# Cache-Aided Two-User Broadcast Channels with State Information at Receivers

Hadi Reisizadeh
Department of ECE
University of Minnesota
Email: hadir@umn.edu

Mohammad Ali Maddah-Ali
Nokia Bell Labs
Holmdel, NJ
Email: mohammad.maddahali@nokia-bell-labs.com

Soheil Mohajer
Department of ECE
University of Minnesota
Email: soheil@umn.edu

***Abstract*—A two-user coded caching problem is studied in a joint source-channel coding framework. A source generates symbols at a certain rate for each file in the database, and a fixed fraction of the symbols are cached at each user. The delivery phase of coded caching takes place over a time-varying erasure broadcast channel, where the channel state information is only available at the receivers. The maximum source rate to keep up with the ergodic rate of both users is characterized.**

## I. Introduction

Coded caching is a promising strategy to overcome the rapidly growing traffic load of networks during their peak-traffic time. A caching system operates in two phases: (1) a placement phase, where each user has access to the database of the server and stores some packets form the database, and (2) a delivery phase, during which the server broadcasts a signal to all the receivers, such that each user be able to decode his desired file from his cache content and his received signal. In [1], authors assumed a *perfect* channel model for the delivery phase, and showed that a significant gain can be achieved by sending coded packets and simultaneously serving multiple users. The gain of caching is a function of the size of the cache available at each user. The exact trade-off between the memory and load of delivery is characterized [2], under the assumption of uncoded placement. In practice, the perfect channel assumption fails, and we are dealing with the randomness of the channel. Coded caching is studied under several wireless channel models, such as erasure and fading channels [3]–[5].

In a cellular system, the channel is time-varying and sending the channel state information (CSI) from the receivers to the transmitter over a feedback link is costly. The *ergodic capacity region* of such broadcast channels with no CSI at the transmitter is studied in [6], for a two-user system and an arbitrary fading distribution. In [6], the fading effect is studied using the deterministic model [7], where depending on the instantaneous channel strength, each receiver only gets the most significant bits, and of the transmit signal. The main challenge for the transmitter here is to allocate the transmit bits/levels to the messages intended to each user, without having the realization of the channels, which consists of the number of bits delivered to each user.

The work is supported in part by the National Science Foundation under Grant CCF-1749981.

In this work, we generalize the result of [6] by providing each user with a cache, that can store a $\mu$ fraction of the entire database. Hence, we are dealing with a *joint source-channel coding* problem: a source keeps generating a number of files at a certain rate. A fixed fraction of each file is available at each receiver. Without having the channel state information, the server is supposed to broadcast information such that each user can decode his desired file. The question is to characterize the maximum source rate for which the delivery rate of the channels (in an ergodic sense) can keep up with the generation of the source. We fully characterize the maximum achievable source rate for different regimes of cache size, by providing an explicit channel (level) allocation for each sub-message, and proving a matching upper bound. While the broadcast channel is not degraded in general, we use an approach similar to [6] to prove the optimality of the scheme.

In the following, we formulate the problem in Section II, and present the main result in Section III. The achievability and converse proofs are provided in Sections IV and V, respectively. The detailed proofs are postponed to the Appendix.

*Notation:* Throughout this paper, we denote the set of integers $\{1, 2, \ldots, N\}$ by $[N]$. For a binary vector of length $q$, i.e., $X \in \mathbb{F}_2^q$, and for a pair of integers $a < b$, we use the short hand notation $X(a : b)$ to denote $[X(a), X(a+1), \ldots, X(b)]^T$.

## II. Problem Formulation

### A. Channel Model

We consider a deterministic version of a 2-user time-varying memoryless fading broadcast channel [7], modeled by

$$Y_{u,t} = D^{q-L_u[t]} X_t = X_t(1 : L_u[t]), \qquad u = 1, 2,$$

where $X_t, Y_{1,t}, Y_{2,t} \in \mathbb{F}_2^q$, and $D$ is a $q \times q$ matrix, given by

$$D = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}.$$

Here $L_u[t]$ with $0 \le L_u[t] \le q$ determines the number of bits delivered to user $u$ at time $t$ (see Fig. 1). The channel state at user $u$, i.e., $\{L_u[t] | t = 1, \ldots, n\}$, is an i.i.d. random sequence generated according to some probability

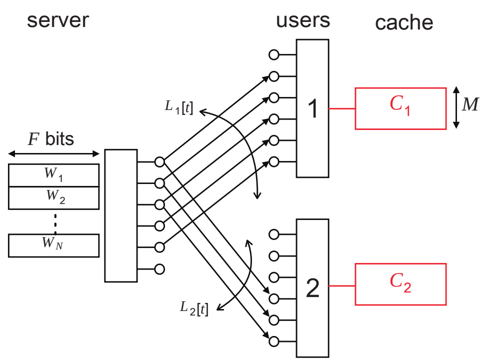


Fig. 1: A time-varying 2-user deterministic channel with cache: $q = 6$, $L_1[t] = 5$, and $L_2[t] = 3$.

