Cache-Aided Content Delivery over Erasure Broadcast Channels

Mohammad Mohammadi Amiri and Deniz Gündüz

Abstract

A cache-aided broadcast network is studied, in which a server delivers contents to a group of receivers over a packet erasure broadcast channel (BC). The receivers are divided into two sets with regards to their channel qualities: the *weak* and *strong* receivers, where all the weak receivers have statistically worse channel qualities than all the strong receivers. The weak receivers, in order to compensate for the high erasure probability they encounter over the channel, 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. Allowing arbitrary erasure probabilities for the weak and strong receivers, a joint caching and channel coding scheme, which divides each file into several subfiles, and applies a different caching and delivery scheme for each subfile, is proposed. It is shown that all the receivers, even those without any cache memories, benefit from the presence of caches across the network. An information theoretic trade-off between the cache size and the achievable rate is formulated. It is shown that the proposed scheme improves upon the state-of-the-art in terms of the achievable trade-off.

I. INTRODUCTION

Wireless content caching is a promising technique to flatten the traffic over the backhaul network by shifting part of the traffic from peak to off-peak periods [1]–[5]. Video-on-demand services for mobile users would particularly benefit from content caching as a few highly popular files are requested by a large number of users over a relatively short time period. Popular contents that are likely to be requested by a majority of the users can be proactively cached 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 users reveal their demands, and the cached contents can be exploited to reduce both the load over the backhaul links and the latency in delivery [6].

A coded proactive content caching and delivery scheme has been proposed by Maddah-Ali and Niesen in [6], [7], where they consider a library of same-size files to be delivered over a noiseless broadcast channel (BC), while the receivers are equipped with cache memories of equal size. They identify a trade-off between the cache size 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 and information theoretic performance bounds have been introduced since then [8]–[15]. Coded caching has since been extended to various network settings, including device-to-device caching [16], [17], online cache placement [18], files with non-uniform popularities [19], [20] and distinct lengths [21], users with non-uniform cache sizes [22], [23], multi-layer caching [24], and caching by users with different distortion requirements [25]–[27].

In contrast to the setting introduced in [6], a noisy channel is considered for the delivery phase in [28]–[35]. Here, we follow the model considered in [31], and assume that the delivery phase takes place over a memoryless packet erasure BC, which models a packetized communication system, where each packet is separately channel coded against errors at the physical layer, so that a packet either arrives at a receiver correctly, or is lost. Communication over Internet is usually modeled as a packet erasure channel. The receivers in the system are grouped into two disjoint sets of weak and strong receivers. All the weak receivers are assumed to have statistically worse channels than the strong receivers, while the users in each set can have arbitrary erasure probabilities. To compensate for the worse channel quality, each weak receiver is equipped with a cache memory of equal size. We consider the case when the number of receivers is not greater than the number of files in the library. Assuming equal-rate files in the library, we derive a trade-off between the size of the caches provided to the weak receivers and the rate of the files, for which any demand combination can be reliably satisfied over the erasure BC. The proposed scheme exploits a novel file subpacketization, and performs a different caching and content delivery scheme to deliver different subpackets over the channel. Moreover, the delivery of the contents to the weak and strong receivers are coupled through the use of a joint encoding scheme to maximally benefit from the available cache memories. We show that, when specified to the homogeneous scenario considered in [31], where the receivers in the same set have the same erasure probability, the proposed scheme outperforms the one in [31].

The rest of this paper is organized as follows. We introduce the system model in Section II. Main results are summarized and compared with the state-of-the-art in Section III. The proposed
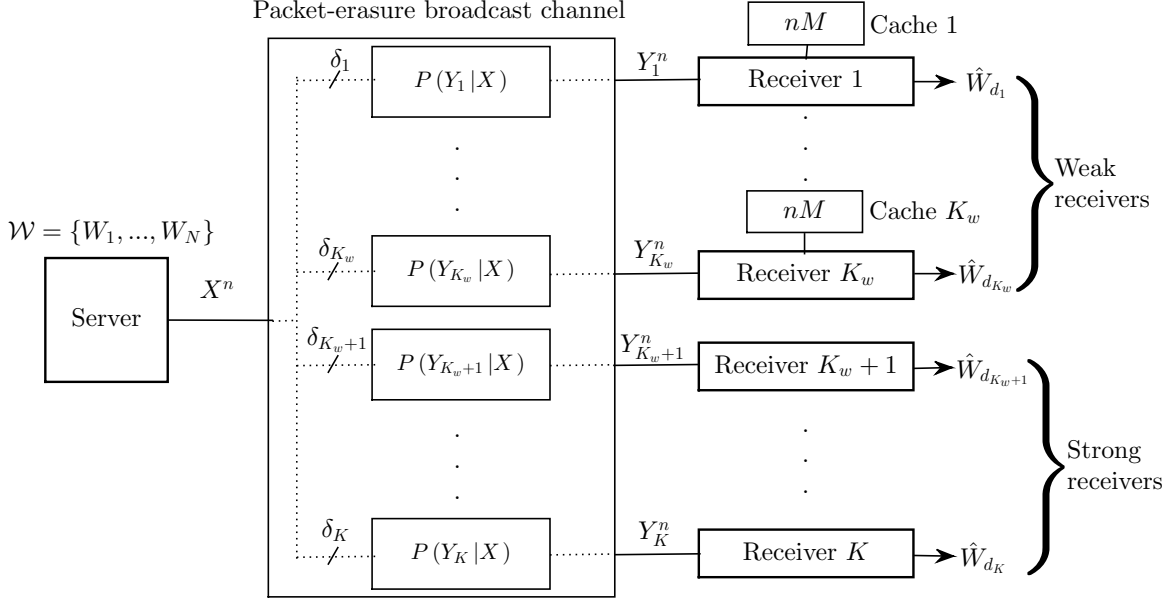
Fig. 1. Cache-aided packet erasure BC. The first K_w receivers have statically worse channels than the last K_s receivers, but each of them is equipped with a cache of normalized size M.

scheme is elaborated and analyzed in Section IV. We conclude the paper in Section V.

Notations: For two integers $i \leq j$, the set $\{i, i+1, ..., j\}$ is denoted by $[i : j]$, and the set $[1 : i]$ is shortly denoted by $[i]$. For two sets $\mathcal{Q}$ and $\mathcal{P}$, $\mathcal{Q} \backslash \mathcal{P}$ is the set of elements in $\mathcal{Q}$ that do not belong to $\mathcal{P}$. Notation $|\cdot|$ represents the cardinality of a set, or the length of a vector. Notation $\oplus$ refers to bitwise XOR operation; and finally, $\binom{j}{i}$ returns the binomial coefficient "j choose i".

II. SYSTEM MODEL AND PRELIMINARIES

We consider a server with a library of N popular files $\mathbf{W} \triangleq (W_1, ..., W_N)$. Each file W_f is distributed uniformly over the set $\left[\left\lceil 2^{nR} \right\rceil\right]$, $\forall f \in [N]$, where R denotes the rate of each file, and n is the number of channel uses during the delivery phase. Receiver k's demand is represented by d_k, where $d_k \in [N]$, $\forall k \in [K]$. The server delivers all the requests $W_{d_1}, ..., W_{d_K}$ to their corresponding receivers simultaneously over a BC.

Following [31], the channel between the server and the receivers is modeled as a memoryless packet erasure BC. For each channel use, the server transmits an F-bit codeword from the alphabet $\mathcal{X} \triangleq \{0, 1\}^F$, and the output alphabet at each receiver is $\mathcal{Y} \triangleq \mathcal{X} \cup \{\Delta\}$, where the erasure symbol Δ corresponds to a packet that is not received at the receiver. Receiver k, for $k \in [K]$, receives the transmitted codeword $x \in \mathcal{X}$ correctly with probability $1 - \delta_k$, and the

erasure symbol Δ with probability δ_k. Thus, given the transmitted codeword $x \in \mathcal{X}$, receiver $k \in [K]$ observes the output $y_k \in \mathcal{Y}$ with the conditional probability

$$P\left(Y_k = y_k \,|X = x\right) = \begin{cases} 1 - \delta_k, & \text{if } y_k = x, \\ \delta_k, & \text{if } y_k = \Delta. \end{cases} \tag{1}$$

Two disjoint sets of receivers, *weak* and *strong* receivers, are considered, grouped according to the erasure probabilities of their channels. These groups may model users located in areas with relatively bad and good network coverage, respectively. We assume that the channel condition of each strong receiver is statistically better than that of each weak receiver; that is, the erasure probability of a strong receiver is lower than any weak receiver. Without loss of generality, we enumerate the receivers in the order of improving channel quality, that is, we have $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_K$. We denote the set of erasure probabilities by $\boldsymbol{\delta} \triangleq \{\delta_1, \delta_2, ..., \delta_K\}$. We denote the first K_w receivers as the *weak receivers*, and the next $K_s = K - K_w$ receivers as the *strong receivers*. To compensate for their poorer channel quality, each weak receiver is equipped with a cache memory of size nM bits, as depicted in Fig. 1. The special case in which all the receivers in the same set have the same erasure probability; that is, all the weak receivers have erasure probability δ_w, and all the strong receivers have erasure probability δ_s, with $\delta_s < \delta_w$, is called the *homogeneous scenario*. The set of erasure probabilities for the homogeneous scenario is represented by $\boldsymbol{\delta}_{ws}$.

