [7]. Here, our goal is to provide an achievable coding scheme that improves upon the state-of-the-art.

III. Achievable Rate-Memory Pairs

A coding scheme as well as an information theoretic upper bound for the above model are presented in [12]. Here, we propose a new coding scheme, which we call the *successive cache and channel coding (SCC)* scheme, which generalizes and improves upon the scheme in [12]. We present the (M, R) pairs achieved by this scheme in Theorem 1 below. The details of the scheme will be described in Section V.

Theorem 1. *For the cache-aided packet erasure broadcast channel described in Section II, a memory-rate pair $\left(M_{(p,q)}, R_{(p,q)}\right)$ is achievable for any $p \in [0 : K_w]$ and $q \in [p : K_w]$, where*

$$M_{(p,q)} \triangleq \frac{NF \sum_{i=p}^{q} \left(i\gamma(p,i)\right)}{\frac{K_w}{1-\delta_w} \sum_{i=p}^{q} \left(\frac{K_w - i}{i+1}\gamma(p,i)\right) + \frac{K_w K_s}{1-\delta_s}}, \quad (3a)$$

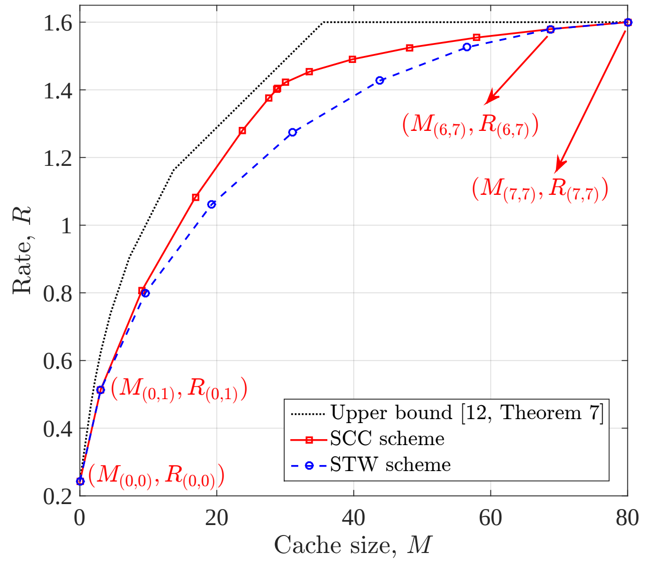

Fig. 1. Comparison of the achievable rates and the upper bound for $K_w = 7$, $K_s = 10$, $N = 50$, $F = 20$, $\delta_s = 0.2$, and $\delta_w = 0.9$.

$$R_{(p,q)} \triangleq \frac{F \sum_{i=p}^{q} \left(\gamma(p,i)\right)}{\frac{1}{1-\delta_w} \sum_{i=p}^{q} \left(\frac{K_w - i}{i+1}\gamma(p,i)\right) + \frac{K_s}{1-\delta_s}}, \quad (3b)$$

with $\gamma(p,i)$ defined as follows:

$$\gamma(p,i) \triangleq \frac{\binom{K_w}{i}}{\binom{K_w}{p} {K_s}^{i-p}} \left(\frac{1-\delta_s}{1-\delta_w} - 1\right)^{i-p}, \quad \text{for } i = p, \dots, q. \quad (3c)$$

The upper convex hull of these $(K_w + 1)(K_w + 2)/2$ memory-rate pairs can also be achieved through memory-sharing.

Next, we compare the achievable rate of the SCC scheme with the rate achieved by the scheme of [12], referred to as the STW scheme. In Fig. 1, we plot the achievable rates of both schemes for the scenario with $K_w = 7$, $K_s = 10$, $N = 50$, $F = 20$, $\delta_s = 0.2$, and $\delta_w = 0.9$. The upper bound on the capacity derived in [12, Theorem 7] is also included. Observe that, for relatively small cache sizes, where the best memory-rate trade-off is achieved by time-sharing between $\left(M_{(0,0)}, R_{(0,0)}\right)$ and $\left(M_{(0,1)}, R_{(0,1)}\right)$, and for relatively large cache sizes, where the best memory-rate trade-off is achieved by time-sharing between $\left(M_{(6,7)}, R_{(6,7)}\right)$ and $\left(M_{(7,7)}, R_{(7,7)}\right)$, both schemes achieve the same rate; however, the proposed SCC scheme achieves a higher rate for a large range of intermediate cache capacities, reducing the gap to the upper bound.

