Successive Refinement to Caching
for Dynamic Content

Pinar Sen
University of California, San Diego
San Diego, USA
Email: psen@ucsd.edu

Michael Gastpar
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
Email: michael.gastpar@epfl.ch

Abstract—To reduce the network load during peak hours, servers deliver partial data to users during the off-peak time of the network before the actual requests are known, which is known as *caching*. This paper studies a single user caching problem in which the file contents are subject to dynamic modifications with respect to a certain probability distribution. To cope with the dynamical nature of the file contents, a successive refinement approach to caching is presented: partial information of the original data is cached first and then if there is a modification, a refinement to the previously cached data is delivered to the user. Given a fixed cache memory, there is a tension between the rates of two cache descriptions. The problem of optimal caching strategies is formulated through a *successive* Gray-Wyner network, the optimal rate region of which is characterized. Some lower and upper bounds on the performance of optimal caching strategies are developed and shown to actually yield closed form solutions for certain classes of file contents.

I. INTRODUCTION

Due to the ever growing number of devices, networks encounter heavy traffic during the peak hours of the day. A recent solution to reduce the network traffic during these busy hours is to deliver partial data for future use before the database server knows which file is requested by the users [1]–[9]. In the standard (static) caching problems, communication is divided into three phases. In the *caching phase*, the database server delivers partial information about the files to users. The database server is informed which of the files are requested by the users in the *request phase*. Finally, in the *delivery phase*, the remaining part of the requested files is delivered. Taking an information theoretical approach, [4], [10] formulated the single user caching problem through its similarity to Gray-Wyner network [11] and discussed the optimal caching strategy. Extension of the theoretical analysis to multi-user caching problems can be found in [2], [5], [10], [12]. Building on rate distortion theory, the counterparts of those caching problems for lossy reconstruction were also discussed in [6]–[9].

On the other hand, the database has a *dynamic* nature in the sense that the contents of the files could be modified or completely changed during the three phases of the caching problem. For example, news websites are continuously updated throughout the day with the most current information. This dynamic nature is taken into account in the more recent work [13], which studied the multicast caching of dynamic content. However, they only consider the benefit of correlation between original and modified files in the delivery phase.

In our setting, we bring a *successive refinement* approach to caching phase and let the database server deliver a refinement to the cache contents after the modification occurs.

For ease of exposition, we concentrate on a database that initially stores two files (X_1, X_2), one of which is to be requested by the user uniformly at random. Our arguments naturally extend to more than two files. As illustrated in Fig. 1, the pair (X_1, X_2) could be modified into the pair (X'_1, X'_2) with a certain probability before the request phase. Given a fixed cache rate/memory, there is a tension between caching for only (X_1, X_2) as if there will be no modification on the content and sparing all the cache rate/memory for (X'_1, X'_2). What would be the optimal caching strategy in such a setup to minimize the expected delivery rate?
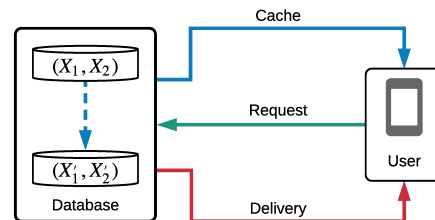

Fig. 1: Caching for dynamic content

Following similar steps to [4], we approach this question from an information theoretical perspective and construct a *successive* Gray-Wyner network to exploit the dynamic nature of the content files. We first characterize the optimal rate region for the successive Gray-Wyner network (cf. Section II) and then use these results to formulate the caching problem for dynamic content (cf. Section III). In particular, we concentrate on a special case in which the modification probability is 0.5 and provide some lower and upper bounds on the performance of optimal caching strategies. For certain classes of file contents, we show that our bounds yield closed form solutions.