Content delivery is performed in two phases [6]. It starts with the placement phase which takes place during the off-peak traffic period, and the caches of the weak receivers are filled without the knowledge of their future demands. Thus, only the weak receivers take part in the placement phase, and the contents of the cache of receiver k, for $k \in [K_w]$, at the end of this phase is denoted by Z_k. Note that, Z_k does not depend on any specific demand combination, and it is instead a function of the library $\mathbf{W}$. The caching function for receiver $k \in [K_w]$ is given by

$$\phi_k : \left[\left\lceil 2^{nR}\right\rceil\right]^N \to \left[\left\lfloor 2^{nM}\right\rfloor\right], \tag{2}$$

which maps the entire library to the cache content Z_k, i.e., $Z_k = \phi_k\left(\mathbf{W}\right)$. 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 X^n over the erasure BC. For a demand vector $\mathbf{d} \triangleq (d_1, ..., d_K)$, a coded delivery function

$$\psi : \left[\left\lceil 2^{nR}\right\rceil\right]^N \times [N]^K \to \mathcal{X}^n \qquad (3)$$

generates a common message X^n as a function of the entire library and the receiver demands, i.e., $X^n = \psi(\mathbf{W}, \mathbf{d})$, to be delivered over the erasure BC. Each receiver $k \in [K]$ observes Y_k^n according to (1). Each weak receiver $k \in [K_w]$ tries to decode W_{d_k} from its channel output Y_k^n along with the content available locally in its cache and the demand vector $\mathbf{d}$, utilizing the decoding function

$$\mu_k : \mathcal{Y}^n \times \left[\left\lfloor 2^{nM}\right\rfloor\right] \times [N]^K \to \left[\left\lceil 2^{nR}\right\rceil\right], \qquad (4)$$

i.e., the reconstructed file by each weak receiver $k \in [K_w]$ is

$$\hat{W}_{d_k} = \mu_k(Y_k^n, Z_k, \mathbf{d}). \qquad (5)$$

On the other hand, to serve the users with demand vector $\mathbf{d}$, each strong receiver $k \in [K_w + 1 : K]$ reconstructs its demand W_{d_k} solely from its channel output Y_k^n through the decoding function

$$\mu_k : \mathcal{Y}^n \times [N]^K \to \left[\left\lceil 2^{nR}\right\rceil\right], \qquad (6)$$

which generates the reconstructed file

$$\hat{W}_{d_k} = \mu_k(Y_k^n, \mathbf{d}). \qquad (7)$$

Definition 1. *An error occurs if $\hat{W}_{d_k} \neq W_{d_k}$ for any $k \in [K]$, and the probability of error is given by*

$$P_e \triangleq \max_{\mathbf{d} \in [N]^K} \Pr\left\{ \bigcup_{k=1}^K \left\{ \hat{W}_{d_k} \neq W_{d_k} \right\} \right\}. \qquad (8)$$

Definition 2. *A memory-rate pair (M, R) is said to be achievable, if for every $\varepsilon > 0$, there exists a large enough n, and corresponding caching function (2), coded delivery function (3), and decoding functions (4) and (6) at weak and strong receivers, respectively, such that $P_e < \varepsilon$.*

Definition 3. *For a given cache size M at the weak receivers, the capacity of the network is defined as*

$$C(M) \triangleq \sup \{R : (M, R) \text{ is achievable}\}. \qquad (9)$$

We would like to note that the capacity of the above caching network remains an open problem even when the delivery channel is an error-free shared bit pipe except for the network with an uncoded cache placement phase [11]. Here, our goal is to identify achievable memory-rate pairs that improve upon the state-of-the-art.

Remark 1. *The system model considered in this paper is inspired by those studied in [30]– [32]. We would like to remark that it is reasonable to assume that cache memories are placed at receivers with relatively weaker coverage. Indeed, it is shown in [31] that placing cache memories at the strong receivers, which already have good coverage, results in a lower capacity. This is mainly due to the definition of the capacity in this framework. Note that, the capacity here characterizes the highest rate of the messages to be delivered to all the receivers in the network. Since a symmetry is imposed across the receivers regarding their requests and file delivery, the system performance is determined by the worst receivers. Therefore, the goal of the cache placement should be to improve the performance of the weak receivers to increase the network capacity. Accordingly, equipping weak receivers with cache memories, and exploiting the coding scheme proposed in this paper also benefits the strong receivers.*

Following well-known results from multi-user information theory are included here for completeness as they will be instrumental in deriving our results later in the paper.

Proposition 1. *[36] The capacity region of a packet erasure BC with K receivers, where file W_i with rate R_i is targeted for receiver i with erasure probability δ_i, for $i = 1, ..., K$, is the closure of the set of non-negative rate tuples $(R_1, ..., R_K)$ that satisfy*

$$\sum_{i=1}^{K} \frac{R_i}{(1 - \delta_i) F} \leq 1, \tag{10}$$

where F denotes the length of the binary channel input.

Next, we consider the packet erasure BC with side information, and provide an achievable rate pair based on the joint encoding scheme of [37]. Here we briefly overview the coding scheme and the proof of achievability, and refer the reader to [37] for details. Consider two receivers with erasure probabilities $\delta_1 \geq \delta_2$. Let W_1 and W_2, distributed uniformly over $\left[2^{nR_1}\right]$ and $\left[2^{nR_2}\right]$, denote the messages targeted for receivers 1 and 2, respectively. We assume that message W_2 is available as side information at receiver 1, the weak receiver. We present a coding scheme

and the corresponding achievable rate region based on the joint encoding scheme of [37]. For a fixed distribution $P(X)$, we generate $2^{n(R_1+R_2)}$ codewords of length n, $x^n(w_1, w_2)$, $w_1 \in [2^{nR_1}]$, $w_2 \in [2^{nR_2}]$, where each entry of each codeword is generated independently according to $P(X)$. The codebook is revealed to the transmitter and the receivers. To transmit particular messages $W_1 = w_1$ and $W_2 = w_2$, the codeword $x^n(w_1, w_2)$ is transmitted over the BC. In the proposed coding scheme, the good receiver, i.e., receiver 2, decodes both messages; and therefore, it tries to find a unique pair $(\hat{w}_1, \hat{w}_2) \in [2^{nR_1}] \times [2^{nR_2}]$, such that $(X^n(\hat{w}_1, \hat{w}_2), y_2^n)$ belongs to the jointly typical set defined in [38]. The probability of decoding error tends to 0 as n goes to infinity, if

$$R_1 + R_2 \leq I(X; Y_2). \tag{11}$$

The first receiver already knows W_2 as side information; therefore, it only needs to decode W_1; thus, it looks for a unique index $\hat{w}_1 \in [2^{nR_1}]$ such that $(X^n(\hat{w}_1, W_2), y_1^n)$ belongs to the typical set [38]. The probability of error tends to 0 as n goes to infinity, if

$$R_1 \leq I(X; Y_1). \tag{12}$$

For the packet erasure BC, both mutual information terms are maximized with a uniform input, and the following conditions are obtained:

$$R_1 \leq (1 - \delta_1) F, \tag{13}$$

$$R_1 + R_2 \leq (1 - \delta_2) F, \tag{14}$$

We can easily generalize this coding scheme to multiple receivers and obtain the achievable rate region stated in the following proposition (also provided in [31]).

Proposition 2. *Consider a packet erasure BC with two disjoint sets of receivers $\mathcal{S}_1$ and $\mathcal{S}_2$, where the channels of the receivers in set $\mathcal{S}_i$ have erasure probability δ_i, for $i = 1, 2$. A common message W_i at rate R_i is to be transmitted to the receivers in set $\mathcal{S}_i$, for $i = 1, 2$, while message W_2 is known to the receivers in set $\mathcal{S}_1$ as side information. With the joint encoding scheme outlined above, rate pairs (R_1, R_2) satisfying the following conditions can be achieved*

$$R_1 \leq (1 - \delta_1) F, \tag{15}$$

$$R_1 + R_2 \leq (1 - \delta_2) F, \tag{16}$$

which is equivalent to

$$\max\left\{\frac{R_1}{(1-\delta_1)\,F}, \frac{R_1+R_2}{(1-\delta_2)\,F}\right\} \leq 1. \tag{17}$$

For notational convenience, in the rest of the paper we use

$$\mathrm{JE}\left((W_1)_{\mathcal{S}_1}, (W_2)_{\mathcal{S}_2}\right) \tag{18}$$

to represent the transmission of message W_1 to the receivers in set $\mathcal{S}_1$, and message W_2 to the receivers in set $\mathcal{S}_2$ using the outlined joint encoding scheme, where $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$, and W_2 is available at all the receivers in $\mathcal{S}_1$ as side information.

III. ACHIEVABLE RATE-MEMORY PAIRS

A coding scheme as well as an information theoretic upper bound on the capacity of the above caching and delivery network are proposed in [31] for the homogeneous scenario. Here, we present a new coding scheme, called the *successive cache-channel coding (SCC)* scheme, for delivery over any packet erasure BC, which is shown to improve upon the one in [31] in the homogeneous scenario. We present the (M, R) pairs achieved by this scheme in Theorem 1 below. The details of the scheme are presented in Section IV.

Theorem 1. *Consider cache-aided delivery of N files over a packet erasure BC with K_w weak and K_s strong receivers, where each weak receiver is equipped with a cache of capacity M. Memory-rate pairs $\left(M_{(p,q)}, R_{(p,q)}\right)$ are achievable for any $p \in [0 : K_w]$ and $q \in [p : K_w]$, where*

$$R_{(p,q)} \triangleq \frac{F \sum\limits_{i=p}^{q} \gamma\left(p, \boldsymbol{\delta}, i\right)}{\sum\limits_{i=p}^{q}\left(\frac{\gamma(p,\boldsymbol{\delta},i)}{\binom{K_w}{i}} \sum\limits_{j=1}^{K_w-i} \frac{\binom{K_w-j}{i}}{1-\delta_j}\right) + \sum\limits_{j=K_w+1}^{K} \frac{1}{1-\delta_j}}, \tag{19a}$$

$$M_{(p,q)} \triangleq \frac{N \sum\limits_{i=p}^{q} i\gamma\left(p, \boldsymbol{\delta}, i\right)}{K_w \sum\limits_{i=p}^{q} \gamma\left(p, \boldsymbol{\delta}, i\right)} R_{(p,q)}, \tag{19b}$$

with $\gamma(p, \boldsymbol{\delta}, i)$ defined as follows:

$$\gamma(p, \boldsymbol{\delta}, i) \triangleq \frac{\binom{K_w}{i}}{\binom{K_w}{p} K_s^{i-p}} \prod_{j=p}^{i-1}\left(\frac{K_s}{\left(1 - \delta_{K_w-j}\right) \sum\limits_{l=K_w+1}^{K} \frac{1}{1-\delta_l}} - 1\right), \quad \text{for } i = p, ..., q. \tag{19c}$$

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

Corollary 1. *For the homogeneous scenario, the achievable memory-rate pairs $(M_{(p,q)}, R_{(p,q)})$, for any $p \in [0 : K_w]$ and $q \in [p : K_w]$, are simplified as follows:*

$$R_{(p,q)} = \frac{F \sum\limits_{i=p}^{q} \gamma(p, \boldsymbol{\delta}_{ws}, i)}{\frac{1}{1-\delta_w} \sum\limits_{i=p}^{q} \left(\frac{K_w - i}{i+1} \gamma(p, \boldsymbol{\delta}_{ws}, i)\right) + \frac{K_s}{1-\delta_s}}, \tag{20a}$$

$$M_{(p,q)} = \frac{N \sum\limits_{i=p}^{q} i\gamma(p, \boldsymbol{\delta}_{ws}, i)}{K_w \sum\limits_{i=p}^{q} \gamma(p, \boldsymbol{\delta}_{ws}, i)} R_{(p,q)}, \tag{20b}$$

where

$$\gamma(p, \boldsymbol{\delta}_{ws}, i) = \frac{\binom{K_w}{i}}{\binom{K_w}{p} K_s^{\,i-p}} \left(\frac{1-\delta_s}{1-\delta_w} - 1\right)^{i-p}, \quad \text{for } i = p, ..., q. \tag{20c}$$