mass function (PMF) $P_{L_u}()$ P[$L_u$ = ]. We denote the complementary cumulative distribution function of $L_u$ by $\bar{F}_{L_u}()$ P[$L_u$ ≥ ]. We assume that the channel state information (CSI) is casually known at the receivers. However, the transmitter only knows the channel statistics $P_{L_1}()$ and $P_{L_2}()$, but not their instantaneous realizations.

*B. Cache Model*

We assume each user is equipped with a cache $C_u$ of size $MF$ bits. The server has access to a database of $N$ independent files $W_1, \ldots, W_N$ each of size $nf$ $F$ bits, i.e.,

$$W_j \in \{1, 2, \ldots, 2^F\}, \quad j \in [N].$$

The system operates in two phases, namely, the *placement phase* and the *delivery phase*. In the placement phase, the cache memory of each user is filled with *uncoded* bits of the files in a central manner, that is,

$$C_u = \bigcup_{j\in[N]} C_{u,j}, \qquad u = 1, 2,$$

where $C_{u,j}$ denotes the part of the cache of user $u$ filled by bits of file $j$. Therefore, we have $H(C_{u,j} | W_j) = 0$, $I(C_{u,} ; W_j) = 0$ for $= j$, and $\sum_{j\in[N]} H(C_{u,j}) \leq MF$ for $u = 1, 2$.

After the completion of the placement phase, each user requests one of the $N$ files, where all files are equally likely to be requested. We denote the request of user $u \in \{1, 2\}$ by $d_u \in [N]$, and define $\mathbf{d} = (d_1, d_2)$[1]. Once the requests are revealed to the server, it forms a broadcasting message $X = \psi_{\mathbf{d}}(W_1, \ldots, W_N; C_1, C_2)$, where

$$\psi^{(n)} : \{1, 2, \ldots, 2^F\}^N \to \{1, 2, \ldots, 2^{qn}\},$$

and transmits it over the broadcast channel during the delivery phase. Upon receiving $Y_u^n = (Y_{u,1}, Y_{u,2}, \ldots, Y_{u,n})$, user $u = 1, 2$ should be able to decode its desired file using its cache content $C_u$ and the received message $Y_u^n$, i.e.,

$$\hat{W}_{d_u} = \varphi_u^{(n)}(Y_u^n, C_u).$$

The overall decoding error probability is defined as

$$P_e^{(n)} \quad \mathrm{P}[\hat{W}_{d_1} = W_{d_1}] + \mathrm{P}[\hat{W}_{d_2} = W_{d_2}].$$

[1]In this work, we consider the worst case scenario, i.e., $d_1 = d_2$.

Definition 1. A source rate $f$ is called achievable if there exists a sequence of placement strategies and encoding and decoding functions $\psi^{(n)}, \varphi_1^{(n)}, \varphi_2^{(n)}$ for a database of files with $F = nf$ bits such that $P_e^{(n)} \to 0$ as $n$ grows.

Our goal is to characterize the maximum achievable source rate $\bar f$ for a deterministic BC with channel statistics $\bar{F}_{L_1}$ and $\bar{F}_{L_2}$, i.e.,

$$\bar f = \max_{f:\text{achievable}} f.$$

## III. Main Results

Before presenting the main result, we discuss the following theorem, in which the achievable rate region of the deterministic BC with no CSIT (and without cache) is characterized.

Theorem 1. *[6, Theorem 2] For an $\omega \geq 0$ we define $C_{\mathrm{NC}}(\omega)$ as the set of $(r_1, r_2)$ satisfying*

$$r_1 \leq R_1(\omega) \qquad \sum_{:g()\leq\omega} \bar{F}_{L_1}(), \tag{1}$$

$$r_2 \leq R_2(\omega) \qquad \sum_{:g()>\omega} \bar{F}_{L_2}(), \tag{2}$$

*where $g() = \frac{\bar{F}_{L_2}()}{\bar{F}_{L_1}()}$. Then the capacity region $C_{\mathrm{NC}}$ of the $q$ bit layered erasure broadcast channel (with no cache) is the convex hull of the union of $C_{\mathrm{NC}}(\omega)$ over all $\omega \geq 0$, that is,*

$$C_{\mathrm{NC}} = \mathrm{conv} \left( \bigcup_{\omega\geq 0} C_{\mathrm{NC}}(\omega) \right).$$

The next theorem is the main result of this paper, in which the maximum achievable source rate of the cache-aided deterministic BC with no CSIT is characterized.

