# Coded Caching with Heterogeneous File Demand Sets — The Insufficiency of Selfish Coded Caching

Chih-Hua Chang and Chih-Chun Wang, Purdue ECE, USA

***Abstract*—This work falls under the broad setting of coded caching with user-dependent file popularity and average-rate capacity analysis. In general, the exact capacity characterization with user-dependent file popularity remains an open problem. For example, user 1 may be interested in files 1 and 2 with probabilities 0.6 and 0.4, respectively, while user 2 may be interested in only files 2, and 3 with probabilities 1/3 and 2/3, respectively, but not interested in file 1 at all. An optimal scheme needs to carefully balance the conflicting interests under the given probabilistic weights. Motivated by this fundamental but intrinsically difficult problem, this work studies the following simplified setting: Each user $k$ is associated with a *file demand set* (FDS) $\Theta_k$; each file in $\Theta_k$ is *equally desired* by user $k$ with probability $\frac{1}{|\Theta_k|}$; and files outside $\Theta_k$ is not desired at all. Different users may have different $\Theta_{k_1} \neq \Theta_{k_2}$, which reflects the user-dependent file popularity. Various capacity results have been derived (mostly for the cases of $K = 2$ users). One surprising byproduct is a proof showing that *selfish coded caching* is insufficient to achieve the capacity. That is, in an optimal coded caching scheme, a user sometimes has to cache the files of which he/she has zero interests.**

## I. Introduction

Coded caching [1] could significantly reduce the worst-case peak-hour transmission time when compared to the traditional uncoded caching solutions. Existing works have characterized the coded caching capacity for some special $N$ and $K$ values [1]–[4] and derived order-optimal capacity expression for general $N$ and $K$ [1], [5]–[7].

While the focus on the worst-case performance is analytically appealing, it is oblivious to the underlying probability distribution of the random requests, and hence may not be able to address the phenomenon that the worst-case situation may only happen infrequently. Recently, there are new results focusing on the *average-rate capacity* [7]–[13]. Under the assumption that all users having the same file popularity profile[1] [7]–[11] proposed new order-optimal coded caching schemes and showed that the traditional (uncoded) highest-popularity-first policy can be strictly suboptimal. One justification of this *user-independent file popularity* is that we can put users of the same preference into a single group and perform coded caching within this group. Various other achievable average rate results have been proposed in [12], [13] under different levels of file/cache size heterogeneity but without the order optimality guarantee. They all assume the same user-independent file popularity setting.

This work was supported in parts by NSF under Grant CCF-1422997, Grant ECCS-1407604, Grant CCF-1618475, and Grant CCF-1816013.

[1]The popularity could be uniform across all $N$ files [7] or vary significantly across all $N$ files [8]–[10].

Nonetheless, the above user-independent file popularity setting becomes less practical if each user has his/her own preference and the number of participating users is small, which is often the case in a traditional uncoded schemes that rely on highly individualized file prediction mechanisms. Motivated by the above observation, this work studies the average-rate capacity with user-dependent file popularity. The results would place coded caching on the same footing as the traditional uncoded solutions and allows for fair comparison between the two.

Specifically, we consider a coded caching system of $N$ files and $K$ users. Each user has his/her own file popularity of the $N$ files. To reduce the complexity and the need of explicitly specifying $K$ distinct probability distributions, one for each user, we assume that each user $k$ is associated with a file demand set (FDS) $\Theta_k$ and is interested in those files in $\Theta_k$ with probability $\frac{1}{|\Theta_k|}$, see Section II for details. The FDS setting reflects user-dependent file popularity by allowing different users having distinct $\Theta_{k_1} \neq \Theta_{k_2}$. We then derive the average-rate capacity results under a variety of this FDS setting.

We note that if user $k$ caches files from his/her own FDS $\Theta_k$, it *directly* reduces the transmission need during the delivery phase. On the other hand, if user $k$ caches files outside $\Theta_k$, the cached file is used to eliminate interference of other user(s), which *indirectly* reduces the transmission rate. An intuitive conjecture is that *since direct rate reduction is more beneficial than the indirect one and since each user $k$ is only interested in files in his/her own FDS $\Theta_k$, the best practice is for user $k$ to store only packets in $\Theta_k$ in order to "directly" benefit from the cached data.* A byproduct of ours is the first proof showing that this conjecture holds for some scenarios but is not true in general, i.e., *selfish coded caching* can be strictly suboptimal.