Remark 2. *As we will explain in Section IV, to achieve the rate-memory pair $(M_{(p,q)}, R_{(p,q)})$ with the proposed SCC scheme, for $p \in [0 : K_w]$ and $q \in [p : K_w]$, each file is divided into $p - q + 1$ non-overlapping subfiles. Here we remark that $\gamma(p, \boldsymbol{\delta}, i) / \sum\limits_{j=p}^{q} \gamma(p, \boldsymbol{\delta}, j)$, for $i \in [p : q]$, indicates the fraction of the total rate R allocated to the $(i-p+1)$-th of these subfiles. Thus, a proportioned rate allocation is performed according to the coefficients $\gamma(p, \boldsymbol{\delta}, p), \gamma(p, \boldsymbol{\delta}, p+1), ..., \gamma(p, \boldsymbol{\delta}, q)$.*

Next, we compare the achievable rate of the SCC scheme for the homogeneous scenario with the scheme of [31], which we will refer to as the STW scheme. In Fig. 2, the achievable memory-rate trade-off of the SCC scheme is compared with the STW scheme when $K_w = 2$, $K_s = 2$, $N = 20$, $F = 10$, $\delta_s = 0.2$, and $\delta_w = 0.8$. The upper bound on the capacity of the cache-aided packet erasure BC derived in [31, Theorem 7] is also included. The SCC scheme outperforms the STW scheme due to the improved achievable memory-rate pair $(M_{(0,2)}, R_{(0,2)})$, which is not achievable by the STW scheme. This improvement can be extended to a wider range of cache sizes through memory-sharing, also reducing the gap to the upper bound.

In Fig. 3, we plot the achievable rates for both schemes in the homogeneous 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 [31, Theorem 7] is also included. Observe that, for relatively small cache
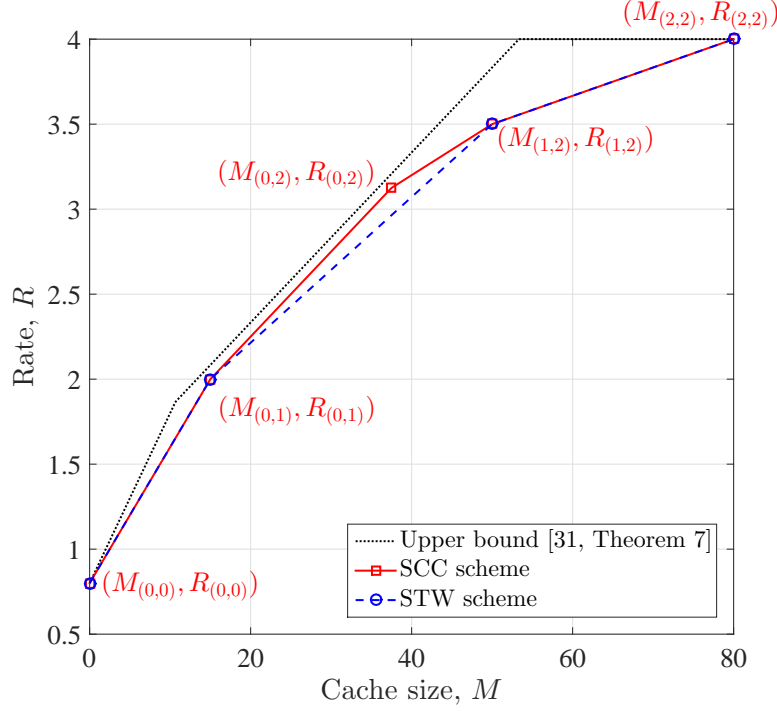
Fig. 2. Lower and upper bounds on the capacity for the homogeneous scenario with $K_w = 2$, $K_s = 2$, $N = 20$, $F = 10$, $\delta_s = 0.2$, and $\delta_w = 0.8$.

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 than STW scheme for all other intermediate cache sizes, and reduces the gap to the upper bound. For a cache capacity of $M = 30$, the SCC scheme provides approximately 15% increase in the achievable rate compared to the STW scheme.

In Fig. 4, the achievable rates of the SCC and STW schemes in the homogeneous scenario are compared for different values of δ_w. System parameters considered in this comparison are $K_w = 20$, $K_s = 10$, $N = 100$, $F = 50$, $\delta_s = 0.2$, and $\delta_w = 0.7, 0.8, 0.9$. Observe that, unlike the STW scheme, the performance of the SCC scheme does not deteriorate notably for intermediate and relatively high cache capacities when δ_w increases, i.e., having worse channel qualities for the weak receivers. This is because the SCC scheme successfully exploits the available cache capacities, and there is little to lose from increasing δ_w when M is sufficiently large. Moreover, the superiority of the SCC scheme over the STW scheme is more pronounced for higher values of
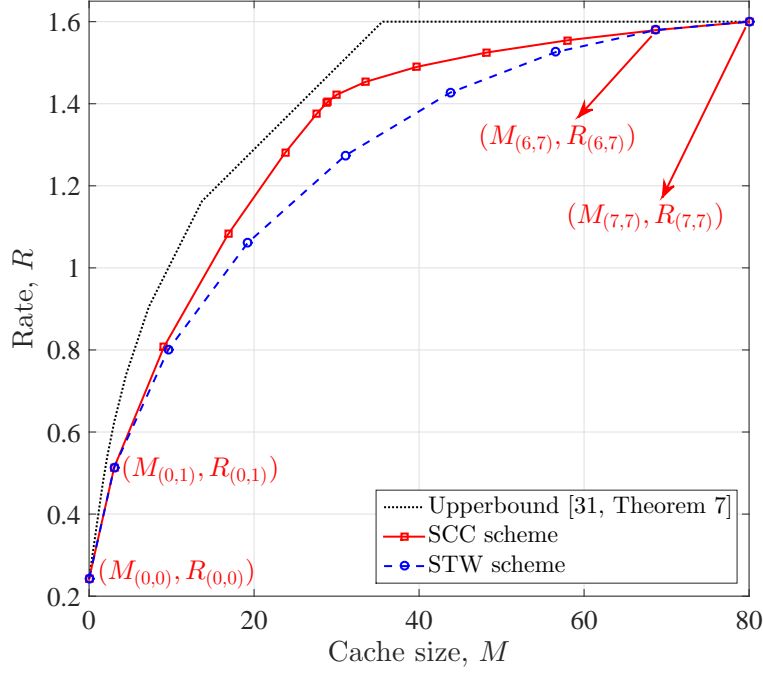
Fig. 3. Lower and upper bounds on the capacity for the homogeneous scenario with $K_w = 7$, $K_s = 10$, $N = 50$, $F = 20$, $\delta_s = 0.2$, and $\delta_w = 0.9$.

δ_w, in which case, exploiting the cache memories available at the weak receivers more effectively through SCC becomes more important.

For the heterogeneous scenario, the capacity of the network under consideration is upper bounded by [31]

$$\min_{\mathcal{S} \subset [K]} \left\{ F \left(\sum_{k \in \mathcal{S}} \frac{1}{1 - \delta_k} \right)^{-1} + \frac{M}{N} |\mathcal{S} \cap [K_w]| \right\}. \tag{21}$$

In Fig. 5, the effect of K_w is considered for the heterogeneous scenario with the system parameters $K = 15$, $N = 100$, $F = 10$, $\delta_k = 0.9 - 0.01k$, for $k = 1, ..., 5$, and $\delta_l = 0.2 - 0.01l$, for $l = 6, ..., 15$. In this figure, the achievable rates are plotted with respect to the total cache capacity of $K_w M$ for four different values for the number of weak receivers in the system, $K_w = 4, 5, 10, 15$. Note that the erasure probabilities are set such that the first 5 receivers have significantly worse channels than the remaining 10 receivers. Note also that the parameter K_w determines which receivers are provided with cache memories. As it can be seen, the setting with $K_w = 5$ achieves significantly higher rates over a wide range of total cache capacities compared to the other settings under consideration. If receiver 5, which has a relatively bad
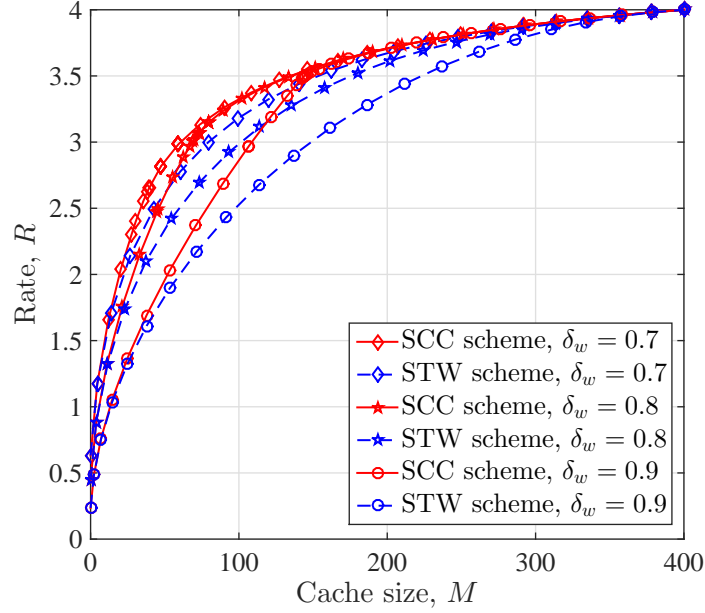
Fig. 4. Lower and upper bounds on the capacity for the homogeneous scenario with $K_w = 20$, $K_s = 10$, $N = 100$, $F = 50$, $\delta_s = 0.2$, and different values for δ_w given by $\delta_w = 0.7, 0.8, 0.9$.

channel quality, is not provided with any cache memory, and only the first 4 receivers are equipped with cache memories, i.e., $K_w = 4$, the performance degrades significantly except for very small values of total cache size. This is because the first five receivers have much worse channel qualities, and the performance depends critically on the caches provided to all these five weak receivers. On the other hand, equipping receivers with relatively good channel qualities with cache memories deteriorates the performance of the system in terms of the achievable rate. Note that this is because the total available cache capacity is allocated across a larger number of receivers. This result confirms that it is more beneficial to allocate cache memories to the receivers with relatively worse channel qualities. The upper bound on the achievable rate for the setting with $K_w = 5$ and $K_s = 10$ is also included in this figure. We observe that the gap between the upper bound and the achievable rate for the same setting is relatively small for a wide range of cache sizes.