II. SUCCESSIVE GRAY-WYNER NETWORK

Consider the *successive* Gray-Wyner network in Fig. 2, in which a tuple of sequences $(X_1^n, X_2^n, X_1'^n, X_2'^n)$ drawn i.i.d. from $p(x_1, x_2, x'_1, x'_2)$ is described by two encoders so that decoder i, having the descriptions M_0 and M_i, can losslessly recover X_i^n and decoder i', having the descriptions M_0, M'_0 and M'_i, can losslessly recover $X_i'^n$ for $i = 1, 2$.
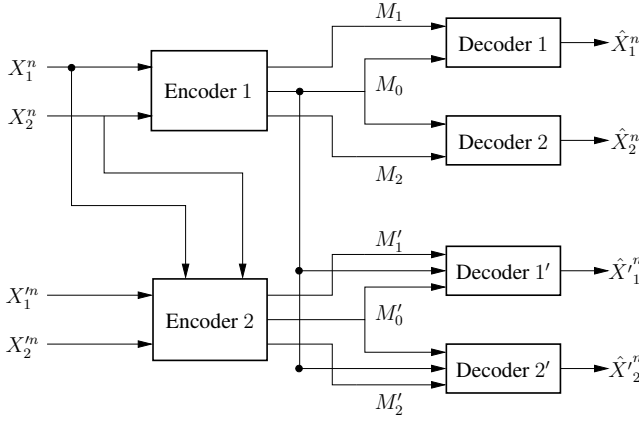
Fig. 2: Successive Gray-Wyner network

An $(nR_0, nR_1, nR_2, nR'_0, nR'_1, nR'_2, n)$ code for the successive Gray-Wyner network consists of two encoders and four decoders such that encoder 1 assigns index tuple $(m_0, m_1, m_2)(x_1^n, x_2^n) \in [2^{nR_0}] \times [2^{nR_1}] \times [2^{nR_2}]$ to each pair of sequences $(x_1^n, x_2^n) \in \mathcal{X}_1^n \times \mathcal{X}_2^n$, encoder 2 assigns index tuple $(m'_0, m'_1, m'_2)(x_1^n, x_2^n, x_1'^n, x_2'^n) \in [2^{nR'_0}] \times [2^{nR'_1}] \times [2^{nR'_2}]$ to each tuple of sequences $(x_1^n, x_2^n, x_1'^n, x_2'^n) \in \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{X}_1'^n \times \mathcal{X}_1'^n$, and decoder i assigns an estimate $\hat{x}_i^n(m_0, m_i)$ to each index pair $(m_0, m_i) \in [2^{nR_0}] \times [2^{nR_i}]$ and decoder i' assigns an estimate $\hat{x}_i'^n(m_0, m'_0, m'_i)$ to each index tuple $(m_0, m'_0, m'_i) \in [2^{nR_0}] \times [2^{nR'_0}] \times [2^{nR'_i}]$ for $i = 1, 2$. The probability of error is defined as $P_e^{(n)} = \mathsf{P}\{(\hat{X}_1^n, \hat{X}_2^n, \hat{X}_1'^n, \hat{X}_2'^n) \neq (X_1^n, X_2^n, X_1'^n, X_2'^n)\}$. A rate tuple $(R_0, R_1, R_2, R'_0, R'_1, R'_2)$ is said to be achievable if there exists a sequence of $(nR_0, nR_1, nR_2, nR'_0, nR'_1, nR'_2, n)$ codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$.

The standard Gray-Wyner network [11], which consists of encoder 1 and decoders 1, 2, is closely related to the standard (static) caching problem [4] in the sense that the common link (M_0) can be seen as the cache and private links can be seen as the updates for the corresponding requests. Similarly, the *successive* Gray-Wyner network captures the dynamic nature of the content. Consider $(X_1'^n, X_2'^n)$ as the modified versions of the original content files (X_1^n, X_2^n). From the caching perspective, M_0 is the cache content generated from the original data (X_1^n, X_2^n). If no modification occurs, user requests one of the files, X_i^n, and corresponding to this request M_i is transmitted as the update. If some modification occurs before the user request, on the other hand, a refinement M'_0 for the cache M_0 is delivered to the user. When the user requests one of the files, $X_i'^n$, corresponding to this request M'_i is transmitted as the update.

A. Optimal Rate Region

Define $\mathscr{R}$ as the set of achievable rate tuples $(R_0, R_1, R_2, R'_0, R'_1, R'_2)$ for the successive Gray-Wyner network. The following theorem presents a single letter characterization of the optimal rate region $\mathscr{R}$.