IV. Example

Consider the setting with $K_w = 3$ weak and $K_s = 2$ strong receivers. In the following, we investigate the achievable memory-rate pair $\left(M_{(0,2)}, R_{(0,2)}\right)$, given by (3) for $p = 0$ and $q = 2$, for an arbitrary demand combination $(d_1, \dots, d_5)$.

Each file W_d, $d \in [N]$, is divided into three subfiles $W_d^{(0)}$, $W_d^{(1)}$ and $W_d^{(2)}$, where subfile $W_d^{(i)}$ has rate $R^{(i)}$, given by

$$R^{(i)} \triangleq \frac{\gamma(0,i)}{\sum_{j=0}^{2}(\gamma(0,j))}R, \quad \text{for } i=0,1,2, \tag{4}$$

where $\gamma(0,i)$ is as defined in (3c). Therefore, we have

$$R^{(0)} + R^{(1)} + R^{(2)} = R. \tag{5}$$

In the placement phase, subfiles $W_1^{(i)}, ..., W_N^{(i)}$ are placed into the caches of $K_w = 3$ weak receivers through the procedure proposed in [1, Algorithm 1] for a cache capacity of iN/K_w, for $i = 0, 1, 2$. Each subfile $W_d^{(i)}$ is further divided into $\binom{3}{i}$ non-overlapping pieces, each at rate $R^{(i)}/\binom{3}{i}$.

$$W_d^{(i)} = \left\{ W_{d,\mathcal{S}}^{(i)} : \mathcal{S} \subset [3], |\mathcal{S}| = i \right\}, \forall d \in [N], \forall i \in [0:2]. \tag{6}$$

For the example under consideration, we have, $\forall d \in [N]$,

$$W_d^{(0)} = \left(W_{d,\emptyset}^{(0)} \right), \tag{7a}$$

$$W_d^{(1)} = \left(W_{d,\{1\}}^{(1)}, W_{d,\{2\}}^{(1)}, W_{d,\{3\}}^{(1)} \right), \tag{7b}$$

$$W_d^{(2)} = \left(W_{d,\{1,2\}}^{(2)}, W_{d,\{1,3\}}^{(2)}, W_{d,\{2,3\}}^{(2)} \right). \tag{7c}$$

The piece $W_{d,\mathcal{S}}^{(i)}$ is placed into the caches of receivers Rx_k, $\forall k \in [K_w] \cap \mathcal{S}$. Therefore, the cache contents of the weak receivers after the placement phase are as follows:

$$Z_1 = \bigcup_{d \in [N]} \left(W_{d,\{1\}}^{(1)}, W_{d,\{1,2\}}^{(2)}, W_{d,\{1,3\}}^{(2)} \right), \tag{8a}$$

$$Z_2 = \bigcup_{d \in [N]} \left(W_{d,\{2\}}^{(1)}, W_{d,\{1,2\}}^{(2)}, W_{d,\{2,3\}}^{(2)} \right), \tag{8b}$$

$$Z_3 = \bigcup_{d \in [N]} \left(W_{d,\{3\}}^{(1)}, W_{d,\{1,3\}}^{(2)}, W_{d,\{2,3\}}^{(2)} \right), \tag{8c}$$

where the required cache capacity for each weak receiver is:

$$M = \left(\frac{R^{(1)}}{3} + \frac{2R^{(2)}}{3} \right) N = \frac{\gamma(0,1) + 2\gamma(0,2)}{3\sum_{j=0}^{2}(\gamma(0,j))} RN. \tag{9}$$

The server tries to satisfy all the demands in the delivery phase with four messages transmitted in a time-division fashion, where the i-th message spans $\beta_i n$ channel uses, such that $\sum_{i=1}^{4} \beta_i = 1$.