IV. THE SUCCESSIVE CACHE-CHANNEL CODING (SCC) SCHEME

Before presenting the SCC scheme for the general heterogeneous scenario, in which we allow the weak and strong receivers to have distinct erasure probabilities, the main ideas behind this scheme are illustrated on an example in the simplified homogeneous scenario.
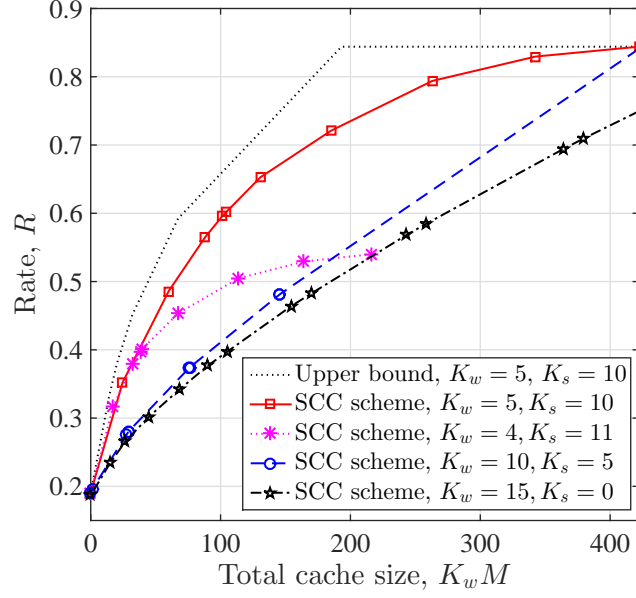
Fig. 5. Lower and upper bounds on the capacity for the heterogeneous scenario with $K = 15$, $N = 100$, $F = 10$, $\delta_k = 0.9 - 0.01k$, for $k = 1, ..., 5$, and $\delta_l = 0.2 - 0.01l$, for $l = 6, ..., 15$ with variable K_w and K_s.

For notational convenience, the i-element subsets of set $[K_w]$ are enumerated by $\mathcal{S}_1^{(i)}, \mathcal{S}_2^{(i)}, ...,$ $\mathcal{S}_{\binom{K_w}{i}}^{(i)}$, i.e.,

$$\mathcal{S}_j^{(i)} \subset [K_w] \text{ and } \left| \mathcal{S}_j^{(i)} \right| = i, \quad \text{for } i \in [0 : K_w], \text{ and } j = 1, ..., \binom{K_w}{i}. \tag{22}$$

Example 1. Consider the cache-aided packet erasure homogeneous BC depicted in Fig. 1 with $K_w = 3$ weak and $K_s = 2$ strong receivers. In the following, we investigate the achievable memory-rate pair $\left(M_{(0,2)}, R_{(0,2)} \right)$, which corresponds to the memory-rate pair in (20) for $p = 0$ and $q = 2$. Each file W_f, $f \in [N]$, is divided into three subfiles $W_f^{(0)}$, $W_f^{(1)}$ and $W_f^{(2)}$, where subfile $W_f^{(i)}$ has a rate of $R^{(i)}$, for $i = 0, 1, 2$, given by

$$R^{(i)} \triangleq \frac{\gamma\left(0, \boldsymbol{\delta}_{ws}, i\right)}{\sum\limits_{j=0}^{2} \gamma\left(0, \boldsymbol{\delta}_{ws}, j\right)} R, \tag{23}$$

where $\gamma\left(0, \boldsymbol{\delta}_{ws}, i\right)$ is as defined in (20c). We have $\sum\limits_{i=0}^{2} R^{(i)} = R$.

Placement phase: In the placement phase, subfiles $W_1^{(i)}, ..., W_N^{(i)}$ are placed in the caches of $K_w = 3$ weak receivers using the procedure in [6, Algorithm 1], specified for a cache capacity of iN/K_w, for $i = 0, 1, 2$. In this cache placement procedure, each subfile $W_f^{(i)}$ is first divided

into $\binom{3}{i}$ non-overlapping pieces, each at a rate of $R^{(i)}/\binom{3}{i}$.

$$W_f^{(i)} = \left(W_{f,\mathcal{S}_1^{(i)}}^{(i)}, W_{f,\mathcal{S}_2^{(i)}}^{(i)}, ..., W_{f,\mathcal{S}_{\binom{3}{i}}^{(i)}}^{(i)}\right), \forall f \in [N], \forall i \in [0:2], \tag{24}$$

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

$$W_f^{(0)} = \left(W_{f,\emptyset}^{(0)}\right), \tag{25a}$$

$$W_f^{(1)} = \left(W_{f,\{1\}}^{(1)}, W_{f,\{2\}}^{(1)}, W_{f,\{3\}}^{(1)}\right), \tag{25b}$$

$$W_f^{(2)} = \left(W_{f,\{1,2\}}^{(2)}, W_{f,\{1,3\}}^{(2)}, W_{f,\{2,3\}}^{(2)}\right). \tag{25c}$$

The piece $W_{f,\mathcal{S}_l^{(i)}}^{(i)}$ is placed in the cache of each receiver $k \in \mathcal{S}_l^{(i)}$, for $l = 1, ..., \binom{3}{i}$. Therefore, the cache contents of the weak receivers after the placement phase are as follows:

$$Z_1 = \bigcup_{f \in [N]} \left(W_{f,\{1\}}^{(1)}, W_{f,\{1,2\}}^{(2)}, W_{f,\{1,3\}}^{(2)}\right), \tag{26a}$$

$$Z_2 = \bigcup_{f \in [N]} \left(W_{f,\{2\}}^{(1)}, W_{f,\{1,2\}}^{(2)}, W_{f,\{2,3\}}^{(2)}\right), \tag{26b}$$

$$Z_3 = \bigcup_{f \in [N]} \left(W_{f,\{3\}}^{(1)}, W_{f,\{1,3\}}^{(2)}, W_{f,\{2,3\}}^{(2)}\right), \tag{26c}$$

where the required cache capacity for each weak receiver is:

$$M = \left(\frac{R^{(1)}}{3} + \frac{2R^{(2)}}{3}\right) N = \frac{\gamma(0, \boldsymbol{\delta}_{ws}, 1) + 2\gamma(0, \boldsymbol{\delta}_{ws}, 2)}{3 \sum_{j=0}^{2} \gamma(0, \boldsymbol{\delta}_{ws}, j)} NR. \tag{27}$$

Delivery phase: The server tries to satisfy all the demands in the delivery phase by sending four distinct messages in an orthogonal fashion, i.e., by time division multiplexing, where the codewords corresponding to the i-th message, $i = 1, ..., 4$, are of length $\beta_i n$ channel uses, such that $\sum_{i=1}^{4} \beta_i = 1$. The contents delivered with each message are illustrated in Table I.

The first message is targeted only for the weak receivers, and its goal is to deliver the missing subfiles of file $W_{d_k}^{(2)}$ to receiver k, $k = 1, 2, 3$, that is, having received this message, each weak receiver should be able to decode the third subfile of its desired file. Exploiting the delivery phase of [6, Algorithm 1] for cache capacity $2N/K_w$, the coded content with message 1 in Table I is delivered to the weak receivers $\{1, 2, 3\}$. Having received message 1 given in Table I, receiver k can recover its missing piece $W_{d_k, [3]\backslash\{k\}}^{(2)}$ of subfile $W_{d_k}^{(2)}$ using its cache contents Z_k.

TABLE I
CONTENTS SENT WITH MESSAGES 1 TO 4 IN THE DELIVERY PHASE OF EXAMPLE 1.

Message 1		$\left(W_{d_1,\{2,3\}}^{(2)} \oplus W_{d_2,\{1,3\}}^{(2)} \oplus W_{d_3,\{1,2\}}^{(2)} \right)$ to receivers $1,2,3$
Message 2	Sub-message 1	JE $\left(\left(W_{d_1,\{2\}}^{(1)} \oplus W_{d_2,\{1\}}^{(1)} \right)_{\{1,2\}}, \left(W_{d_4,\{1,2\}}^{(2)}, W_{d_5,\{1,2\}}^{(2)} \right)_{\{4,5\}} \right)$
	Sub-message 2	JE $\left(\left(W_{d_1,\{3\}}^{(1)} \oplus W_{d_3,\{1\}}^{(1)} \right)_{\{1,3\}}, \left(W_{d_4,\{1,3\}}^{(2)}, W_{d_5,\{1,3\}}^{(2)} \right)_{\{4,5\}} \right)$
	Sub-message 3	JE $\left(\left(W_{d_2,\{3\}}^{(1)} \oplus W_{d_3,\{2\}}^{(1)} \right)_{\{2,3\}}, \left(W_{d_4,\{2,3\}}^{(2)}, W_{d_5,\{2,3\}}^{(2)} \right)_{\{4,5\}} \right)$
Message 3	Sub-message 1	JE $\left(\left(W_{d_1,\emptyset}^{(0)} \right)_{\{1\}}, \left(W_{d_4,\{1\}}^{(1)}, W_{d_5,\{1\}}^{(1)} \right)_{\{4,5\}} \right)$
	Sub-message 2	JE $\left(\left(W_{d_2,\emptyset}^{(0)} \right)_{\{2\}}, \left(W_{d_4,\{2\}}^{(1)}, W_{d_5,\{2\}}^{(1)} \right)_{\{4,5\}} \right)$
	Sub-message 3	JE $\left(\left(W_{d_3,\emptyset}^{(0)} \right)_{\{3\}}, \left(W_{d_4,\{3\}}^{(1)}, W_{d_5,\{3\}}^{(1)} \right)_{\{4,5\}} \right)$
Message 4		$\left(W_{d_4,\emptyset}^{(0)} \right)$ to receiver 4, and $\left(W_{d_5,\emptyset}^{(0)} \right)$ to receiver 5

Thus, together with its cache content, receiver k can recover subfile $W_{d_k}^{(2)}$, for $k = 1, 2, 3$.