## II. Problem Formulation

We consider a coded caching system with one server and $K$ users. The server has access to $N$ files $W_1, \dots, W_N$, each having the same file size $F$ bits. Each $W_n$ is independently and uniformly randomly distributed over $\{0,1\}^F$. We use $Z_k$ to denote the cache content of user $k$, which is of size $M_k$ bits. Without loss of generality, we assume $M_k \in [0, NF]$. For any positive integer $x$, we define $[x] \triangleq \{1, \dots, x\}$.

The operation of the system consists of the *placement phase* and the *delivery phase*. In the placement phase, each user $k$ populates its cache content by

$$Z_k = \phi_k(W_1, \dots, W_N),\ \forall k \in [K] \tag{1}$$

where $\phi_k$ is the caching function of user $k$. In the delivery phase, each user $k$ sends a request $d_k \in [N]$ to the server, i.e., user $k$ demands file $W_{d_k}$. We denote the probability mass function (pmf) of the random request $d_k$ by $p^{[k]}_{d_k}$. The joint pmf of the demand pattern of $K$ users $\vec{d} \triangleq (d_1, \ldots, d_K) \in [N]^K$ is then $p_{\vec{d}} = p^{[1]}_{d_1} \cdots p^{[K]}_{d_K}$.

After receiving the demand index vector $\vec{d}$, the server broadcasts an encoded signal

$$X_{\vec{d}} = \psi(\vec{d}, W_1, \ldots, W_N) \quad (2)$$

of $R_{\vec{d}}$ bits with encoding function $\psi$ using an error-free link to all $K$ users. Each user $k$ then uses $X_{\vec{d}}$ as well as his/her cache content $Z_k$ to decode the requested file

$$\hat{W}_{d_k} = \mu_k(\vec{d}, X_{\vec{d}}, Z_k), \quad (3)$$

where $\mu_k$ is the decoding function of user $k$. A coded caching scheme is completely specified by $K$ caching functions $\{\phi_k\}$, one encoding function $\psi$, and $K$ decoding functions $\{\mu_k\}$.

**Definition 1.** *The* file demand set *(FDS) of user $k$ is defined as $\Theta_k \triangleq \{n \in [N] : p^{[k]}_n > 0\}$, which is the set of files that user $k$ desires with a strictly positive probability.*

**Definition 2.** *A coded caching scheme is* zero-error feasible *if $\hat{W}_{d_k} = W_{d_k}$ for all $k \in [K]$, all $d_k \in \Theta_k$ in the FDS, and all $(W_1, \cdots, W_N) \in \{0,1\}^{NF}$.*

Throughout this manuscript, we consider exclusively zero-error feasible schemes.

**Definition 3.** *A coded caching scheme is* selfish *if we replace all $K$ encoding functions $\phi_k$ in* (1) *by*

$$Z_k = \phi_k(\{W_n : n \in \Theta_k\}), \ \forall k \in [K]. \quad (4)$$

*Namely, each user $k$ only stores the files that he/she is interested, thus the name* selfish*. In contrast, the original, more general design using* (1) *is referred to as an* unselfish *scheme.*

**Definition 4.** *The worst-case rate of a coded caching scheme is defined as*

$$R^* = \max_{\forall \vec{d}: d_k \in \Theta_k} R_{\vec{d}}. \quad (5)$$

**Definition 5.** *The average-rate of a coded caching scheme is defined as*

$$\bar{R} = \sum_{\forall \vec{d}, d_k \in \Theta_k} p_{\vec{d}} R_{\vec{d}}. \quad (6)$$

*The uniform-average-rate of a scheme is defined as*

$$\tilde{R} = \frac{1}{\prod_{k=1}^{K} |\Theta_k|} \sum_{\forall \vec{d}, d_k \in \Theta_k} R_{\vec{d}}. \quad (7)$$

$\tilde{R}$ can be viewed as a first-order approximation of the average-rate $\bar{R}$ that replaces the joint distribution $p_{\vec{d}}$ with a uniform distribution over the FDS $\prod_{k=1}^{K} \Theta_k$ (rather the simplest, uniform distribution over $[N]^K$ [7].). In [14], an exact characterization of $\bar{R}$ has been provided for the 2-user/2-file setting, which involves detailed discussion of up to 28 different cases that depends on the underlying values of $(M_1, M_2)$ and $p_{\vec{d}}$. Instead of focusing on the exact $\bar{R}$, in this work we focus on the simplified, more tractable quantities $\tilde{R}$ and $R^*$ but allow the $N$ value to be $\geq 2$.