The first message is targeted only for the weak receivers, and the missing subfiles of file $W_{d_k}^{(2)}$ are delivered to Rx_k, $\forall k \in [K_w]$. Exploiting the delivery phase algorithm of [1, Algorithm 1] for cache capacity $2N/K_w$, the following content is sent to the weak receivers:

$$W_{d_1,\{2,3\}}^{(2)} \oplus W_{d_2,\{1,3\}}^{(2)} \oplus W_{d_3,\{1,2\}}^{(2)}, \tag{10}$$

where $\oplus$ denotes the bitwise XOR operation. Observe that the content in (10) has rate $R^{(2)}/3$. The capacity region of the standard packet erasure broadcast channel [15] suggests that all the weak receivers can decode (10), for n large enough, if

$$\frac{R^{(2)}}{3(1-\delta_w)F} \leq \beta_1. \tag{11}$$

Having received (10), Rx_k, $k \in [K_w]$, can receive its missing piece $W_{d_k,\{1,2,3\}\setminus k}^{(2)}$ of the subfile $W_{d_k}^{(2)}$ using its cache contents Z_k. Thus, along with its cache content, Rx_k can recover subfile $W_{d_k}^{(2)}$.

With the second message of the delivery phase, the server delivers the contents

$$\left\{ W_{d_l,\mathcal{S}}^{(2)} : \mathcal{S} \subset \{1,2,3\}, |\mathcal{S}| = 2 \right\} \tag{12}$$

to strong receiver Rx_l, $l \in [K_w + 1 : K]$, and the missing contents

$$\left\{ W_{d_k,\mathcal{S}}^{(1)} : \mathcal{S} \subset \{1,2,3\} \setminus k, |\mathcal{S}| = 1 \right\} \tag{13}$$

to weak receiver Rx_k, $k \in [K_w]$. The missing contents in (13) are delivered to the weak receivers by using the delivery phase algorithm proposed in [1, Algorithm 1] corresponding to cache capacity N/K_w. Therefore, the following contents

$$W_{d_1,\{2\}}^{(1)} \oplus W_{d_2,\{1\}}^{(1)}, W_{d_1,\{3\}}^{(1)} \oplus W_{d_3,\{1\}}^{(1)}, W_{d_2,\{3\}}^{(1)} \oplus W_{d_3,\{2\}}^{(1)} \tag{14}$$

are transmitted to the weak receivers. The goal is to deliver the contents in (12) to the strong receivers, while simultaneously delivering the contents in (14) to the weak receivers. The transmission is performed over three orthogonal time periods. In the first period, contents

$$W_{d_4,\{1,2\}}^{(2)}, W_{d_5,\{1,2\}}^{(2)} \tag{15}$$

are delivered to the strong receivers Rx_4 and Rx_5, while

$$W_{d_1,\{2\}}^{(1)} \oplus W_{d_2,\{1\}}^{(1)} \tag{16}$$

is delivered to the weak receivers Rx_1 and Rx_2. Recall that, both receivers Rx_1 and Rx_2 already have the contents in (15) in their caches, which act as side information. Hence, the transmission can be carried out by using the joint encoding scheme of [16]. For notational convenience, we use

$$\mathrm{JE}\left((A_1)_{\mathcal{S}_1}, (A_2)_{\mathcal{S}_2} \right) \tag{17}$$

to represent the transmission of content A_1 to the receivers in $\mathcal{S}_1 \subset [K]$, and content A_2 to the receivers in set $\mathcal{S}_2 \subset [K]$ through joint encoding, where $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$, and A_2 is available at all the receivers in $\mathcal{S}_1$ as side information. Using the notation in (17), in the first part of the second message of the delivery phase, the transmission

$$\mathrm{JE}\left(\left(W_{d_1,\{2\}}^{(1)} \oplus W_{d_2,\{1\}}^{(1)} \right)_{\{1,2\}}, \left(W_{d_4,\{1,2\}}^{(2)}, W_{d_5,\{1,2\}}^{(2)} \right)_{\{4,5\}} \right) \tag{18}$$

is performed. In the second period, contents

$$W_{d_4,\{1,3\}}^{(2)}, W_{d_5,\{1,3\}}^{(2)}, \tag{19}$$

which are available in the caches of receivers Rx_1 and Rx_3

as side information, are delivered to Rx_4 and Rx_5, while

$$W_{d_1,\{3\}}^{(1)} \oplus W_{d_3,\{1\}}^{(1)} \tag{20}$$

is delivered to Rx_1 and Rx_3. Hence, the following transmission is performed:

$$\mathrm{JE}\left(\left(W_{d_1,\{3\}}^{(1)} \oplus W_{d_3,\{1\}}^{(1)}\right)_{\{1,3\}}, \left(W_{d_4,\{1,3\}}^{(2)}, W_{d_5,\{1,3\}}^{(2)}\right)_{\{4,5\}}\right). \tag{21}$$