Through the second message of the delivery phase, the server simultaneously delivers subfile $W_{d_l}^{(2)}$ to strong receiver l, $l = 4, 5$, and the missing bits of subfile $W_{d_k}^{(1)}$ to weak receiver k, $k = 1, 2, 3$. The content targeted to the weak receivers is delivered by using the delivery phase of [6, Algorithm 1] for the cache capacity of N/K_w; that is, the contents

$$\left\{ W_{d_1,\{2\}}^{(1)} \oplus W_{d_2,\{1\}}^{(1)}, W_{d_1,\{3\}}^{(1)} \oplus W_{d_3,\{1\}}^{(1)}, W_{d_2,\{3\}}^{(1)} \oplus W_{d_3,\{2\}}^{(1)} \right\} \tag{28}$$

are transmitted to the weak receivers. Therefore, the goal is to deliver $W_{d_l}^{(2)}$ to strong receiver l, $l = 4, 5$, while delivering the contents in (28) to the weak receivers in parallel. The transmission is performed by sending three sub-messages, transmitted over orthogonal time periods. With the first sub-message of message 2 given in Table I, receivers 1 and 2 receive $W_{d_1,\{2\}}^{(1)} \oplus W_{d_2,\{1\}}^{(1)}$ since they both have $W_{d_4,\{1,2\}}^{(2)}$ and $W_{d_5,\{1,2\}}^{(2)}$ in their caches as side information. Accordingly, receiver 1 and receiver 2 can recover $W_{d_1,\{2\}}^{(1)}$ and $W_{d_2,\{1\}}^{(1)}$, respectively. On the other hand, with the joint encoding scheme, $W_{d_4,\{1,2\}}^{(2)}$ and $W_{d_5,\{1,2\}}^{(2)}$ are directly delivered to receiver 4 and receiver 5, respectively. With the second sub-message of message 2 in Table I, $W_{d_4,\{1,3\}}^{(2)}$ and $W_{d_5,\{1,3\}}^{(2)}$, which are available in the caches of receivers 1 and 3 as side information, are delivered to receivers 4 and 5, while $W_{d_1,\{3\}}^{(1)} \oplus W_{d_3,\{1\}}^{(1)}$ is delivered to receivers 1 and 3. By receiving sub-message

2, receiver 1 and receiver 3 can obtain $W_{d_1,\{3\}}^{(1)}$ and $W_{d_3,\{1\}}^{(1)}$, respectively. Finally, sub-message 3 of message 2 aims to deliver $W_{d_4,\{2,3\}}^{(2)}$ and $W_{d_5,\{2,3\}}^{(2)}$, which are in the cache of receivers 2 and 3, to receivers 4 and 5, respectively, and $W_{d_2,\{3\}}^{(1)} \oplus W_{d_3,\{2\}}^{(1)}$ to receivers 2 and 3 by the joint encoding scheme. Having received coded content $W_{d_2,\{3\}}^{(1)} \oplus W_{d_3,\{2\}}^{(1)}$, receiver 2 and receiver 3 can recover $W_{d_2,\{3\}}^{(1)}$ and $W_{d_3,\{2\}}^{(1)}$, respectively. Thus, having received message 2, each weak receiver k, $k = 1, 2, 3$, can recover all the missing bits of subfile $W_{d_k}^{(1)}$ of its request, while each strong receiver l, $l = 4, 5$, can obtain subfile $W_{d_l}^{(2)}$ of its request.

The third message of the delivery phase is designed to deliver $W_{d_l}^{(1)}$ to strong receiver l, $l = 4, 5$, and $W_{d_k}^{(0)}$ is delivered to weak receiver k, $k = 1, 2, 3$. Third message is also divided into three sub-messages, transmitted over orthogonal time periods. With sub-message k, given in Table I, $W_{d_4,\{k\}}^{(1)}$ and $W_{d_5,\{k\}}^{(1)}$, both of which are available locally at receiver k as side information, $k = 1, 2, 3$, are delivered to receivers 4 and 5, respectively, while $W_{d_k,\emptyset}^{(0)}$ is delivered to receiver k. Therefore, with the third message in Table I, each weak receiver k, $k = 1, 2, 3$, can obtain $W_{d_k,\emptyset}^{(0)}$, while each strong receiver l, $l = 4, 5$, can recover $W_{d_l}^{(1)}$. Thus, after receiving message 3 in Table I, the demands of the weak receivers are fully satisfied.

The last and fourth message of the delivery phase is generated only for the strong receivers with the goal of delivering them the missing bits of their demands, in particular, subfile $W_{d_l,\emptyset}^{(0)}$ is delivered to each strong receiver l, $l = 4, 5$.

Observe that message 1 in Table I has a rate of $R^{(2)}/3$. The capacity region of the standard packet erasure BC presented in Proposition 1 suggests that all the weak receivers can decode message 1, for n large enough, if

$$\frac{R^{(2)}}{3\left(1 - \delta_w\right)F} \leq \beta_1. \tag{29}$$

From Table I, with each sub-message of the second message, messages of rate $2R^{(2)}/3$, available at the weak receivers as side information, are delivered to the strong receivers; while, simultaneously, a common message at rate $R^{(1)}/3$ is transmitted to the weak receivers. Overall, $\left(W_{d_4}^{(2)}, W_{d_5}^{(2)}\right)$ and the contents in (28) with a total rate of $2R^{(2)}$ and $R^{(1)}$ are delivered to the strong and weak receivers, respectively, through three different sub-messages by using the joint encoding scheme of [37] that exploits the side information at the weak receivers. Using the achievable rate region of the joint encoding scheme for the packet erasure channels stated in Proposition 2, $\left(W_{d_4}^{(2)}, W_{d_5}^{(2)}\right)$ and the contents in (28) can be simultaneously decoded by the

strong and weak receivers, respectively, for n large enough, if

$$\max\left\{\frac{R^{(1)}}{(1-\delta_w)\,F}, \frac{R^{(1)}+2R^{(2)}}{(1-\delta_s)\,F}\right\} \leq \beta_2. \tag{30}$$

From the expressions for $R^{(1)}$ and $R^{(2)}$ in (23), it can be verified that the two terms in the maximization in (30) are equal for the setting under consideration. Thus, the condition in (30) can be simplified as

$$\frac{R^{(1)}}{(1-\delta_w)\,F} \leq \beta_2. \tag{31}$$

According to Table I, with each sub-message of message 3, a message at rate $R^{(0)}$ is targeted for the weak receivers, while message at rate $2R^{(1)}/3$, available locally at the weak receivers, is aimed for the strong receivers. Therefore, through joint encoding scheme over three periods, messages with a total rate of $3R^{(0)}$ are delivered to the weak receivers, while the strong receivers receive a total rate of $2R^{(1)}$. According to Proposition 2, all the weak and strong receivers can decode their messages, for n large enough, if

$$\max\left\{\frac{3R^{(0)}}{(1-\delta_w)\,F}, \frac{3R^{(0)}+2R^{(1)}}{(1-\delta_s)\,F}\right\} \leq \beta_3. \tag{32}$$

Again, from the expressions of $R^{(0)}$ and $R^{(1)}$ in (23), it can be verified that, when $K_w = 3$ and $K_s = 2$, (32) can be simplified as

$$\frac{3R^{(0)}}{(1-\delta_w)\,F} \leq \beta_3. \tag{33}$$

From the capacity region of the standard erasure BC in Proposition 1, each receiver l, $l = 4, 5$, can decode $W_{d_l,\emptyset}^{(0)}$, delivered with message 4, successfully for n sufficiently large, if

$$\frac{2R^{(0)}}{(1-\delta_s)\,F} \leq \beta_4. \tag{34}$$

Combining (29), (31), (33), (34), and the fact that $\sum_{i=1}^{4} \beta_i = 1$, we have the condition

$$\frac{R^{(2)}}{3\,(1-\delta_w)\,F} + \frac{R^{(1)}}{(1-\delta_w)\,F} + \frac{3R^{(0)}}{(1-\delta_w)\,F} + \frac{2R^{(0)}}{(1-\delta_s)\,F} \leq 1. \tag{35}$$

By replacing $R^{(i)}$ with the expressions from (23), for $i = 0, 1, 2$, and using the fact that

$\gamma\left(0,\boldsymbol{\delta}_{ws},0\right)=1$, (35) is reduced to

$$R\leq\frac{\sum\limits_{j=0}^{2}\gamma\left(0,\boldsymbol{\delta}_{ws},j\right)F}{\frac{1}{1-\delta_{w}}\left(3+\gamma\left(0,\boldsymbol{\delta}_{ws},1\right)+\frac{1}{3}\gamma\left(0,\boldsymbol{\delta}_{ws},2\right)\right)+\frac{2}{1-\delta_{s}}}.\tag{36}$$

Observe that, the term on the right hand side of inequality (36) is $R_{(0,2)}$, which is given by (20a). The cache size of each weak receiver exploited by our coding scheme is given by (27), and we have $M=M_{(0,2)}$, where $M_{(0,2)}$ is defined as in (20b). Thus, the memory-rate pair $\left(M_{(0,2)},R_{(0,2)}\right)$ given by (20) is achievable for the setting under consideration. □

Next, we present the SCC scheme for a general heterogeneous scenario, achieving the memory-rate pair $\left(M_{(p,q)},R_{(p,q)}\right)$ given by (19), for any $p\in[0:K_{w}]$ and $q\in[p:K_{w}]$. For a given (p,q) pair, where $p\in[0:K_{w}]$ and $q\in[p:K_{w}]$, each file W_{f}, $f\in[N]$, is divided into $(q-p+1)$ non-overlapping subfiles, represented by

$$W_{f}=\left(W_{f}^{(p)},...,W_{f}^{(q)}\right),\tag{37}$$

where subfile $W_{f}^{(i)}$, for $i\in[p:q]$, has a rate of

$$R^{(i)}\triangleq\frac{\gamma\left(p,\boldsymbol{\delta},i\right)}{\sum\limits_{j=p}^{q}\gamma\left(p,\boldsymbol{\delta},j\right)}R,\tag{38}$$

such that $\sum\limits_{i=p}^{q}R^{(i)}=R$. Note that, $\gamma\left(p,\boldsymbol{\delta},i\right)$, for $i\in[p:q]$, is given in (19c).

In the sequel, the placement and delivery phases of the SCC scheme are explained, and the achievability of the memory-rate pair in (19) is proven.