Theorem 1: The optimal rate region $\mathscr{R}$ consists of the rate tuples such that

$$R_0 \geq I(X_1, X_2; W), \tag{1a}$$
$$R_i \geq H(X_i|W), \quad i = 1, 2, \tag{1b}$$
$$R'_0 \geq I(X_1, X_2, X'_1, X'_2; V|W), \tag{1c}$$
$$R'_j \geq H(X'_j|W, V), \quad j = 1, 2, \tag{1d}$$

for some conditional pmfs $p(w|x_1, x_2), p(v|w, x_1, x_2, x'_1, x'_2)$ with $|\mathcal{W}| \leq |\mathcal{X}_1||\mathcal{X}_2| + 2$ and $|\mathcal{V}| \leq |\mathcal{W}||\mathcal{X}_1||\mathcal{X}_2||\mathcal{X}'_1||\mathcal{X}'_2| + 2$.

Proof sketch: For achievability, we use joint typicality encoding to find a $w^n(m_0), m_0 \in [2^{nR_0}]$ that is jointly typical with (x_1^n, x_2^n). The index m_0 is sent to all four decoders. Given $w^n(m_0)$, we assign indices $m_1 \in [2^{nR_1}]$ and $m_2 \in [2^{nR_2}]$ to the sequences in $\mathcal{T}_\epsilon^{(n)}(X_1|w^n(m_0))$ and $\mathcal{T}_\epsilon^{(n)}(X_2|w^n(m_0))$, respectively, and send them to decoders 1 and 2, respectively. Similarly, at encoder 2, given $w^n(m_0)$, we find a $v^n(m'_0), m'_0 \in [2^{nR'_0}]$, that is jointly typical with $(w^n, x_1^n, x_2^n, x_1'^n, x_2'^n)$. Given $(w^n(m_0), v^n(m'_0))$, we assign indices $m'_1 \in [2^{nR'_1}]$ and $m'_2 \in [2^{nR'_2}]$ to the sequences in $\mathcal{T}_\epsilon^{(n)}(X'_1|w^n(m_0), v^n(m'_0))$ and $\mathcal{T}_\epsilon^{(n)}(X'_2|w^n(m_0), v^n(m'_0))$, respectively, and send them to decoders 1' and 2', respectively. The converse follows from standard arguments with the auxiliary random variables $W_i := (M_0, X_1^{i-1}, X_2^{i-1})$ and $V_i := (M'_0, X_1'^{i-1}, X_2'^{i-1})$, $i \in [n]$ and from the fact that $(M_0, X_{li}, X_{2i}, X'_{1i}, X'_{2i}) \to (X_1^{i-1}, X_2^{i-1}) \to (X_1'^{i-1}, X_2'^{i-1})$ form a Markov chain. Finally, the cardinality bound on $\mathcal{W}$ and $\mathcal{V}$ can be shown using the convex cover method in [14]. ∎

Remark 1: The optimal rate region $\mathscr{R}$ for a pair of two files in Theorem 1 extends naturally to a pair of k-files (X^k, X'^k), for which the rate region $\mathscr{R}$ is the set of rate tuples such that

$$R_0 \geq I(X^k; W),$$
$$R_i \geq H(X_i|W), \quad i \in [k],$$
$$R'_0 \geq I(X^k, X'^k; V|W),$$
$$R'_j \geq H(X'_j|W, V), \quad j \in [k],$$

for some conditional pmfs $p(w|x^k)$ and $p(v|w, x^k, x'^k)$ with $|\mathcal{W}| \leq \prod_{i=1}^k |\mathcal{X}_i| + 2$ and $|\mathcal{V}| \leq |\mathcal{W}| \prod_{i=1}^k \prod_{j=1}^k |\mathcal{X}_i||\mathcal{X}'_j| + 2$. Using this general result, one could also extend the proceeding arguments to more than two files. We, however, maintain our discussion using two files for the ease of exposition.

B. On Lower Boundaries of the Optimal Rate Region

To better understand the boundaries of the optimal rate region $\mathscr{R}$, we now provide some lower bounds.

Corollary 1: If $(R_0, R_1, R_2, R'_0, R'_1, R'_2) \in \mathscr{R}$, then

$$R_0 + R_1 + R_2 \geq H(X_1, X_2), \tag{2}$$
$$R_0 + R'_0 + R'_1 + R'_2 \geq H(X'_1, X'_2). \tag{3}$$