Finally, in the third period, Rx_4 and Rx_5 aim to recover

$$W_{d_4,\{2,3\}}^{(2)}, W_{d_5,\{2,3\}}^{(2)}, \tag{22}$$

and Rx_2 and Rx_3, while having (22) as side information, try to decode

$$W_{d_2,\{3\}}^{(1)} \oplus W_{d_3,\{2\}}^{(1)}, \tag{23}$$

after receiving the transmission

$$\mathrm{JE}\left(\left(W_{d_2,\{3\}}^{(1)} \oplus W_{d_3,\{2\}}^{(1)}\right)_{\{2,3\}}, \left(W_{d_4,\{2,3\}}^{(2)}, W_{d_5,\{2,3\}}^{(2)}\right)_{\{4,5\}}\right). \tag{24}$$

Observe that, in each period, a message of rate $2R^{(2)}/3$, available at the weak receivers as side information, is transmitted to the strong receivers; while, simultaneously, a message at rate $R^{(1)}/3$ is transmitted to the weak receivers. Overall, the contents in (12) and (14) with a total rate of $2R^{(2)}$ and $R^{(1)}$ are delivered to the strong and weak receivers, respectively, over three orthogonal periods by using the joint encoding scheme of [16] that exploits the side information at the weak receivers. Using the achievable rate region of the joint encoding scheme for the packet erasure channels [12], contents in (12) and (14) can be simultaneously decoded by the strong and weak receivers, respectively, for n large enough, if

$$\max\left\{\frac{R^{(1)}}{(1-\delta_w)F}, \frac{R^{(1)}+2R^{(2)}}{(1-\delta_s)F}\right\} \leq \beta_2. \tag{25}$$

With the particular values for $R^{(1)}$ and $R^{(2)}$ specified in (4), it can be verified that the two terms in the maximization in (25) are equal. Thus, the condition in (25) can be written as

$$\frac{R^{(1)}}{(1-\delta_w)F} \leq \beta_2. \tag{26}$$

Observe that, Rx_1 can obtain the messages $W_{d_1,\{2\}}^{(1)}$ and $W_{d_1,\{3\}}^{(1)}$ after receiving (16) and (20), respectively. Rx_2 can recover $W_{d_2,\{1\}}^{(1)}$ and $W_{d_2,\{3\}}^{(1)}$ after receiving (16) and (23), respectively. Finally, decoding (20) and (23) allows Rx_3 to recover $W_{d_3,\{1\}}^{(1)}$ and $W_{d_3,\{2\}}^{(1)}$, respectively. Thus, at the end of the second message, weak receiver Rx_k, $k \in [K_w]$, has received all the missing bits of its requested subfile $W_{d_k}^{(1)}$, while strong receiver, Rx_l, $l \in [K_w+1:K]$ has recovered subfile $W_{d_l}^{(2)}$.

With the third message sent in the delivery phase, content

$$\left\{W_{d_l,\mathcal{S}}^{(1)} : \mathcal{S} \subset \{1,2,3\}, |\mathcal{S}| = 1\right\} \tag{27}$$

is delivered to strong receiver Rx_l, $l \in [K_w+1:K]$, while $W_{d_k,\emptyset}^{(0)}$ is delivered to weak receiver Rx_k, $k \in [K_w]$. The transmission takes place over three orthogonal time periods. In the first period,

$$W_{d_4,\{1\}}^{(1)}, W_{d_5,\{1\}}^{(1)}, \tag{28}$$

available locally at Rx_1 as side information, are delivered to Rx_4 and Rx_5, while $W_{d_1,\emptyset}^{(0)}$ is delivered to Rx_1. Therefore, the transmission can be performed through joint encoding

$$\mathrm{JE}\left(\left(W_{d_1,\emptyset}^{(0)}\right)_{\{1\}}, \left(W_{d_4,\{1\}}^{(1)}, W_{d_5,\{1\}}^{(1)}\right)_{\{4,5\}}\right). \tag{29}$$

In the second period, the goal is to send

$$W_{d_4,\{2\}}^{(1)}, W_{d_5,\{2\}}^{(1)} \tag{30}$$

to Rx_4 and Rx_5, while delivering $W_{d_2,\emptyset}^{(0)}$ to Rx_2, which has access to (30) as side information. Joint encoding is performed with

$$\mathrm{JE}\left(\left(W_{d_2,\emptyset}^{(0)}\right)_{\{2\}}, \left(W_{d_4,\{2\}}^{(1)}, W_{d_5,\{2\}}^{(1)}\right)_{\{4,5\}}\right). \tag{31}$$