A. Placement Phase

In the placement phase, for each set of subfiles $W_{1}^{(i)},...,W_{N}^{(i)}$ a cache placement procedure, corresponding to the one proposed in [6, Algorithm 1] for a cache capacity of iN/K_{w}, is performed, $\forall i\in[p:q]$; that is, each subfile $W_{f}^{(i)}$ is partitioned into $\binom{K_{w}}{i}$ independent equal-rate pieces,

$$W_{f}^{(i)}=\left(W_{f,\mathcal{S}_{1}^{(i)}}^{(i)},W_{f,\mathcal{S}_{2}^{(i)}}^{(i)},...,W_{f,\mathcal{S}_{\binom{K_{w}}{i}}^{(i)}}^{(i)}\right),\forall f\in[N],\forall i\in[p:q].\tag{39}$$

The piece $W^{(i)}_{f,\mathcal{S}^{(i)}_l}$ of rate $R^{(i)}/\binom{K_w}{i}$ is cached by receivers $k \in \mathcal{S}^{(i)}_l$, for $l = 1, ..., \binom{K_w}{i}$. Thus, the content placed in the cache of each weak receiver $k \in [K_w]$ is given by

$$Z_k = \bigcup_{f \in [N]} \bigcup_{i \in [p:q]} \bigcup_{l \in \left[\binom{K_w}{i}\right]:k \in \mathcal{S}^{(i)}_l} W^{(i)}_{f,\mathcal{S}^{(i)}_l}. \tag{40}$$

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

$$M = N \sum_{i=p}^{q} \binom{K_w - 1}{i - 1} \frac{R^{(i)}}{\binom{K_w}{i}} = \frac{N}{K_w} \sum_{i=p}^{q} i R^{(i)} = \frac{N \sum\limits_{i=p}^{q} i\gamma(p,\boldsymbol{\delta},i)}{K_w \sum\limits_{i=p}^{q} \gamma(p,\boldsymbol{\delta},i)} R. \tag{41}$$

B. Delivery Phase

In the delivery phase, the goal is to satisfy all the demands for an arbitrary demand combination $(d_1, ..., d_K)$. The delivery phase consists of $(q - p + 2)$ different messages, transmitted over orthogonal time periods, where the codewords of the i-th message are of length $\beta_i n$ channel uses, for $i = 1, ..., q - p + 2$, such that $\sum\limits_{i=1}^{q-p+2} \beta_i = 1$.

The first message of the delivery phase is only targeted for the weak receivers, and the goal is to deliver the missing bits of subfile $W^{(q)}_{d_k}$ to receiver k, $\forall k \in [K_w]$. It is to be noted that, for $q = K_w$, based on the cache contents in (40), all the weak receivers have all the subfiles $W^{(q)}_f$, $\forall f \in [N]$; therefore, no message needs to be delivered. In the sequel, we consider $q < K_w$. The first message of the delivery phase is transmitted over $\binom{K_w}{q+1}$ orthogonal time slots, where in each slot, a sub-message is delivered to a group of $q+1$ weak receivers. Sub-message j is a codeword of length $\beta_{1,j} n$ channel uses, and is targeted to the receivers in set $\mathcal{S}^{(q+1)}_j$, for $j = 1, ..., \binom{K_w}{q+1}$, such that $\sum\limits_{j=1}^{\binom{K_w}{q+1}} \beta_{1,j} = \beta_1$. Following the procedure in [6, Algorithm 1], the content delivered by sub-message j is given by

$$V^{(q)}_j \triangleq \bigoplus_{k \in \mathcal{S}^{(q+1)}_j} W^{(q)}_{d_k,\mathcal{S}^{(q+1)}_j \setminus \{k\}}, \quad \text{for } j = 1, ..., \binom{K_w}{q+1}. \tag{42}$$

After receiving $V^{(q)}_j$, each receiver $k \in \mathcal{S}^{(q+1)}_j$ can obtain $W^{(q)}_{d_k,\mathcal{S}^{(q+1)}_j \setminus \{k\}}$, for $j = 1, ..., \binom{K_w}{q+1}$, i.e., the missing bits of subfile $W^{(q)}_{d_k}$ of its desired file, having access to Z_k given in (40). Thus, together with its cache content, receiver $k \in [K_w]$ can recover $W^{(q)}_{d_k}$.

The delivery technique performed to transmit messages $2, 3, ..., q - p + 1$ follows the same procedure. With the message $q - i + 1$ of length $\beta_{q-i+1}n$ channel uses, the server delivers the missing bits of subfile $W_{d_k}^{(i)}$ to each weak receiver k, $k \in [K_w]$, and $W_{d_l}^{(i+1)}$ to each strong receiver l, $l \in [K_w + 1 : K]$, for $i = q - 1, q - 2, ..., p$.[1] Message $(q - i + 1)$ is delivered through $\binom{K_w}{i+1}$ sub-messages, transmitted over orthogonal time periods, where sub-message j is of length $\beta_{q-i+1,j}n$ channel uses, such that $\sum_{j=1}^{\binom{K_w}{i+1}} \beta_{q-i+1,j} = \beta_{q-i+1}$. With the j-th sub-message, using the coded delivery procedure in [6, Algorithm 1], the coded content

$$V_j^{(i)} \triangleq \bigoplus_{k \in \mathcal{S}_j^{(i+1)}} W_{d_k, \mathcal{S}_j^{(i+1)} \setminus \{k\}}^{(i)}, \tag{43}$$

is delivered to the weak receivers in set $\mathcal{S}_j^{(i+1)}$, while

$$\left\{ W_{d_{K_w+1}, \mathcal{S}_j^{(i+1)}}^{(i+1)}, ..., W_{d_K, \mathcal{S}_j^{(i+1)}}^{(i+1)} \right\}, \tag{44}$$

is delivered to the strong receivers, for $i = q - 1, ..., p$, and $j = 1, ..., \binom{K_w}{i+1}$. Observe that, after receiving sub-message $V_j^{(i)}$, each receiver $k \in \mathcal{S}_j^{(i+1)}$ can obtain $W_{d_k, \mathcal{S}_j^{(i+1)} \setminus \{k\}}^{(i)}$, for $j = 1, ..., \binom{K_w}{i+1}$, i.e., the missing bits of subfile $W_{d_k}^{(i)}$ of its desired file, for $i = q - 1, ..., p$. Note that, the content in (44), which is targeted to the strong receivers, is known by each weak receiver in set $\mathcal{S}_j^{(i+1)}$. Therefore, the j-th sub-message of message $q - i + 1$ can be transmitted using joint encoding:

$$\text{JE}\left(\left(V_j^{(i)} \right)_{\mathcal{S}_j^{(i+1)}}, \left(W_{d_{K_w+1}, \mathcal{S}_j^{(i+1)}}^{(i+1)}, ..., W_{d_K, \mathcal{S}_j^{(i+1)}}^{(i+1)} \right)_{[K_w+1:K]} \right), \quad \text{for } i = q - 1, ..., p,$$

$$\text{and } j = 1, ..., \binom{K_w}{i+1}. \tag{45}$$

However, to increase the efficiency of the delivery phase, the j-th sub-message is delivered via K_s orthogonal time periods, such that in the m-th period a codeword of length $\beta_{q-i+1,j,m}n$ channel uses is transmitted, where $\sum_{m=1}^{K_s} \beta_{q-i+1,j,m} = \beta_{q-i+1,j}$. Coded content $V_j^{(i)}$, targeted for receivers in set $\mathcal{S}_j^{(i+1)}$, is divided into K_s non-overlapping equal-rate pieces

$$V_j^{(i)} = \left(V_{j,1}^{(i)}, ..., V_{j,K_s}^{(i)} \right), \tag{46}$$

---

[1]For example, with the second message, subfile $W_{d_l}^{(q)}$ is delivered to each strong receiver $l \in [K_w + 1 : K]$, and subfile $W_{d_k}^{(q-1)}$ to each weak receiver $k \in [K_w]$. With the third message, subfile $W_{d_l}^{(q-1)}$ is delivered to each strong receiver $l \in [K_w + 1 : K]$, and subfile $W_{d_k}^{(q-2)}$ to each weak receiver $k \in [K_w]$, and so on so forth.

TABLE II
CONTENTS SENT WITH MESSAGES 1 TO $q - p + 2$ IN THE DELIVERY PHASE OF THE SCC SCHEME FOR THE HETEROGENEOUS SCENARIO.

Message index	Sub-message index	Delivered content
1	1	$\left(V_1^{(q)}\right)$ to receivers in $\mathcal{S}_1^{(q+1)}$
	$\vdots$	$\vdots$
	$\binom{K_w}{q+1}$	$\left(V_{\binom{K_w}{q+1}}^{(q)}\right)$ to receivers in $\mathcal{S}_{\binom{K_w}{q+1}}^{(q+1)}$
$t = 2, ..., $ $q - p + 1$	1	$\text{JE}\left(\left(V_{1,m}^{(q-t+1)}\right)_{\mathcal{S}_1^{(q-t+2)}}, \left(W_{d_{K_w+m}, \mathcal{S}_1^{(q-t+2)}}^{(q-t+2)}\right)_{\{K_w+m\}}\right)$ in m-th time period, for $m = 1, ..., K_s$
	$\vdots$	$\vdots$
	$\binom{K_w}{q-t+2}$	$\text{JE}\left(\left(V_{\binom{K_w}{q-t+2},m}^{(q-t+1)}\right)_{\mathcal{S}_{\binom{K_w}{q-t+2}}^{(q-t+2)}}, \left(W_{d_{K_w+m}, \mathcal{S}_{\binom{K_w}{q-t+2}}^{(q-t+2)}}^{(q-t+2)}\right)_{\{K_w+m\}}\right)$ in m-th time period, for $m = 1, ..., K_s$
$q - p + 2$		$\left(W_{d_l}^{(p)}\right)$ to receiver $l \in [K_w + 1 : K]$

and the delivery over the m-th time period is performed by joint encoding:

$$\text{JE}\left(\left(V_{j,m}^{(i)}\right)_{\mathcal{S}_j^{(i+1)}}, \left(W_{d_{K_w+m}, \mathcal{S}_j^{(i+1)}}^{(i+1)}\right)_{\{K_w+m\}}\right), \quad \text{for } m = 1, ..., K_s. \tag{47}$$

We note that, after receiving messages 2 to $q - p + 1$, each weak receiver $k \in [K_w]$ can obtain subfiles $\left(W_{d_k}^{(q-1)}, W_{d_k}^{(q-2)}, ..., W_{d_k}^{(p)}\right)$, while each strong receiver $l \in [K_w + 1 : K]$ can decode subfiles $\left(W_{d_l}^{(q)}, W_{d_l}^{(q-1)}, ..., W_{d_l}^{(p+1)}\right)$; therefore, together with message 1, the demand of weak receivers are fully satisfied. However, strong receiver $l \in [K_w + 1 : K]$ only requires to receive subfile $W_{d_l}^{(p)}$.