To see this, fix two pmfs $p(w|x_1, x_2), p(v|w, x_1, x_2, x'_1, x'_2)$ (or equivalently, fix a pmf $p(w, v|x_1, x_2, x'_1, x'_2)$ such that $(X'_1, X'_2) \to (X_1, X_2) \to W$ form a Markov chain). Then,

$$R_0 + R_1 + R_2 \geq I(X_1, X_2; W) + H(X_1|W) + H(X_2|W)$$
$$= H(X_1, X_2) + I(X_1; X_2|W) \overset{(a)}{\geq} H(X_1, X_2),$$

where (a) holds with equality if and only if $X_1 \to W \to X_2$ form a Markov chain. For example, if we choose $p(w|x_1, x_2)$ as the achiever of the Wyner common information between X_1 and X_2, which is defined as [15]

$$C(X_1; X_2) := \min_{\substack{p(w|x_1, x_2) \\ I(X_1; X_2|W)=0}} I(X_1, X_2; W),$$

then (2) is achieved with equality. Similarly for (3), we have

$$R_0 + R_0' + R_1' + R_2'$$
$$\geq H(X_1', X_2') + I(X_1'; X_2'|W, V) + I(X_1, X_2; W, V|X_1', X_2')$$
$$\overset{(b)}{\geq} H(X_1', X_2'),$$

where (b) holds with equality if and only if $X_1' \to (W, V) \to X_2'$ and $(X_1, X_2) \to (X_1', X_2') \to (W, V)$ form Markov chains. For example, if we let $W = \emptyset$ and let $p(v|x_1', x_2')$ be the achiever of the Wyner common information between X_1' and X_2', then (3) is achieved with equality.

Intiutively, Corollary 1 expresses the fact that the communication system in Fig.2 cannot perform better than the optimistic case where decoders 1 and 2 cooperates as well as decoders 1' and 2'. A natural question then arises: are the lower bounds in (2) and (3) simultaneously achievable? The arguments in the proof of Corollary 1 implies that there exists a rate tuple $(R_0, R_1, R_2, R_0', R_1', R_2') \in \mathscr{R}$ simultaneously achieving the lower bounds in (2) and (3) if and only if there exists a pmf $p(w, v|x_1, x_2, x_1', x_2')$ such that

$$(X_1', X_2') \to (X_1, X_2) \to W, \quad (4a)$$
$$(X_1, X_2) \to (X_1', X_2') \to W, \quad (4b)$$
$$X_1 \to W \to X_2, \quad (4c)$$
$$(X_1, X_2) \to (X_1', X_2', W) \to V, \quad (4d)$$
$$X_1' \to (W, V) \to X_2' \quad (4e)$$

form Markov chains. We can simplify these constraints using the following lemma.

Lemma 1: For any given pmf $p(w|x_1, x_2)$, there exist a pmf $p(v|w, x_1, x_2, x_1', x_2')$ such that

$$I(X_1, X_2; V|X_1', X_2', W) = I(X_1'; X_2'|W, V) = 0.$$

It is easy to justify Lemma 1 by letting $V = (X_2', X_2')$ as one example among many. Consequently, the condition to achieve (2) and (3) with equality reduces to finding a pmf $p(w|x_1, x_2, x_1', x_2')$ such that (4a)-(4c) form Markov chains. A simple example can be constructed as follows.

Example 1: Suppose that X_1 and X_2 are independent, i.e., $p(x_1, x_2, x_1', x_2') = p(x_1)p(x_2)p(x_1', x_2'|x_1, x_2)$. Let $W = \emptyset$ and let $p(v|x_1', x_2')$ be the achiever of the Wyner common information between X_1' and X_2'. It is then easy to see that (4) form Markov chains and thus (2) and (3) can be achieved.

If X_1 and X_2 are correlated unlike Example 1, is it still possible to find such a W? The answer does in fact depend on the distribution of the whole content, $p(x_1, x_2, x_1', x_2')$. In the following, we present a necessary condition on the content distribution to achieve the lower bounds in Corollary 1.

Proposition 1: Suppose that (X_1, X_2) are not independent. Let G be a bipartite graph with vertex set $\mathcal{A} \cup \mathcal{B}$ where $\mathcal{A} = \mathcal{X}_1 \times \mathcal{X}_2$ and $\mathcal{B} = \mathcal{X}_1' \times \mathcal{X}_2'$ such that there is an edge between two vertices (t, t') if and only if $p(x_1, x_2, x_1', x_2') > 0$ where $t = (x_1, x_2)$ and $t' = (x_1', x_2')$. If G is connected, then the lower bounds in (2) and (3) cannot be achieved simultaneously.