In the last period, the following transmission is performed:

$$\mathrm{JE}\left(\left(W_{d_3,\emptyset}^{(0)}\right)_{\{3\}}, \left(W_{d_4,\{3\}}^{(1)}, W_{d_5,\{3\}}^{(1)}\right)_{\{4,5\}}\right), \tag{32}$$

where

$$W_{d_4,\{3\}}^{(1)}, W_{d_5,\{3\}}^{(1)} \tag{33}$$

are aimed for the strong receivers Rx_4 and Rx_5, and Rx_3 receives $W_{d_3,\emptyset}^{(0)}$ while having (33) as side information. With three orthogonal transmissions (29), (31) and (32), each corresponding to a submessage of the third message of the delivery phase, a message of rate $R^{(0)}$ is targeted for the weak receivers, while a message of rate $2R^{(1)}/3$, available locally at all the weak receivers, is aimed for the strong receivers. Therefore, through joint encoding, a total rate of $3R^{(0)}$ is delivered to the weak receivers, while the strong receivers receive a total rate of $2R^{(1)}$ over these three periods. All the weak and strong receivers can decode their messages, for n large enough, if

$$\max\left\{\frac{3R^{(0)}}{(1-\delta_w)F}, \frac{3R^{(0)}+2R^{(1)}}{(1-\delta_s)F}\right\} \leq \beta_3. \tag{34}$$

Again, from the expressions of $R^{(0)}$ and $R^{(1)}$ in (4), it can be verified that (34) can be simplified as

$$\frac{3R^{(0)}}{(1-\delta_w)F} \leq \beta_3. \tag{35}$$

Therefore, if (35) holds, each receiver Rx_k, $k \in [K_w]$, obtains $W_{d_k,\emptyset}^{(0)}$, while each receiver Rx_l, $l \in [K_w+1:K]$, obtains $W_{d_l}^{(1)}$. Thus, with the third message of the delivery phase, the demands of the weak receivers are fully satisfied.

The fourth and last message of the delivery phase is destined only to the strong receivers, and the goal is to deliver them the missing bits of their demands, in particular, the subfiles

$W^{(0)}_{d_k,\emptyset}$, which are of rate $R^{(0)}$. From the capacity region of the standard erasure broadcast channel [15], each receiver Rx_l, $l \in [K_w+1:K]$, can decode $W^{(0)}_{d_l,\emptyset}$ successfully for n sufficiently large, if

$$\frac{2R^{(0)}}{(1-\delta_s)\,F} \le \beta_4. \tag{36}$$

Combining (11), (26), (35), (36), and the fact that $\sum_{i=1}^{4} \beta_i = 1$, we obtain the condition

$$\frac{R^{(2)}}{3\,(1-\delta_w)\,F} + \frac{R^{(1)}}{(1-\delta_w)\,F} + \frac{3R^{(0)}}{(1-\delta_w)\,F} + \frac{2R^{(0)}}{(1-\delta_s)\,F} \le 1. \tag{37}$$

By replacing $R^{(i)}$ with the expressions from (4), for $i=0,1,2$, and using the fact that $\gamma(0,0)=1$, (37) corresponds to

$$R \le \frac{\sum_{j=0}^{2} (\gamma\,(0,j))F}{\frac{1}{1-\delta_w}\left(3+\gamma\,(0,1)+\frac{1}{3}\gamma\,(0,2)\right)+\frac{2}{1-\delta_s}}. \tag{38}$$

Observe that, the term in the right hand side of the inequality (38) is $R_{(0,2)}$, which is given by (3b). By setting $R = R_{(0,2)}$, the equivalent cache size M given by (9) is $M = M_{(0,2)}$, where $M_{(0,2)}$ is defined in (3a). Thus, the memory-rate pair $\left(M_{(0,2)},\ \ R_{(0,2)}\right)$ given by (3) is achievable for the setting under consideration.