Theorem 2. *The maximum achievable source rate of a 2-user time-varying deterministic broadcast channel with per-file cache rate $\mu = M/N$, for $\mu \leq \frac{1}{2}$ is given by*

$$\bar f = \min \left\{ \min_{\omega\geq 1} \frac{\omega R_1(\omega) + R_2(\omega)}{1 - 2\mu + \omega(1-\mu)}, \min_{0\leq\omega\leq 1} \frac{R_1(\omega) + \frac{1}{\omega} R_2(\omega)}{(1-2\mu) + \frac{1}{\omega}(1-\mu)} \right\}, \tag{3}$$

*and for $\mu \geq \frac{1}{2}$ is characterized by*

$$\bar f = \min \left\{ \frac{\sum_{=1}^{q} \bar{F}_{L_1}()}{1-\mu}, \frac{\sum_{=1}^{q} \bar{F}_{L_2}()}{1-\mu} \right\}. \tag{4}$$

## IV. The Achievable Scheme

This section is dedicated to show the achievability of the source rates defined in Theorem 2.

*A.* $0 \leq \mu \leq \frac{1}{2}$

Without loss of generality, assume the minimum in (3) is obtained in the first term, and by some $\omega \geq 1$. That is

$$\bar f = \frac{G(\omega)}{(1-2\mu) + \omega(1-\mu)},$$

where we define $G(\omega)$ $\omega R_1(\omega) + R_2(\omega)$.

In the placement phase we split each file $W_j$ into three *non-overlapping* subfiles $W_{j,1}$, $W_{j,2}$, $W_{j,\varnothing}$ of sizes $\mu F$, $\mu F$, and

$(1-2\mu)F$, respectively. Then, each user $u \in \{1,2\}$ stores all subfiles $W_{j,u}$ for $j \in [N]$, i.e., $C_{u,j} = W_{j,u}$, for $j \in [N]$.

Upon receiving the request vector $\boldsymbol{d} = (d_1, d_2)$, the server needs to send $(W_{d_1,2}, W_{d_1,\varnothing})$ to User 1, and $(W_{d_2,1}, W_{d_2,\varnothing})$ to User 2. To this end, we can consider the private messages $W_{d_1,\varnothing}$ and $W_{d_2,\varnothing}$, intended for Users 1 and 2, respectively, and the common message $W_{d_1,2} \oplus W_{d_2,1}$ intended for both users. Note that the size of the private messages is $(1-2\mu)f^\star n$ bits, and the size of the common message is $\mu f^\star n$.

Intuitively, the achievability proof of $f^\star$ is equivalent to allocation of levels of the deterministic channel to the two private messages and the common message, so that each user can decode the intended messages. We start by sorting all the $q$ levels of the channel according to $g(\ell) = \frac{\bar{F}_{L_2}(\ell)}{\bar{F}_{L_1}(\ell)}$ in an *increasing* order, and labeling them by $\{\ell_1, \dots, \ell_q\}$. Therefore, we have $g(\ell_1) \le g(\ell_2) \le \dots \le g(\ell_q)$. We also define $\omega_i = g(\ell_i)$ for $i = 1, \dots, q$, implying $\omega_1 \le \omega_2 \le \dots \le \omega_q$. We first have the following lemma, which is proved in Appendix A.

**Lemma 1.** If $\omega^\star$ is the minimizer of (3), then $\omega^\star \in \{\omega_1, \omega_2, \dots, \omega_q\}$, and therefore, there exists $k \in [q]$ such that

$$(1-2\mu)f^\star + \omega_k(1-\mu)f^\star = G(\omega_k). \tag{5}$$

The next lemma (proved in Appendix A) shows the existence of two disjoint subsets of levels that can guarantee rates of $R_1^p = (1-2\mu)f^\star$ and $R_2^p = (1-2\mu)f^\star$ for Users 1 and 2, respectively. Note that the lemma determines the levels allocated to transmission of the private messages.

**Lemma 2.** There exist integers $i, k \in \{1, \dots, q\}$ with $i \le k$ (where $k$ is defined in Lemma 1), and constants $\alpha, \beta \in [0,1]$ such that

$$\sum_{j=1}^{i-1} \bar{F}_{L_1}(\ell_j) + \beta \bar{F}_{L_1}(\ell_i) = R_1^p = (1-2\mu)f^\star,$$
$$\alpha \bar{F}_{L_2}(\ell_k) + \sum_{j=k+1}^{q} \bar{F}_{L_2}(\ell_j) = R_2^p = (1-2\mu)f^\star.$$

The remaining levels will be used for multicasting the common message. The rate supported for User $u$ is given by

$$R_u^c \triangleq \bar{\beta} \cdot \bar{F}_{L_u}(\ell_i) + \sum_{j=i+1}^{k-1} \bar{F}_{L_u}(\ell_j) + \bar{\alpha} \cdot \bar{F}_{L_u}(\ell_k), \tag{6}$$

for $u = 1, 2$, where $\bar{\alpha} = 1-\alpha$, and $\bar{\beta} = 1-\beta$. It remains to show that the common message can be reliably sent to both users over these levels. This is formally stated in the following lemma, which is proved in Appendix A.