Proof: We prove by contradiction. Suppose that there is a pmf $p(w|x_1, x_2, x_1', x_2')$ such that (4a)-(4c) form Markov chains and that the described bipartite graph G is connected. The Markov chains in (4a) and (4b) implies that for any $w \in \mathcal{W}$, $p(w|x_1, x_2, x_1', x_2')$ is a constant over each connected component of G. Since G is connected, then $p(w|x_1, x_2, x_1', x_2') = p(w)$, $\forall (x_1, x_2, x_1', x_2') \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_1' \times \mathcal{X}_2'$. Thus, W is independent of (X_1, X_2, X_1', X_2'), which contradicts with the Markov chain in (4c) since X_1 and X_2 are correlated. ∎

The necessary condition in Proposition 1 is in fact closely related to Gács-Korner common information [16], which is another well-known quantity proposed to measure the common information between two random variables and is defined as

$$K(X; Y) := \max_{\substack{p(w|x, y): \\ W \to X \to Y, \\ W \to Y \to X}} I(X, Y; W).$$

[17, Corollary 1] provides a way to compute $K(X; Y)$. Let G be a bipartite graph with vertex set $\mathcal{X} \cup \mathcal{Y}$ such that there is an edge between two vertices (x, y) if and only if $p(x, y) > 0$ and let W be the labels of the connected components of G. Then, $K(X; Y) = H(W)$. Therefore, $K(X; Y) = 0$ if and only if G is connected. Letting $X \leftarrow (X_1, X_2)$ and $Y \leftarrow (X_1', X_2')$ implies the following in our setting.

Corollary 2: For a correlated pair of (X_1, X_2), it is not possible to achieve (2) and (3) if $K(X_1, X_2; X_1', X_2') = 0$.

Inspired by Proposition 1, we construct an example where the information carried over the common link (M_0) is useful for all decoders, which allows the achievability of (2) and (3).

Example 2: Suppose the pmf $p(x_1, x_2, x_1', x_2')$ is given as

$(X_1, X_2)/(X_1', X_2')$	00	11	01	10
00	$\bar{\alpha}/2$	0	0	0
11	0	$\bar{\alpha}/2$	0	0
01	0	0	$\alpha/2 - \beta$	β
10	0	0	β	$\alpha/2 - \beta$

Note that both (X_1, X_2) and (X_1', X_2') are doubly symmetric binary source with parameter α. Let

$$W|x_1, x_2 = \begin{cases} \text{Bern}(q), & \text{if } (x_1, x_2) = (0, 0) \\ \text{Bern}(\bar{q}), & \text{if } (x_1, x_2) = (1, 1) \\ \text{Bern}(0.5), & \text{otherwise} \end{cases}, \quad (5)$$

where $q = 0.5 - 0.5\sqrt{1 - 2\alpha}/(1 - \alpha)$, which is the achiever of both of the Wyner common informations $C(X_1; X_2)$ and $C(X_1'; X_2')$ (refer to [10], [15] for the proof). Therefore, letting $V = \emptyset$ yield the achievability for the lower bounds in (2)-(3).

C. Relation to Conditional Wyner's Common Information

Given a fixed strategy for Encoder 1, what would be the optimal strategy for Encoder 2 to minimize the total rate of the descriptions transmitted from Encoder 2 to Decoders 1' and

2'? First, Theorem 1 implies that given a pmf $p(w|x_1, x_2)$, a rate tuple $(R_0, R_1, R_2, R'_0, R'_1, R'_2) \in \mathcal{R}$ must satisfy

$$R'_0 + R'_1 + R'_2 \geq H(X'_1, X'_2 | W).$$

Therefore, given a pmf $p(w|x_1, x_2)$, we define

$$R'^*_0(W) = \min_{\substack{(R_0, R_1, R_2, R'_0, R'_1, R'_2) \in \mathcal{R} \\ R'_0 + R'_1 + R'_2 = H(X'_1, X'_2 | W)}} R'_0.$$

As a corollary of Theorem 1, one can prove the following.