V. The Successive Cache and Channel Coding (SCC) Scheme

In this section, we present the SCC scheme for a general setting, achieving the memory-rate pair $\left(M_{(p,q)}, R_{(p,q)}\right)$ given by (3), for any $p \in [0:K_w]$, and $q \in [p:K_w]$. For each $p \in [0:K_w]$, and $q \in [p:K_w]$, each file W_d, $\forall d$, is divided into $(q-p+1)$ non-overlapping subfiles, represented by

$$W_d = \left(W_d^{(p)}, ..., W_d^{(q)}\right), \tag{39}$$

where subfile $W_d^{(i)}$, for $i \in [p:q]$, is at rate

$$R^{(i)} \triangleq \frac{\gamma\,(p,i)}{\sum_{j=p}^{q} (\gamma\,(p,j))} R, \tag{40}$$

which satisfy

$$\sum_{i=p}^{q} R^{(i)} = R. \tag{41}$$

A. Placement Phase

In the placement phase, for each set of subfiles $W_1^{(i)}, ..., W_N^{(i)}$, $i = p, p+1, \ldots, q$, a cache placement procedure corresponding to the one proposed in [1, Algorithm 1] for a cache capacity of iN/K_w is performed; that is, each subfile $W_d^{(i)}$ is partitioned into $\binom{K_w}{i}$ independent equal-rate pieces,

$$W_d^{(i)} = \left(W_{d,\mathcal{S}}^{(i)} : \mathcal{S} \subset [K_w]\,, |\mathcal{S}| = i\right), \forall d \in [N]\,, \forall i \in [p:q]\,. \tag{42}$$

The piece $W_{d,\mathcal{S}}^{(i)}$ of rate $R^{(i)}/\binom{K_w}{i}$ is then cached by receivers Rx_k, $\forall k \in [K_w] \cap \mathcal{S}$. Thus, the content placed in the cache of each weak receiver Rx_k, $k \in [K_w]$, is given by

$$Z_k = \bigcup_{d\in[N]} \bigcup_{i\in[p:q]} \left\{W_{d,\mathcal{S}}^{(i)} : k \in \mathcal{S}, \mathcal{S} \subset [K_w]\,, |\mathcal{S}| = i\right\}. \tag{43}$$

Accordingly, $\binom{K_w-1}{i-1}$ pieces, each of rate $R^{(i)}/\binom{K_w}{i}$, corresponding to each subfile $W_d^{(i)}$, are cached by weak receiver Rx_k, $k \in [K_w]$, which requires a cache capacity of

$$M = N \sum_{i=p}^{q} \binom{K_w-1}{i-1} \frac{R^{(i)}}{\binom{K_w}{i}} = N \sum_{i=p}^{q} \frac{iR^{(i)}}{K_w}. \tag{44}$$

An example of the cache placement phase for $K_w=3$, $p=0$, and $q=2$ is given in (8).

B. Delivery Phase

In the delivery phase, the goal is to satisfy any arbitrary demand combination $(d_1, ..., d_K)$. The delivery phase is performed over $(q-p+2)$ orthogonal transmissions, where the i-th transmission is performed over $\beta_i n$ channel uses, for $i = 1, ..., q-p+2$, such that $\sum_{i=1}^{q-p+2} \beta_i = 1$.

The first message of the delivery phase is destined only for the weak receivers, and the server delivers the missing bits of subfile $W_{d_k}^{(q)}$ to receiver Rx_k, $k \in [K_w]$. The first message is obtained through the coded delivery procedure proposed in [1, Algorithm 1] for a cache capacity of qN/K_w. The coded contents

$$\left\{\bigoplus\nolimits_{k\in\mathcal{S}} W_{d_k,\mathcal{S}\backslash k}^{(q)} : \mathcal{S} \subset [K_w]\,, |\mathcal{S}| = q+1\right\} \tag{45}$$

are delivered to receivers Rx_k, $k \in [K_w]$. Having received (45) correctly along with their cached contents, weak receiver Rx_k, $k \in [K_w]$, can recover

$$\left\{W_{d_k,\mathcal{S}}^{(q)} : \mathcal{S} \subset [K_w] \backslash k, |\mathcal{S}| = q\right\}, \tag{46}$$

i.e., all the missing bits of subfile $W_{d_k}^{(q)}$. An example of the contents delivered with the first message of the delivery phase for $K_w = 3$ and $q = 2$ can be found in (10).