## III. Main Results

Sections III-A and III-B present several exact capacity results. Then we present in Section III-C converse and achievability results that do not yet have a matching counterpart (thus not necessarily tight).

### A. When Selfish and Unselfish Designs Are Equally Powerful

In this subsection, we outline several special cases for which selfish and unselfish designs are equally powerful.

**Proposition 1.** *If $\Theta_{k_1} \cap \Theta_{k_2} = \emptyset$ for all distinct $k_1, k_2 \in [K]$, then selfish and unselfish designs are equally powerful and achieve the same $R^*$ and $\tilde{R}$.*

This proposition shows if no two users are interested in a common file, each user can act as if he/she is the sole user in the system.

**Proposition 2** ($\Theta_1 = \Theta_2$)**.** *Consider $K = 2$ users, $N \geq 2$ files, and $\Theta_1 = \Theta_2 = [N]$. By definition, there is no difference between selfish and unselfish designs. Then the $\tilde{R}$ is tightly characterized*[2] *by*

$$\tilde{R} \geq F - (M_1/N) \quad \text{(Q1)}$$

$$\tilde{R} \geq F - (M_2/N) \quad \text{(Q2)}$$

$$\tilde{R} \geq \frac{2N-1}{N}F - \frac{2N-2}{N^2}M_1 - \frac{1}{N}M_2 \quad \text{(Q3)}$$

$$\tilde{R} \geq \frac{2N-1}{N}F - \frac{1}{N}M_1 - \frac{2N-2}{N^2}M_2 \quad \text{(Q4)}$$

This proposition is the average-rate counterpart of the worst-case setting in [3] for the $K = 2$ and arbitrary $N \geq 2$ case. The relationship of $\tilde{R}$ versus $(M_1, M_2)$ is illustrated in Fig. 1. The x-axis (resp. y-axis) is for the $M_1$ (resp. $M_2$) value. The inequalities (Q1) to (Q4) are marked in the corresponding regions. There are seven vertices $t_1$ to $t_7$ and each vertex is labeled by a tuple $(M_1, M_2, \tilde{R})$, where $(M_1, M_2)$ describe the location and the third coordinate describe the corresponding exact uniform-average-rate capacity $\tilde{R}$.

We then consider the simplest scenario when $\{\Theta_k\}$ are distinct and overlap with each other.

**Proposition 3** ($\Theta_1 = \{1\}, \Theta_2 = [N]$)**.** *Consider $K = 2$ users and $\Theta_1 = \{1\}$ and $\Theta_2 = [N]$. The selfish and unselfish designs have identical $\tilde{R}$, which is tightly characterized by:*

$$\tilde{R} \geq F - M_1 \quad \text{(Q5)}$$

$$\tilde{R} \geq F - (M_2/N) \quad \text{(Q6)}$$

$$\tilde{R} \geq \frac{2N-1}{N}F - \frac{N-1}{N}M_1 - \frac{1}{N}M_2 \quad \text{(Q7)}$$

[2]We use the statement *tightly characterized* when we can derive a matching pair of the converse and achievability results, i.e., it characterizs capacity.

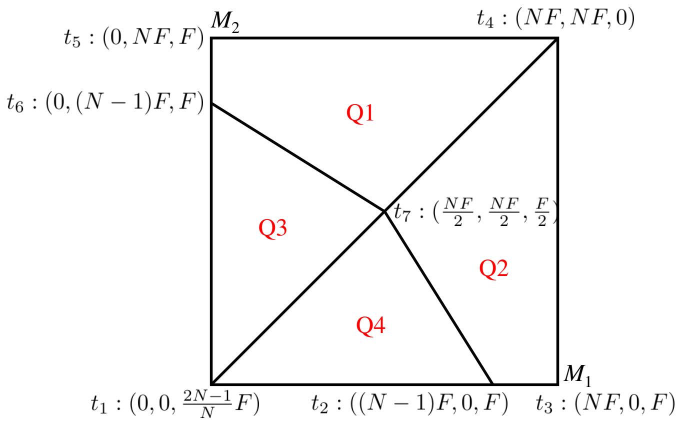


Fig. 1. The capacity $\tilde{R}$ of both the selfish and unselfish designs w. $\Theta_1 = \Theta_2 = [N]$.