Corollary 3: Given a pmf $p(w|x_1, x_2)$, the minimum rate

$$R'^*_0(W) = C(X'_1; X'_2 | W),$$

where $C(X'_1; X'_2 | W)$ denotes the conditional Wyner's common information [18] and is defined by

$$C(X'_1; X'_2 | W) := \min_{\substack{p(v|x'_1, x'_2, w) \\ X'_1 - (W, V) - X'_2}} I(X'_1, X'_2; V | W).$$

[18] studied the Gray-Wyner network with side information W^n and showed that the minimum common rate to achieve the sum rate of $H(X'_1, X'_2 | W)$ is equal to $C(X'_1; X'_2 | W)$. Although W^n is not such a side information but related to (X'^n_1, X'^n_2) through (X^n_1, X^n_2) in our successive setting, we have the same result implying that W^n behaves like a side information for Encoder 2 and Decoders 1'-2'.

III. Dynamic Caching Problem

We now utilize the successive Gray-Wyner network discussed in the preceding section to formulate the dynamic caching problem illustrated in Fig 1. We concentrate on the simplest case where the database has a pair of files (X^n_1, X^n_2) drawn i.i.d. from the pmf $p(x_1, x_2)$ over a finite alphabet and one of the files is to be requested by the user randomly and equally likely. It is, however, easy to extend our arguments to more than two files. In the static caching setting [4], [12], contents of the files stays the same during all the three phases of caching, request and delivery. In the dynamic setting, however, we assume that the files (X^n_1, X^n_2) are modified to (X'^n_1, X'^n_2) with probability p before the request phase, where we assume that the tuple $(X^n_1, X^n_2, X'^n_1, X'^n_2)$ is distributed i.i.d. with respect to the pmf $p(x_1, x_2, x'_1, x'_2)$ over a finite alphabet. Given a fixed amount of data rate (or memory) for the total cache information, assuming that the encoder is given the probability distribution $p(x_1, x_2, x'_1, x'_2)$ and the modification probability at the beginning, what would be the optimal caching and delivery strategy to minimize the expected delivery rate? In this section, we address this optimization problem for modification probability $p = 0.5$.

Define the caching function

$$R(C) := \min_{\substack{p(w|x_1, x_2), \\ p(v|w, x_1, x_2, x'_1, x'_2): \\ I(X_1, X_2, X'_1, X'_2; V, W) \leq C}} \frac{1}{4}[H(X'_1 | W, V) + H(X'_2 | W, V)$$
$$+ H(X_1 | W) + H(X_2 | W)].$$

It is easy to see that the solution to this minimization problem occurs at the boundary of $I(X_1, X_2, X'_1, X'_2; V, W) = C$.

It is, however, not trivial to solve this optimization in general. Nonetheless, for some class of content distributions, $p(x_1, x_2, x'_1, x'_2)$, we can characterize the closed form solution.

A. Examples

We now present specific classes of file contents and discuss the corresponding optimization for the caching function.

Example 3 (Nested Content): Suppose that $H(X_1 | X_2) = H(X_2 | X'_1) = H(X'_1 | X'_2) = 0$. Define

$$\bar{R}(C) := \frac{1}{4}\left\{ \sum_{i=1}^{2}[H(X_i) - C]^+ + \sum_{j=1}^{2}[H(X'_j) - C]^+ \right\}, \quad (6)$$

where $[a]^+ := \max\{0, a\}$. Note that $\bar{R}(C)$ is in general a lower bound on $R(C)$. For this example, however, $R(C) = \bar{R}(C)$. To see this, first cache (X_1, X_2) up to the cache rate (via W). If $C > H(X_1, X_2) = H(X_2)$, cache (X'_1, X'_2) given X_2 (via V) until the total cache rate C is exhausted.

Example 4 (Partially Nested Content): Suppose that the pmf $p(x_1, x_2, x'_1, x'_2) = p(x_1, x_2)p(x'_1 | x_1)p(x'_2 | x_2)$ and $H(X_1, X_2 | X'_1, Y'_2) = 0$. If $C \geq C(X_1; X_2)$, then

$$R(C) = \frac{[H(X_1, X_2) - C]^+ + H(X'_1, X'_2) - C}{4}.$$

The converse follows from the general lower bound in Proposition 2, Section III-B. For the achievability, notice that if $X_1 \to W \to X_2$ form a Markov chain for a pmf $p(w|x_1, x_2)$, then $X'_1 \to W \to X'_2$ also form a Markov chain. Therefore, we choose $p(w|x_1, x_2)$ and $p(v|w, x_1, x_2, x'_1, x'_2)$ as follows. If $C \leq H(X_1, X_2)$, let $V = \emptyset$ and