The delivery procedures performed for transmissions 2 to $q-p+1$ of the delivery phase are similar. With the message delivered in transmission i, the server delivers the content

$$\left\{W_{d_l,\mathcal{S}}^{(q-(i-2))} : \mathcal{S} \subset [K_w]\,, |\mathcal{S}| = q-(i-2)\right\} \tag{47}$$

to strong receiver Rx_l, $l \in [K_w+1:K]$, for $i = 2, ..., q-p+1$, while at the same time, delivering the missing bits of content

$$\left\{W_{d_k,\mathcal{S}}^{(q-(i-1))} : \mathcal{S} \subset [K_w]\,, |\mathcal{S}| = q-(i-1)\right\} \tag{48}$$

to weak receiver Rx_k, $k \in [K_w]$, through the coded delivery scheme proposed in [1, Algorithm 1] for a cache capacity of $(q-(i-1))N/K_w$. That is, the XOR-ed contents

$$\left\{\bigoplus\nolimits_{k\in\mathcal{S}} W_{d_k,\mathcal{S}\backslash k}^{(q-(i-1))} : \mathcal{S} \subset [K_w]\,, |\mathcal{S}| = q-(i-2)\right\} \tag{49}$$

are sent to weak receivers. Observe that, having received (47) correctly, strong receiver Rx_l, $l \in [K_w+1:K]$, can recover subfile $W_{d_l}^{(q-(i-2))}$. Furthermore, after receiving (49) correctly, weak receiver Rx_k, $k \in [K_w]$, can recover the missing bits of subfile $W_{d_k}^{(q-(i-1))}$ given by

$$\left\{W_{d_k,\mathcal{S}}^{(q-(i-1))} : \mathcal{S} \subset [K_w] \setminus k, |\mathcal{S}| = q-(i-1)\right\}. \qquad (50)$$

We note that, from (47) and (49), the number of contents targeted for the sets of strong and weak receivers are equal, and given by $\binom{K_w}{q-(i-2)}$. Thus, the i-th transmission of the delivery phase can be performed over $\binom{K_w}{q-(i-2)}$ orthogonal periods through time-sharing, each of which corresponds to a set $\mathcal{S} \subset [K_w]$ such that $|\mathcal{S}| = q-(i-2)$, for $i = 2, ..., q-p+1$. During each period of the i-th transmission of the delivery phase, corresponding to a set of receivers $\mathcal{S} \subset [K_w]$, the contents

$$\left\{W_{d_{K_w+1},\mathcal{S}}^{(q-(i-2))}, ..., W_{d_K,\mathcal{S}}^{(q-(i-2))}\right\}, \qquad (51)$$

which are available to all the weak receivers in $\mathcal{S}$ as side information, are delivered to the strong receivers, while

$$\left\{\bigoplus\nolimits_{k\in\mathcal{S}} W_{d_k,\mathcal{S}\setminus k}^{(q-(i-1))}\right\} \qquad (52)$$

is delivered to the weak receivers in set $\mathcal{S}$. This transmission is performed using the joint encoding scheme of [16]. After transmission $q-p+1$ of the delivery phase, strong receiver Rx_l can recover subfiles $\left(W_{d_l}^{(q)}, ..., W_{d_l}^{(p+1)}\right)$, for $l \in [K_w+1:K]$; while, weak receiver Rx_k, for $k \in [1:K_w]$, can recover subfiles $\left(W_{d_k}^{(q-1)}, ..., W_{d_k}^{(p)}\right)$. Hence, together with the contents delivered in the first transmission, the demands of the weak receivers are satisfied completely after transmission $q-p+1$ of the delivery phase, and each strong receiver Rx_l, $l \in [K_w+1:K]$, only requires subfile $W_{d_l}^{(p)}$.

In the last transmission of the delivery phase, only the strong receivers are targeted, and content $W_{d_l}^{(p)}$ is delivered to strong receiver Rx_l, $l \in [K_w+1:K]$. Observe that, together with the contents delivered through transmission 2 to $q-p+1$ of the delivery phase, the demand of each strong receiver is also satisfied.

It is illustrated in [17] that using the proposed SCC scheme, all the receiver demands $(d_1, ..., d_K) \in [N]^K$ can be satisfied with $(M, R) = (M_{(p,q)}, R_{(p,q)})$, for any $p \in [0:K_w]$ and $q \in [p:K_w]$, given by (3). Hence, the memory-rate pairs $(M_{(p,q)}, R_{(p,q)})$ given by (3) are achievable.