**Lemma 3.** The channel rates defined in (6) satisfy

$$R_2^c \ge R_1^c = \mu f^\star. \tag{7}$$

*B.* $\frac{1}{2} < \mu \le 1$

In this regime, the *placement* phase consists of splitting each file $W_j$ into three subfiles $W_{j,1}$, $W_{j,2}$ and $W_{j,\{1,2\}}$ of sizes, $(1-\mu)fn$, $(1-\mu)fn$, and $(2\mu-1)fn$ bits, respectively. Then user $u \in \{1,2\}$ will store $W_{j,u}$ and $W_{j,\{1,2\}}$ for each file $j \in [N]$. Upon receiving the vector $\boldsymbol{d} = (d_1, d_2)$, the server needs to only multicast a common message $W_{d_1,2} \oplus W_{d_2,1}$ to both users. The size of this common message is $(1-\mu)fn$. Hence, the maximum achievable source rate is given the capacity of the channel of the weak user, which is

$$f^\star = \min\left( \frac{\sum_{\ell=1}^q \bar{F}_{L_1}(\ell)}{1-\mu}, \frac{\sum_{\ell=1}^q \bar{F}_{L_2}(\ell)}{1-\mu} \right).$$

## V. Proof of the Optimality

Next, we derive an upper bound on the achievable source rate, stated following lemma, and proved in Appendix A.

**Lemma 4.** For any *uncoded* cache placement with cache contents $C_1$ and $C_2$, the source rate of the physically degraded memoryless BC described by $P_{Y_1,Y_2|X}$ is the union over all pairs of $(U, X)$ such that $U \leftrightarrow X \leftrightarrow Y_2 \leftrightarrow Y_1$ satisfying

$$f \le f_1 \triangleq \frac{I(U;Y_1)}{1-\mu}, \qquad f \le f_2 \triangleq \frac{I(X;Y_2|U)}{1-2\mu}.$$

Now, we are ready to prove the optimality of $f^\star$ in Theorem 2. We start with the case of $\mu \le \frac{1}{2}$. Note that for arbitrary $\bar{F}_{L_1}$ and $\bar{F}_{L_2}$, the broadcast channel is not degraded. In order to apply Lemma 4, we first enhance the channel of User 2 by replacing the fading distribution $L_2$ by $\tilde{L}_2$, defined as,

$$\bar{F}_{\tilde{L}_2}(\ell) = \min\left(1, \max(\bar{F}_{L_2}(\ell), \omega \bar{F}_{L_1}(\ell))\right),$$

for some $\omega \ge 1$, and define $\tilde{Y}_2 = D^{q-\tilde{L}_2} X = X(1:\tilde{L}_2)$. Now, we have $X \leftrightarrow \tilde{Y}_2 \leftrightarrow Y_1$. Moreover, note that in an erasure channel, the receiver observation contains the channel realization, i.e., $(Y_1, L_1)$ and $(\tilde{Y}_2, \tilde{L}_2)$. Then, given the fact that CSI is available at the receivers, the terms in Lemma 4 for the deterministic channel of interest will be simplified to

$$I(U;Y_1,L_1) = I(U;X(1:L_1),L_1) = \sum_{i=1}^q P_{L_1}(i) I(U;X(1:i))$$
$$= \sum_{i=1}^q P_{L_1}(i) \sum_{\ell=1}^i I(U;X(\ell)|X(1:\ell-1))$$
$$= \sum_{\ell=1}^q I(U;X(\ell)|X(1:\ell-1)) \sum_{i=\ell}^q P_{L_1}(i)$$
$$= \sum_{\ell=1}^q \bar{F}_{L_1}(\ell) I(U;X(\ell)|X(1:\ell-1))$$
$$= \sum_{\ell=1}^q \bar{F}_{L_1}(\ell)[H(X(\ell)|X(1:\ell-1)) - H(X(\ell)|U,X(1:\ell-1))].$$

Similarly, we have

$$I(X;\tilde{Y}_2,\tilde{L}_2|U) = \sum_{\ell=1}^q \bar{F}_{\tilde{L}_2}(\ell) H(X(\ell)|U,X(1:\ell-1)).$$