$$W = \begin{cases} W^*, & \text{with probability } \beta \\ (X_1, X_2), & \text{with probability } \bar{\beta} \end{cases},$$

where W^* denotes the achiever of the Wyner's common information $C(X_1; X_2)$ and β is such that $I(X_1, X_2; W) = C(X_1; X_2)\beta + H(X_1, X_2)\bar{\beta} = C$. If $C \geq H(X_1, X_2)$, let $W = (X_1, X_2)$ and

$$V = \begin{cases} (X'_1, X'_2), & \text{with probability } \alpha \\ \emptyset, & \text{with probability } \bar{\alpha} \end{cases},$$

where α is such that $H(X_1, X_2) + \alpha H(X'_1, X'_2 | X_1, X_2) = C$. Fig. 3 shows the characterization of caching function $R(C)$ for $C \geq C(X_1; X_2)$. For $C < C(X_1; X_2)$, we provide a lower bound that follows from Proposition 2 and an upper bound that follows from time sharing between $(W, V) = (\emptyset, \emptyset)$ and $(W, V) = (W^*, \emptyset)$, where W^* denotes the achiever of the Wyner's common information $C(X_1; X_2)$.

In this example, as a result of the correlation between the original and modified files, total cache rate (or memory) is initially exhausted to cache (X_1, X_2) as if the encoder does not aware of a possible modification. If the cache rate is larger than $H(X_1, X_2)$, then additional information about (X'_1, X'_2) is cached when modification occurs (if any).

Example 5 (Pairwise Independent Content): Suppose that $p(x_1, x_2, x'_1, x'_2) = p(x_1, x_2)p(x'_1, x'_2)$. If $C \geq C(X_1; X_2) + C(X'_1; X'_2)$, then

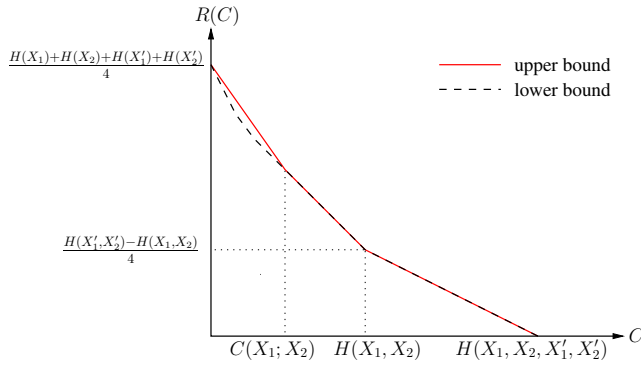$$R(C) = \frac{H(X_1, X_2) + H(X'_1, X'_2) - C}{4}.$$

Fig. 3: Bounds on caching function for Example 4

The converse follows from the general lower bound in Proposition 2, Section III-B. The achievability follows by time sharing between $(W, V) = ((X_1, X_2), (X_1', X_2'))$ and $(W, V) = (W^*, W'^*)$ where W^* and W'^* denote the achievers of $C(X_1; X_2)$ and $C(X_1'; X_2')$, respectively.

This example corresponds to caching two independent libraries using a shared cache memory that has to be split between them. As the achieavability implies, some of the cache rate/memory should be spared for the modified files.

B. Arbitrarily Correlated Contents

For arbitrarily correlated contents, we have the following lower bound on the caching function $R(C)$.

Proposition 2: For any $p(x_1, x_2, x_1', x_2')$,

$$R(C) \geq \max \left\{ \frac{\bar{R}(C)}{4}, \frac{H(X_1, X_2) + H(X_1', X_2') - 2C}{4}, \right.$$
$$\left. \frac{H(X_1, X_2, X_1', X_2') - C}{4} \right\}, \quad (7)$$

where $\bar{R}(C)$ is as defined in (6). Equality at $R(C) = 0.25[H(X_1, X_2, X_1', X_2') - C]$ holds if and only if

$$C \geq C^* := \min[I(X_1, X_2; W) + I(X_1', X_2'; V|W)],$$

where the minimum is over all pmfs $p(w, v|x_1, x_2, x_1', x_2')$ such that $(X_1', X_2') \to (X_1, X_2) \to W$, $(X_1, X_2) \to W \to (X_1', X_2')$, $X_1 \to W \to X_2$, $X_1' \to (W, V) \to X_2'$, and $(X_1, X_2) \to (X_1', X_2', W) \to V$ form Markov chains.
The proof follows by standard arguments.