Remark 1. *We note that, the achievable STW scheme studied in [12] is a special case of the SCC scheme with $q = p+1$. The delivery phase of the STW scheme is performed over 3 orthogonal transmissions. The SCC scheme utilizes a more flexible caching and coding scheme which applies a finer subpacketization compared to [12], together with the joint encoding scheme of [16], which is also used in [12] enabling all the receivers to exploit the cache capacities of the weak receivers.*

VI. Conclusions

We have studied a cache-enabled packet erasure broadcast channel, over which a server delivers contents simultaneously to a set of receivers, each demanding one content from a finite library. We have considered two disjoint sets of receivers: a set of weak receivers with erasure probability δ_w, and a set of the strong receivers with erasure probability $\delta_s < \delta_w$. To compensate for the weak channel, each weak receiver is equipped with a cache memory of equal size. We have characterized a lower bound on the capacity which denotes the highest rate of the contents in the library that can be delivered reliably to all the receivers under any demand combination. Building upon the scheme proposed in [12], we have proposed an improved joint caching and channel coding scheme that achieves a higher rate by finer subpacketization of the files in the library, which enables each receiver, even the strong receivers without cache memories, to benefit further from the cache memories available at the weak receivers. This model illustrates that storage can be converted into spectral efficiency if it is placed strategically across the network, and exploited intelligently.

References

[1] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inform. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[2] ——, "Decentralized caching attains order optimal memory-rate trade-off," *IEEE/ACM Trans. Netw*, vol. 23, no. 4, pp. 1029–1040, Apr. 2014.

[3] Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: Improved bounds for users with small buffers," *IET Communications*, vol. 10, no. 17, pp. 2315–2318, Nov. 2016.

[4] M. Mohammadi Amiri and D. Gündüz, "Improved delivery rate-cache capacity trade-off for centralized coded caching," in *Proc. IEEE ISITA*, Monterey, CA, Oct. 2016, pp. 26–30.

[5] ——, "Fundamental limits of coded caching: Improved delivery rate-cache capacity trade-off," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 806–815, Feb. 2017.

[6] K. Wan, D. Tuninetti, and P. Piantanida, "On caching with more users than files," in *Proc. IEEE ISIT*, Barcelona, Spain, Jul. 2016, pp. 135–139.

[7] M. Mohammadi Amiri, Q. Yang, and D. Gündüz, "Coded caching for a large number of users," in *Proc. IEEE ITW*, Cambridge, UK, Sep. 2016, pp. 171–175.

[8] C. Tian and J. Chen, "Caching and delivery via interference elimination," in *Proc. IEEE ISIT*, Barcelona, Spain, Jul. 2016, pp. 830–834.

[9] J. Gomez-Vilardebo, "Fundamental limits of caching: Improved bounds with coded prefetching," *arXiv:1612.09071v2 [cs.IT]*, Jan. 2017.

[10] W. Huang, S. Wang, L. Ding, F. Yang, and W. Zhang, "The performance analysis of coded cache in wireless fading channel," *arXiv:1504.01452v1 [cs.IT]*, Apr. 2015.

[11] R. Timo and M. Wigger, "Joint cache-channel coding over erasure broadcast channels," in *Proc. IEEE ISWCS*, Brussels, Belgium, Aug. 2015, pp. 201–205.

[12] S. Saeedi Bidokhti, R. Timo, and M. Wigger, "Noisy broadcast networks with receiver caching," *arXiv:1605.02317v1 [cs.IT]*, May 2016.

[13] M. Maddah-Ali and U. Niesen, "Cache-aided interference channels," *arXiv: 1510.06121v1 [cs.IT]*, Oct. 2015.

[14] N. Naderializadeh, M. Maddah-Ali, and A. S. Avestimehr, "Fundamental limits of cache-aided interference management," *arXiv:1602.04207v1 [cs.IT]*, Feb. 2016.

[15] R. L. Urbanke and A. D. Wyner, "Packetizing for the erasure broadcast channel with an internet application," in *Proc. Int'l. Conf. Combinatorics, Inform. Theory and Stat.*, Portland, ME, May 1997, p. 93.

[16] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1469–1482, Apr. 2006.

[17] M. Mohammadi Amiri and D. Gündüz, "Cache-aided data delivery over erasure broadcast channels," *arXiv:1702.05454v1 [cs.IT]*, Feb. 2017.