Hence, we can upper bound the following weighted sum as

$$\begin{aligned} f(\omega) &= \omega(1-\mu)f + (1-2\mu)f \\ &\le \omega I(U;Y_1,L_1) + I(X;\tilde{Y}_2,\tilde{L}_2|U) \\ &= \sum_{\ell=1}^q (\tilde{g}(\ell) - \omega)\bar{F}_{L_1}(\ell) H(X(\ell)|X(1:\ell-1),U) \\ &\quad + \omega \sum_{\ell=1}^q \bar{F}_{L_1}(\ell) H(X(\ell)|X(1:\ell-1)), \end{aligned} \tag{8}$$

where $\tilde{g}(\ell) = \frac{F_{\tilde{L}_2}(\ell)}{F_{L_1}(\ell)}$. The terms in the second summation in (8) will be maximized by an be i.i.d. Bernoulli random variable choice for $X_1, \ldots, X_q$. The terms in the first summation can be maximized if

$$H(X(\ell)|X(1:\ell-1), U) = \begin{cases} 1 & \tilde{g}(\ell) > \omega, \\ 0 & \tilde{g}(\ell) \le \omega, \end{cases} \quad (9)$$

which can be satisfied by an optimum choice for $U$, given by

$$U = \{X(\ell)|\tilde{g}(\ell) \le \omega\}.$$

Therefore, we have

$$\begin{aligned} f(\omega) &\le \omega \sum_{\ell:\tilde{g}(\ell)\le\omega} F_{L_1}(\ell) + \sum_{\ell:\tilde{g}(\ell)>\omega} \tilde{g}(\ell)\bar{F}_{L_1}(\ell) \\ &\overset{(a)}{=} \omega \sum_{\ell:g(\ell)\le\omega} F_{L_1}(\ell) + \sum_{\ell:g(\ell)>\omega} F_{\tilde{L}_2}(\ell) \\ &\overset{(b)}{=} \omega \sum_{\ell:g(\ell)\le\omega} F_{L_1}(\ell) + \sum_{\ell:g(\ell)>\omega} F_{L_2}(\ell) \\ &= \omega R_1(\omega) + R_2(w), \end{aligned} \quad (10)$$

where $R_1(\omega)$ and $R_2(\omega)$ are defined in (1) and (2) and $(a)$ holds since $\tilde{g}(\ell)\bar{F}_{L_1}(\ell) = \bar{F}_{\tilde{L}_2}(\ell)$ and $\{\ell|\tilde{g}(\ell) > \omega\} = \{\ell|g(\ell) > \omega\}$, $(b)$ follows the fact that $F_{\tilde{L}_2}(\ell) = \bar{F}_{L_1}(\ell)$ whenever $g(\ell) > \omega$. Therefore, (10) proves the first minimization in Theorem 2.

For $0 \le \omega \le 1$, we can repeat the steps in (8) through (10) by swapping the labels of the users and replacing $\omega$ by $\frac{1}{\omega}$. Under this reversed labels, we now enhance the channel of User 1, and get the second minimization in Theorem 2. The details of the proof are omitted due to the page limit.

Finally, for $\mu > \frac{1}{2}$, we explore the cut-set bound, i.e.,

$$\begin{aligned} (1-\mu)f \le I(U; Y_1, L_1) &= \sum_{\ell=1}^{q} F_{L_1}(\ell) I(U; X(\ell)|X(1:\ell-1)) \\ &\le \sum_{\ell=1}^{q} F_{L_1}(\ell) H(X(\ell)|X(1:\ell-1)) \le \sum_{\ell=1}^{q} F_{L_1}(\ell). \end{aligned}$$

Similarly, for the other user, we get $(1-\mu)f \le \sum_{\ell=1}^{q} F_{L_2}(\ell)$. This completes the proof.

## References


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

[2] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1281–1296, 2018.

[3] S. S. Bidokhti, M. Wigger, and A. Yener, "Benefits of cache assignment on degraded broadcast channels," in *Information Theory (ISIT), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 1222–1226.

[4] M. M. Amiri and D. Gündüz, "Cache-aided content delivery over erasure broadcast channels," *IEEE Transactions on Communications*, vol. 66, no. 1, pp. 370–381, 2018.

[5] K.-H. Ngo, S. Yang, and M. Kobayashi, "Scalable content delivery with coded caching in multi-antenna fading channels," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 548–562, 2018.

[6] N. David and R. D. Yates, "Fading broadcast channels with state information at the receivers," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3453–3471, 2012.

[7] A. S. Avestimehr, S. N. Diggavi, and N. David, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information theory*, vol. 57, no. 4, pp. 1872–1905, 2011.