Remark 2: Letting $W = (X_1, X_2)$, it is easy to see that

$$C^* \leq H(X_1, X_2) + C(X_1'; X_2'|X_1, X_2).$$

Therefore, if $C \geq H(X_1, X_2) + C(X_1'; X_2'|X_1, X_2)$, by Proposition 2, $R(C) = 0.25[H(X_1, X_2, X_1', X_2') - C]$. The achievability follows by $W = (X_1, X_2)$ and by time sharing between $V = V^*$ and $V = (X_1', X_2')$, where V^* denotes the achiever of $C(X_1'; X_2'|X_1, X_2)$. Intuitively, if the total cache rate (or memory) is large enough, caching the whole original contents is useful even if modification occurs because of the correlation between the original and modified files.

If the total cache rate (or memory) is limited, characterization of caching function is rather complicated. Instead, we next establish an upper bound on $R(C)$.

Lemma 2: Let W^* denote the achiever of Wyner's common information $C(X_1; X_2)$. For any $p(x_1, x_2, x_1', x_2')$, if $C \geq C(X_1; X_2) + C(X_1'; X_2'|W^*)$, then

$$R(C) \leq \frac{H(X_1, X_2) + H(X_1', X_2') - C - I(X_1', X_2'; W^*)}{4}.$$
$$(8)$$

Proof: It follows by time sharing between $W = W^*$ and $W = (X_1, X_2)$, and by setting V as the achiever of the conditional Wyner's common information $C(X_1'; X_2'|W)$. ∎

By Proposition 2 and Lemma 2, the gap Δ between the lower bound in (7) and the upper bound in (8) satisfies

$$\Delta \leq \min\{C, I(X, Y; X', Y')\}/4,$$

if $C(X; Y) + C(X'; Y'|W^*) \leq C \leq C^*$.

ACKNOWLEDGMENT

REFERENCES

[1] Y. Birk and T. Kol, "Coding on demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2825–2830, Jun. 2006.

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

[3] U. Niesen and M. A. Maddah-Ali, "Coded caching with nonuniform demands," in *IEEE Int. Conf. Comput. Commun. (INFOCOM)*, April 2014, pp. 221–226.

[4] C. Wang, S. H. Lim, and M. Gastpar, "Information-theoretic caching," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2015, pp. 1776–1780.

[5] P. Hassanzadeh, A. Tulino, J. Llorca, and E. Erkip, "Correlation-aware distributed caching and coded delivery," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2016, pp. 166–170.

[6] P. Hassanzadeh, E. Erkip, J. Llorca, and A. Tulino, "Distortion-memory tradeoffs in cache-aided wireless video delivery," in *Proc. 53th Ann. Allerton Conf. Comm. Control Comput.*, Sep. 2015, pp. 1150–1157.

[7] G. Op 't Veld and M. C. Gastpar, "Caching gaussians: Minimizing total correlation on the graywyner network," *Proceedings of the 50th Annual Conference on Information Systems and Sciences (CISS)*, p. 6, 2016.

[8] Q. Yang and D. Gndz, "Centralized coded caching for heterogeneous lossy requests," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2016, pp. 405–409.

[9] R. Timo, S. S. Bidokhti, M. Wigger, and B. C. Geiger, "A rate-distortion approach to caching," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1957–1976, March 2018.

[10] C.-Y. Wang, *Function Computation over Networks Efficient Information Processing for Cache and Sensor Applications*. Lausanne: EPFL, 2015.

[11] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 53, no. 9, pp. 1681–1721, 1974.

[12] S. H. Lim, C. Wang, and M. Gastpar, "Information-theoretic caching: The multi-user case," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7018–7037, Nov 2017.

[13] P. Hassanzadeh, A. M. Tulino, J. Llorca, and E. Erkip, "On coding for cache-aided delivery of dynamic correlated content," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 8, pp. 1666–1681, Aug 2018.

[14] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[15] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[16] P. Gács and J. Körner, "Common information is far less than mutual information," *Probl. Control Inf. Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[17] R. Ahlswede and J. Körner, *Appendix: On Common Information and Related Characteristics of Correlated Information Sources*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 664–677.

[18] A. Lapidoth and M. Wigger, "Conditional and relevant common information," in *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, Nov 2016, pp. 1–5.