The last message delivers subfile $W_{d_l}^{(p)}$ to the strong receiver $l \in [K_w + 1 : K]$ using the channel coding scheme for standard packet erasure BCs.

The contents delivered with each message in the delivery phase for the heterogeneous scenario are summarized in Table II.

Remark 3. *We remark here that, instead of multicasting XORed contents to groups of weak receivers, one can also use a higher dimensional joint encoding scheme, in which case each*

weak receiver can decode its missing part directly by using the parts available in its cache as side information. However, we stick to the coding scheme presented here as it makes the connection to the original delivery scheme of Maddah-Ali and Niesen in [6] more explicit.

C. Achievable Memory-Rate Pair Analysis (Proof of Theorem 1)

The rate of the coded content targeted to a group of weak receivers for each message of the delivery phase is allocated such that it can be decoded by the weakest receiver among the intended group of weak receivers.

With sub-message j of message 1 of length $\beta_{1,j} n$ channel uses, $V_j^{(q)}$, which is given in (42), a message of rate $R^{(q)}/\binom{K_w}{q}$ is transmitted to the weak receivers in $\mathcal{S}_j^{(q+1)}$, for $j = 1, ..., \binom{K_w}{q+1}$. The rate of $V_j^{(q)}$ is adjusted such that the weakest receiver in $\mathcal{S}_j^{(q+1)}$ can decode it successfully, i.e.,

$$\frac{R^{(q)}}{\binom{K_w}{q}} \leq \beta_{1,j} \left(1 - \max_{r \in \mathcal{S}_j^{(q+1)}} \{\delta_r\} \right) F, \quad \text{for } j = 1, ..., \binom{K_w}{q+1}, \tag{48}$$

which after summing over all the sets $\mathcal{S}_j^{(q+1)}$, for $j = 1, ..., \binom{K_w}{q+1}$, one can obtain

$$\frac{R^{(q)}}{\binom{K_w}{q}} \sum_{r=1}^{K_w - q} \frac{\binom{K_w - r}{q}}{1 - \delta_r} \leq \sum_{j=1}^{\binom{K_w}{q+1}} \beta_{1,j} F = \beta_1 F. \tag{49}$$

Note that with the codeword given in (47), $V_{j,m}^{(i)}$, targeted for the receivers in $\mathcal{S}_j^{(i+1)}$, is of rate $R^{(i)}/\left(K_s \binom{K_w}{i} \right)$, while $W_{d_{K_w+m}, \mathcal{S}_j^{(i+1)}}^{(i+1)}$, destined for receiver $K_w + m$, is of rate $R^{(i+1)}/\binom{K_w}{i+1}$, for $m = 1, ..., K_s$, $i = q - 1, ..., p$ and $j = 1, ..., \binom{K_w}{i+1}$. Proposition 2 suggests that the codeword in (47) can be decoded correctly by the intended receivers if

$$\max \left\{ \frac{R^{(i)}/\left(K_s \binom{K_w}{i} \right)}{\left(1 - \max_{r \in \mathcal{S}_j^{(i+1)}} \{\delta_r\} \right) F}, \frac{R^{(i)}/\left(K_s \binom{K_w}{i} \right) + R^{(i+1)}/\binom{K_w}{i+1}}{\left(1 - \delta_{K_w+m} \right) F} \right\} \leq \beta_{q-i+1,j,m}, \text{for } m = 1, ..., K_s, \tag{50}$$

where the rate of $V_{j,m}^{(i)}$ is limited by the weakest receiver in $\mathcal{S}_j^{(i+1)}$, for $i = q - 1, ..., p$, and

$j = 1, ..., \binom{K_w}{i+1}$. By summing up all the K_s inequalities in (50), we have

$$\max \left\{ \frac{R^{(i)}/\binom{K_w}{i}}{\left(1 - \max\limits_{r \in \mathcal{S}_j^{(i+1)}} \{\delta_r\}\right) F}, \left(\frac{R^{(i)}}{K_s \binom{K_w}{i}} + \frac{R^{(i+1)}}{\binom{K_w}{i+1}} \right) \sum_{m=1}^{K_s} \frac{1}{(1 - \delta_{K_w+m}) F} \right\} \leq \beta_{q-i+1,j},$$

$$\text{for } i = q - 1, ..., p, \text{ and } j = 1, ..., \binom{K_w}{i+1}. \quad (51)$$

By the choice of (38), and the fact that

$$\gamma(p, \boldsymbol{\delta}, i + 1) = \frac{\binom{K_w}{i+1}}{\binom{K_w}{i} K_s} \left(\frac{K_s}{(1 - \delta_{K_w-i}) \sum\limits_{l=K_w+1}^{K} \frac{1}{1-\delta_l}} - 1 \right) \gamma(p, \boldsymbol{\delta}, i), \quad (52)$$

which follows from the definition in (19c), the second term of the maximization in (51) is reduced to

$$\frac{R^{(i)}/\binom{K_w}{i}}{(1 - \delta_{K_w-i}) F}. \quad (53)$$

Thus, (51) is simplified as follows:

$$\max \left\{ \frac{R^{(i)}/\binom{K_w}{i}}{\left(1 - \max\limits_{r \in \mathcal{S}_j^{(i+1)}} \{\delta_r\}\right) F}, \frac{R^{(i)}/\binom{K_w}{i}}{(1 - \delta_{K_w-i}) F} \right\} \leq \beta_{q-i+1,j}, \quad \text{for } i = q - 1, ..., p, \ j = 1, ..., \binom{K_w}{i+1}.$$

$$(54)$$

Note that $\left| \mathcal{S}_j^{(i+1)} \right| = i + 1$; hence, for $i = q - 1, ..., p$,

$$\max_{r \in \mathcal{S}_j^{(i+1)}} \{\delta_r\} \geq \delta_{K_w-i}, \quad \forall j \in \left[\binom{K_w}{i+1} \right]. \quad (55)$$

From (55), (54) is reduced to

$$\frac{R^{(i)}/\binom{K_w}{i}}{\left(1 - \max\limits_{r \in \mathcal{S}_j^{(i+1)}} \{\delta_r\}\right) F} \leq \beta_{q-i+1,j}, \quad \text{for } i = q - 1, ..., p, \ j = 1, ..., \binom{K_w}{i+1}, \quad (56)$$

which holds for every $j \in \left[\binom{K_w}{i+1} \right]$, each corresponding to a different $(i + 1)$-element subset

$\mathcal{S}_j^{(i+1)}$. After summing up over all values of j, one can obtain

$$\frac{R^{(i)}}{\binom{K_w}{i}} \sum_{r=1}^{K_w-i} \frac{\binom{K_w-r}{i}}{1-\delta_r} \leq \sum_{j=1}^{\binom{K_w}{i+1}} \beta_{q-i+1,j} F = \beta_{q-i+1} F, \quad \text{for } i = q-1, ..., p. \tag{57}$$

According to Proposition 1, each receiver l, $l \in [K_w + 1 : K]$, can decode subfile $W_{d_k}^{(p)}$ of rate $R^{(p)}$, delivered by the last message, correctly, if

$$R^{(p)} \sum_{l=K_w+1}^{K} \frac{1}{1-\delta_l} \leq \beta_{q-p+2} F. \tag{58}$$

By combining inequalities (49), (57) and (58), we have

$$\sum_{i=p}^{q} \left(\frac{R^{(i)}}{\binom{K_w}{i}} \sum_{j=1}^{K_w-i} \frac{\binom{K_w-j}{i}}{1-\delta_j} \right) + R^{(p)} \sum_{j=K_w+1}^{K} \frac{1}{1-\delta_j} \leq \sum_{i=p-1}^{q} \beta_{q-i+1} F = F. \tag{59}$$

Finally, by replacing $R^{(i)}$, for $i = p, ..., q$, with the expression in (38), one can obtain

$$R \leq \frac{F \sum_{i=p}^{q} \gamma(p, \boldsymbol{\delta}, i)}{\sum_{i=p}^{q} \left(\frac{\gamma(p,\boldsymbol{\delta},i)}{\binom{K_w}{i}} \sum_{j=1}^{K_w-i} \frac{\binom{K_w-j}{i}}{1-\delta_j} \right) + \sum_{j=K_w+1}^{K} \frac{1}{1-\delta_j}}, \tag{60}$$

which, together with the cache capacity of each weak receiver, M, given in (41), proves the achievability of the memory-rate pairs $\left(M_{(p,q)}, R_{(p,q)}\right)$ in (19).