## Appendix A
## Proof of the Lemmas

*Proof of Lemma 1.* Assume otherwise holds, and we have [2] $\omega_s < \omega^\star < \omega_{s+1}$ for some $s$. Pick a pair $(\omega^-, \omega^+)$ such that $\omega_s < \omega^- < \omega^\star < \omega^+ < \omega_{s+1}$. Note that functions $R_1(\omega)$ and $R_2(\omega)$ remain constant over the interval $(\omega_s, \omega_{s+1})$, and we thus can denote such constants by $A$ and $B$, respectively. Similarly, we can define $C = 1-\mu$ and $D = 1-2\mu$. Then for any $\omega \in (\omega_s, \omega_{s+1})$ we define $h(\omega) = (A\omega + B)/(C\omega + D)$ which is a monotonic function, and hence we either have $h(\omega^-) < h(\omega^\star)$ or $h(\omega^+) < h(\omega^\star)$, which is in contradiction with optimality of $w^\star$. □

*Proof of Lemma 2.* It is easy to verify that

$$\begin{aligned} G(\omega_{j+1}) - G(\omega_j) &= \omega_{j+1} F_{L_1}(\ell_{j+1}) + (\omega_{j+1} - \omega_j)R_1(\omega_j) \\ &\quad - F_{L_2}(\ell_{j+1}) = (\omega_{j+1} - \omega_j)R_1(\omega_j), \end{aligned}$$

for every $i$. Moreover, optimality of $f^\star$ implies $(1-2\mu)f^\star + \omega_{k+1}(1-\mu)f^\star \le G(\omega_{k+1})$, which together with (5) implies

$$\begin{aligned} (\omega_{k+1} - \omega_k)(1-\mu)f^\star &\le G(\omega_{k+1}) - G(\omega_k) \\ &= (\omega_{k+1} - \omega_k)R_1(\omega_k). \end{aligned} \quad (11)$$

Similarly, from $(1-2\mu)f^\star + \omega_{k-1}(1-\mu)f^\star \le G(\omega_{k-1})$ and (5) we get

$$\begin{aligned} (\omega_k - \omega_{k-1})(1-\mu)f^\star &\ge G(\omega_k) - G(\omega_{k-1}) \\ &= (\omega_k - \omega_{k-1})R_1(\omega_{k-1}). \end{aligned} \quad (12)$$

From (11) and (12) we conclude $R_1(\omega_{k-1}) \le (1-\mu)f^\star \le R_1(\omega_k)$, which implies

$$\begin{aligned} (1-\mu)f^\star &= R_1(\omega_k) - \alpha\bar{F}_{L_1}(\ell_k) \\ &= \sum_{j=1}^{k-1} F_{L_1}(\ell_j) + \bar{\alpha} \cdot F_{L_1}(\ell_k) \end{aligned} \quad (13)$$

for some $\alpha \in [0,1]$. Then, from $(1-2\mu)f^\star \le (1-\mu)f^\star$, we can conclude existence of some $i \le k$ and $\beta \in [0,1]$ such that

$$(1-2\mu)f^\star = \sum_{j=1}^{i-1} F_{L_1}(\ell_j) + \beta\bar{F}_{L_1}(\ell_i). \quad (14)$$

Moreover, plugging (13) into (5) we get

$$\begin{aligned} (1-2\mu)f^\star &= G(\omega_k) - \omega R_1(\omega_k) + \omega_k \alpha\bar{F}_{L_1}(\ell_k) \\ &= R_2(\omega_k) + \omega_k \alpha\bar{F}_{L_1}(\ell_k) = R_2(\omega_k) + \alpha\bar{F}_{L_2}(\ell_k), \end{aligned}$$

which is the claim of the lemma. □

*Proof of Lemma 3.* Subtracting (14) from (13) we immediately get $R_1^c = \mu f^\star$. Next, note that

$$\begin{aligned} G(w_i) &= \omega_i \left[ \sum_{j=1}^{i-1} F_{L_1}(\ell_j) + \beta\bar{F}_{L_1}(\ell_i) \right] + \omega_i \bar{\beta} \cdot F_{L_1}(\ell_i) \\ &\quad + \left[ \sum_{j=i+1}^{k-1} F_{L_2}(\ell_j) + \bar{\alpha}F_{L_2}(\ell_k) \right] + \left[ \alpha\bar{F}_{L_2}(\ell_k) + \sum_{j=k+1}^{q} F_{L_2}(\ell_j) \right] \\ &\overset{(a)}{=} \omega_i R_1^p + R_2^c + R_2^p = (1+\omega_i)(1-2\mu)f^\star + R_2^c, \end{aligned}$$

[2]We define $\omega_0 = 1$ and $\omega_{q+1} = +\infty$ for completeness.

where $(a)$ follows the fact that $\omega_i F_{L_1}(\ell_i) = \bar{F}_{L_2}(\ell_i)$. Then recall that $f^{\star}$ is the minimum value obtained by (3). We can distinguish the following two cases.