Remark 4. *We remark that, for the homogeneous scenario, the codeword in (47) can be received by the targeted receivers correctly if*

$$\max \left\{ \frac{R^{(i)} / \left(K_s \binom{K_w}{i}\right)}{(1-\delta_w) F}, \frac{R^{(i)} / \left(K_s \binom{K_w}{i}\right) + R^{(i+1)} / \binom{K_w}{i+1}}{(1-\delta_s) F} \right\} \leq \beta_{q-i+1,j,m},$$

$$\text{for } i = q-1, ..., p, \ j = 1, ..., \binom{K_w}{i+1}, \text{ and } m = 1, ..., K_s, \tag{61}$$

from which we obtain

$$\max \left\{ \frac{R^{(i)}}{\binom{K_w}{i}(1-\delta_w) F}, \frac{R^{(i)} / \binom{K_w}{i} + K_s R^{(i+1)} / \binom{K_w}{i+1}}{(1-\delta_s) F} \right\} \leq \sum_{m=1}^{K_s} \beta_{q-i+1,j,m} = \beta_{q-i+1,j},$$

$$\text{for } i = q-1, ..., p, \text{ and } j = 1, ..., \binom{K_w}{i+1}. \tag{62}$$

By adding all $\binom{K_w}{i+1}$ sub-messages, each weak receiver and each strong receiver can decode their

targeted contents after receiving message $q - i + 1$, if

$$\max \left\{ \frac{\frac{K_w - i}{i+1} R^{(i)}}{(1 - \delta_w) F}, \frac{\frac{K_w - i}{i+1} R^{(i)} + K_s R^{(i+1)}}{(1 - \delta_s) F} \right\} \leq \sum_{j=1}^{\binom{K_w}{i+1}} \beta_{q-i+1,j} = \beta_{q-i+1}, \quad for \ i = q-1, ..., p.$$

(63)

According to (38), we have

$$R^{(i+1)} = \frac{\gamma(p, \boldsymbol{\delta}_{ws}, i+1)}{\gamma(p, \boldsymbol{\delta}_{ws}, i)} R^{(i)} = \frac{K_w - i}{K_s(i+1)} \left(\frac{1 - \delta_s}{1 - \delta_w} - 1 \right) R^{(i)},$$

(64)

where we used

$$\gamma(p, \boldsymbol{\delta}_{ws}, i+1) = \frac{K_w - i}{K_s(i+1)} \left(\frac{1 - \delta_s}{1 - \delta_w} - 1 \right) \gamma(p, \boldsymbol{\delta}_{ws}, i)$$

(65)

from (20c). By substituting (64) into (63), the second term of the maximization in (63) can be rewritten as

$$\frac{\frac{K_w - i}{i+1} R^{(i)} + \frac{K_w - i}{i+1} \left(\frac{1 - \delta_s}{1 - \delta_w} - 1 \right) R^{(i)}}{(1 - \delta_s) F} = \frac{\frac{K_w - i}{i+1} R^{(i)}}{(1 - \delta_w) F},$$

(66)

which is equal to the first term of the maximization in (63). Thus, (63) can be simplified as

$$\frac{\frac{K_w - i}{i+1} R^{(i)}}{(1 - \delta_w) F} \leq \beta_{q-i+1}, \quad for \ i = q-1, ..., p.$$

(67)

By combining (49), (67), and (58), we have

$$\sum_{i=p}^{q} \left(\frac{\frac{K_w - i}{i+1} R^{(i)}}{(1 - \delta_w) F} \right) + \frac{K_s R^{(p)}}{(1 - \delta_s) F} \leq \sum_{i=p-1}^{q} \beta_{q-i+1} = 1.$$

(68)

Replacing $R^{(i)}$, for $i = p, ..., q$, with the expression in (38) for the homogeneous scenario results in

$$R \leq \frac{F \sum_{i=p}^{q} \gamma(p, \boldsymbol{\delta}_{ws}, i)}{\frac{1}{1 - \delta_w} \sum_{i=p}^{q} \left(\frac{K_w - i}{i+1} \gamma(p, \boldsymbol{\delta}_{ws}, i) \right) + \frac{K_s}{1 - \delta_s}},$$

(69)

where $\gamma(p, \boldsymbol{\delta}_{ws}, i)$ is given in (20c). Observe that (69) together with the cache capacity of each weak receiver given in (41) confirm the achievability of the memory-rate pair in (20) for the homogeneous scenario.

Remark 5. *It is to be noted that the achievable STW scheme introduced in [31] is a special case of the SCC scheme for the homogeneous scenario with $q = p + 1$. The delivery phase of the*

STW scheme is performed by delivering three messages in orthogonal time periods. The SCC scheme utilizes a more flexible caching and coding scheme which applies a finer subpacketization compared to [31], together with the joint encoding scheme of [37], also used in [31], enabling all the receivers to exploit the cache capacities of the weak receivers.

Remark 6. *The performance gain from the proposed SCC scheme is relatively higher when the erasure probabilities of the two sets of receivers are more disparate, i.e., the difference between the erasure probability of the best receiver among the weak receivers and the erasure probability of the worst receiver among the strong receivers is larger (see Fig. 5). Furthermore, the SCC scheme is relatively robust to skewed erasure probabilities across the strong receivers; this is due to the subpacketization performed to deliver the sub-messages of messages 2 to $q - p + 1$, where the sub-messages targeted to strong receivers are delivered over orthogonal time slots. On the other hand, it suffers more from skewed erasure probabilities across the weak receivers, which is due to the multicast nature of cache-aided coded content delivery to a group of weak receivers, which should be decoded by all of them simultaneously. Accordingly, the reliable rate at which the coded content is delivered to the intended set of weak receivers, is limited by the worst one.*

V. Conclusions

We have studied cache-enabled content delivery over a packet erasure BC with arbitrary erasure probabilities. We have defined the capacity of this network as the maximum common rate of contents in the library, which allows reliable delivery to all the receivers, independent of their demands. We have derived a lower bound on the capacity by proposing a novel caching and delivery scheme, which enables each receiver, even the strong receivers without a cache memory, to benefit from the cache memories available at the weak receivers. The proposed scheme utilizes a finer subpacketization of the files in the library, and provides a better exploitation of the available cache memories with a higher achievable rate than the state-of-the-art [31], which focuses on the homogeneous scenario, where the receivers in the same set all have the same erasure probability. This model and the presented results illustrate that even limited storage can be converted into spectral efficiency in noisy communication network, benefiting the whole network, if it is placed strategically across the network, and exploited intelligently.

References

[1] L. W. Dowdy and D. V. Foster, "Comparative models of the file assignment problem," *ACM Comput. Surv.*, vol. 14, pp. 287–313, Jun. 1982.

[2] K. C. Almeroth and M. H. Ammar, "The use of multicast delivery to provide a scalable and interactive video-on-demand service," *IEEE J. Sel. Areas Commun.*, vol. 14, no. 6, pp. 1110–1122, Aug. 1996.

[3] A. Meyerson, K. Munagala, and S. Plotkin, "Web caching using access statistics," in *Proc. ACM-SIAM SODA*, Washington, D.C., USA, Jan. 2001, pp. 354–363.

[4] I. Baev, R. Rajaraman, and C. Swamy, "Approximation algorithms for data placement problems," *SIAM J. Comput.*, vol. 38, no. 4, pp. 1411–1429, Jul. 2008.

[5] S. Borst, V. Gupta, and A. Walid, "Distributed caching algorithms for content distribution networks," in *Proc. IEEE INFOCOM*, San Diego, CA, Mar. 2010, pp. 1–9.

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

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

[8] 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.

[9] M. Mohammadi Amiri and D. Gündüz, "Fundamental limits of coded caching: Improved delivery rate-cache capacity trade-off," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 806–815, Feb. 2017.

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

[11] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," *arXiv:1609.07817v1 [cs.IT]*, Sep. 2016.

[12] C. Tian, "Symmetry, outer bounds, and code constructions: A computer-aided investigation on the fundamental limits of caching," *arXiv:1611.00024v1 [cs.IT]*, Oct. 2016.

[13] A. Sengupta, R. Tandon, and T. C. Clancy, "Improved approximation of storage-rate tradeoff for caching via new outer bounds," in *Proc. IEEE Int'l Symp. on Inform. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 1691–1695.

[14] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," in *Proc. IEEE Int'l Symp. on Inform. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 1696–1700.

[15] C.-Y. Wang, S. H. Lim, and M. Gastpar, "A new converse bound for coded caching," in *Proc. Information Theory and Applications Workshop (ITA)*, La Jolla, CA, Jan.-Feb. 2016, pp. 1–6.

[16] M. Gregori, J. Gomez-Vilardebo, J. Matamoros, and D. Gündüz, "Wireless content caching for small cell and D2D networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1222–1234, May 2016.

[17] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless D2D networks," *IEEE Trans. Inform. Theory*, vol. 62, no. 2, pp. 849–869, Dec. 2015.

[18] R. Pedarsani, M. A. Maddah-Ali, and U. Niesen, "Online coded caching," in *Proc. IEEE Int'l Conf. Commun. (ICC)*, Sydney, Australia, Jun. 2014, pp. 1878–1883.

[19] U. Niesen and M. A. Maddah-Ali, "Coded caching with nonuniform demands," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Toronto, ON, Apr. 2014, pp. 221–226.

[20] J. Zhang, X. Lin, and X. Wang, "Coded caching under arbitrary popularity distributions," in *Proc. Infor. Theory Applications Workshop (ITA)*, San Diego, CA, Feb. 2015, pp. 98–107.

[21] J. Zhang, X. Lin, C. C. Wang, and X. Wang, "Coded caching for files with distinct file sizes," in *Proc. IEEE Int'l Symp. on Inform. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 1686–1690.

[22] S. Wang, W. Li, X. Tian, and H. Liu, "Coded caching with heterogenous cache sizes," *arXiv:1504.01123v3 [cs.IT]*, Apr. 2015.

[23] M. Mohammadi Amiri, Q. Yang, and D. Gündüz, "Decentralized coded caching with distinct cache capacities," *arXiv:1601.05690v1 [cs.IT]*, Nov. 2016.

[24] N. Karamchandani, U. Niesen, M. A. Maddah-Ali, and S. Diggavi, "Hierarchical coded caching," in *Proc. IEEE Int'l Symp. on Inform. Theory (ISIT)*, Honolulu, HI, Jun. 2014, pp. 2142–2146.

[25] Q. Yang and D. Gündüz, "Centralized coded caching for heterogeneous lossy requests," *arXiv:1604.08178v1 [cs.IT]*, Apr. 2016.

[26] P. Hassanzadeh, E. Erkip, J. Llorca, and A. Tulino, "Distortion-memory tradeoffs in cache-aided wireless video delivery," *arXiv:1511.03932v1 [cs.IT]*, Nov. 2015.

[27] R. Timo, S. B. Bidokhti, M. Wigger, and B. Geiger, "A rate-distortion approach to caching," in *Proc. International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2016.

[28] J. Zhang and P. Elia, "Fundamental limits of cache-aided wireless BC: Interplay of coded-caching and CSIT feedback," *IEEE Trans. Inform. Theory*, vol. 63, no. 5, pp. 3142–3160, May 2017.

[29] A. Ghorbel, M. Kobayashi, and S. Yang, "Content delivery in erasure broadcast channels with cache and feedback," *IEEE Trans. Inform. Theory*, vol. 62, no. 11, pp. 6407–6422, Nov. 2016.

[30] R. Timo and M. Wigger, "Joint cache-channel coding over erasure broadcast channels," in *Proc. IEEE Int'l Symp. on Wirel. Commun. Systems (ISWCS)*, Brussels, Belgium, Aug. 2015, pp. 201–205.

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

[32] 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.

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

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

[35] J. Zhang and P. Elia, "Wireless coded caching: A topological perspective," *arXiv:1606.08253v1 [cs.IT]*, Jun. 2016.

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

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

[38] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge, UK: Cambridge University Press, 2012.