(i) $\omega_i \geq 1$ : In this case from (3) we can write

$$f^{\star} \leq \frac{\omega_i R_1(\omega_i) + R_2(\omega_i)}{(1-2\mu) + \omega_i(1-\mu)} = \frac{(1+\omega_i)(1-2\mu)f^{\star} + R_2^c}{(1-2\mu) + \omega_i(1-\mu)},$$

which implies $R_2^c \geq \omega_i \mu f^{\star} \geq \mu f^{\star}$.

(ii) $\omega_i \leq 1$ : Alternatively, from (3) we have

$$f^{\star} \leq \frac{R_1(\omega_i) + \frac{1}{\omega_i} R_2(\omega_i)}{(1-2\mu) + \frac{1}{\omega_i}(1-\mu)} = \frac{\omega_i R_1(\omega_i) + R_2(\omega_i)}{\omega_i(1-2\mu) + (1-\mu)} = \frac{(1+\omega_i)(1-2\mu)f^{\star} + R_2^c}{\omega_i(1-2\mu) + (1-\mu)},$$

which implies $R_2^c \geq \mu f^{\star}$. □

*Proof of Lemma 4.* We follow an approach similar to that of [3]. By Fano's inequality, we have

$$H(W_{d_1}|Y_1^n, C_1) \leq n\epsilon_n, \qquad H(W_{d_2}|Y_2^n, C_2) \leq n\epsilon_n,$$

for some $\epsilon_n \to 0$ as $n \to \infty$. For given caches $C_1$ and $C_2$ and the sequence of requests $\boldsymbol{d} = (d_1, d_2)$, we have

$$\begin{aligned}
nf - n\epsilon_n &\leq H(W_{d_1}) - n\epsilon_n \leq I(W_{d_1}; Y_1^n, C_1) \\
&= I(W_{d_1}; Y_1^n|C_1) + I(W_{d_1}; C_1) \\
&= \sum_{i=1}^{n} I(W_{d_1}; Y_{1,i}|Y_1^{i-1}, C_1) + I(W_{d_1}; C_1) \\
&\leq \sum_{i=1}^{n} I(W_{d_1}, Y_1^{i-1}; Y_{1,i}|C_1) + I(W_{d_1}; C_1) \\
&= nI(W_{d_1}, Y_1^{Q-1}; Y_{1,Q}|C_1, Q) + I(W_{d_1}; C_1) \\
&\leq nI(W_{d_1}, C_1, Y_1^{Q-1}, Q; Y_{1,Q}) + I(W_{d_1}; C_1) \\
&= nI(U; Y_{1,Q}) + I(W_{d_1}; C_1) \\
&= nI(U; Y_1) + I(W_{d_1}; C_1),
\end{aligned} \tag{15}$$

where $Q$ is a random variable uniformly distributed over $[n]$ and independent of $(W_{d_1}, W_{d_2}, X^n, Y_1^n, Y_2^n, C_1, C_2)$ and $U = (W_{d_1}, C_1, Y_1^{Q-1}, Q)$. Similarly, we have

$$\begin{aligned}
nf - n\epsilon_n &= H(W_{d_2}) - n\epsilon_n \leq I(W_{d_2}; Y_2^n, C_2) \\
&= I(W_{d_2}; Y_2^n|C_2) + I(W_{d_2}; C_2) \\
&\leq I(W_{d_2}; Y_2^n, W_{d_1}, C_1|C_2) + I(W_{d_2}; C_2) \\
&= I(W_{d_2}; Y_2^n|W_{d_1}, C_1, C_2) + I(W_{d_2}; W_{d_1}|C_1, C_2) \\
&\quad + I(W_{d_2}; C_1|C_2) + I(W_{d_2}; C_2) \\
&\overset{(a)}{=} I(W_{d_2}; Y_2^n|W_{d_1}, C_1, C_2) + I(W_{d_2}; C_1, C_2) \\
&= \sum_{i=1}^{n} I(W_{d_2}; Y_{2,i}|W_{d_1}, C_1, C_2, Y_2^{i-1}) + I(W_{d_2}; C_1, C_2) \\
&\leq \sum_{i=1}^{n} I(W_{d_2}; Y_{2,i}, Y_1^{i-1}|W_{d_1}, C_1, C_2, Y_2^{i-1}) + I(W_{d_2}; C_1, C_2) \\
&\overset{(b)}{=} \sum_{i=1}^{n} I(W_{d_2}; Y_{2,i}|W_{d_1}, C_1, C_2, Y_1^{i-1} Y_2^{i-1}) + I(W_{d_2}; C_1, C_2) \\
&\leq \sum_{i=1}^{n} I(W_{d_2}, C_2, Y_2^{i-1}; Y_{2,i}|W_{d_1}, C_1, Y_1^{i-1}) + I(W_{d_2}; C_1, C_2) \\
&= nI(W_{d_2}, C_2, Y_2^{Q-1}; Y_{2,Q}|W_{d_1}, C_1, Y_1^{Q-1}, Q) + I(W_{d_2}; C_1, C_2) \\
&\leq nI(X_Q, W_{d_2}, C_2, Y_2^{Q-1}; Y_{2,Q}|U) + I(W_{d_2}; C_1, C_2) \\
&\overset{(c)}{=} nI(X_Q; Y_{2,Q}|U) + I(W_{d_2}; C_1, C_2) \\
&= nI(X; Y_2|U) + I(W_{d_2}; C_1, C_2).
\end{aligned} \tag{16}$$

In the above chain of inequalities, (a) holds since for an uncoded cache placement and independent files we have

$$\begin{aligned}
I(W_{d_2}; W_{d_1}|C_1, C_2) &= H(W_{d_2}|C_1, C_2) - H(W_{d_2}|W_{d_1}, C_1, C_2) \\
&= H(W_{d_2}|C_{1,d_1}, C_{2,d_1}) - H(W_{d_2}|C_{1,d_1}, C_{2,d_1}) = 0.
\end{aligned}$$

Moreover, $(b)$ follows the degradedness of the channel, $(W_{d_1}, W_{d_2}, C_1, C_2) \leftrightarrow X_i \leftrightarrow Y_{2,i} \leftrightarrow Y_{1,i}$, which implies

$$\begin{aligned}
I(W_{d_2}; Y_1^{i-1}|W_{d_1}, C_1, C_2, Y_2^{i-1}) &= H(Y_1^{i-1}|W_{d_1}, C_1, C_2, Y_2^{i-1}) \\
&\quad - H(Y_1^{i-1}|W_{d_1}, W_{d_2}, C_1, C_2, Y_2^{i-1}) \\
&= H(Y_1^{i-1}|Y_2^{i-1}) - H(Y_1^{i-1}|Y_2^{i-1}) = 0.
\end{aligned}$$

Finally, $(c)$ holds since condition on $X_Q$, the received signal $Y_{2,Q}$ is independent of all other variables, i.e.,

$$\begin{aligned}
I(W_{d_2}, C_2, Y_2^{Q-1}; Y_{2,Q}|X_Q, U) &= H(Y_{2,Q}|X_Q, U) - H(Y_{2,Q}|X_Q, U, W_{d_2}, C_2, Y_2^{Q-1}) \\
&= H(Y_{2,Q}|X_Q) - H(Y_{2,Q}|X_Q) = 0.
\end{aligned}$$

Since each file is equally likely to be requested, taking average of (15) and (16) over all *distinct* requests $d_1$ and $d_2$, i.e., $\mathcal{D} = \{(d_1, d_2) : d_1 \neq d_2, d_1, d_2 \in [N]\}$ provides an upper bound for any achievable rate $f$, i.e.,

$$nf - n\epsilon_n \leq nI(U; Y_1) + \frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{d} \in \mathcal{D}} I(W_{d_1}; C_1), \tag{17}$$

$$nf - n\epsilon_n \leq nI(X; Y_2|U) + \frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{d} \in \mathcal{D}} I(W_{d_2}; C_1, C_2). \tag{18}$$

For the last terms in inequalities, we have

$$\begin{aligned}
\frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{d} \in \mathcal{D}} I(W_{d_1}; C_1) &= \frac{1}{N} \sum_{j=1}^{N} I(W_j; C_1) \\
&= \frac{1}{N} \sum_{j=1}^{N} H(C_{1,j}) = \frac{1}{N} H(C_1) \leq \frac{1}{N} MF = \mu f.
\end{aligned} \tag{19}$$

and

$$\begin{aligned}
\frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{d} \in \mathcal{D}} I(W_{d_1}; C_1, C_2) &= \frac{1}{N} \sum_{j=1}^{N} I(W_j; C_1, C_2) \\
&= \frac{1}{N} \sum_{j=1}^{N} H(C_{1,j}, C_{2,j}) = \frac{1}{N} H(C_1, C_2) \leq \frac{2MF}{N} = 2\mu f.
\end{aligned} \tag{20}$$

Plugging (19) and (20) into (17) and (18), respectively, we get the desired bounds. □