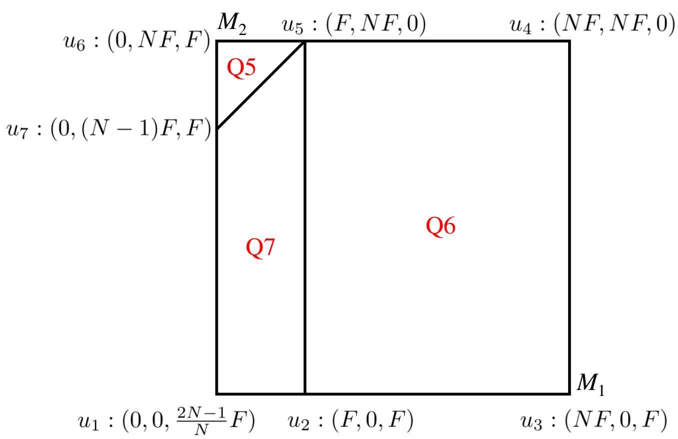


Fig. 2. The capacity $\tilde{R}$ of both the selfish and unselfish designs w. $\Theta_1 = \{1\}$ and $\Theta_2 = [N]$.

The relationship of $\tilde{R}$ versus $(M_1, M_2)$ is illustrated in Fig. 2. Comparing Propositions 2 and 3, it is clear that when the FDS $\Theta_1$ reduces from $[N]$ to $\{1\}$, the capacity $\tilde{R}$ reduces since an optimal scheme can now take advantage of the fact that user 1 is only interested in file 1.

### *B. Insufficiency of Selfish Designs*

**Proposition 4** (Unselfish w. $\Theta_1 = \{1, 2\}, \Theta_2 = \{1, 2, 3\}$)**.** *Consider $K = 2$ users and $\Theta_1 = \{1, 2\}$ and $\Theta_2 = \{1, 2, 3\}$. $\tilde{R}$ of the unselfish schemes is tightly characterized by*

$$\tilde{R} \geq F - M_1/2 \tag{P1}$$

$$\tilde{R} \geq F - M_2/3 \tag{P2}$$

$$\tilde{R} \geq \frac{5F}{4} - \frac{M_1}{4} - \frac{M_2}{4} \tag{P3}$$

$$\tilde{R} \geq \frac{3F}{2} - \frac{M_1}{3} - \frac{M_2}{3} \tag{P4}$$

$$\tilde{R} \geq \frac{5F}{3} - \frac{M_1}{2} - \frac{M_2}{3} \tag{P5}$$

$$\tilde{R} \geq \frac{5F}{3} - \frac{M_1}{3} - \frac{M_2}{2} \tag{P6}$$

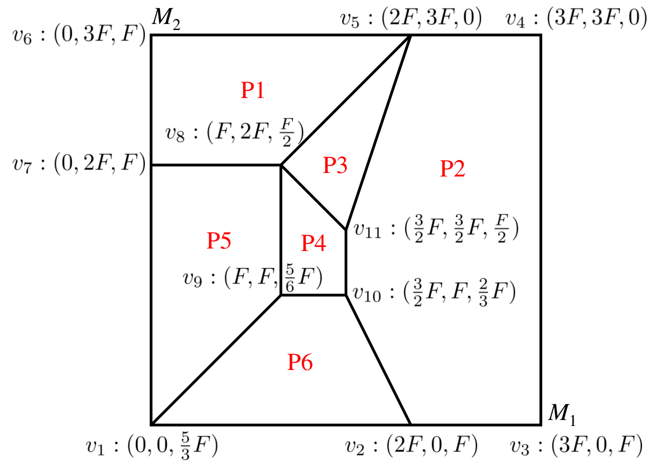


Fig. 3. The unselfish capacity $\tilde{R}$ w. $\Theta_1 = \{1, 2\}$ and $\Theta_2 = \{1, 2, 3\}$.

The relationship of the unselfish capacity $\tilde{R}$ versus $(M_1, M_2)$ is illustrated in Fig. 3.

**Proposition 5** (Selfish w. $\Theta_1 = \{1, 2\}, \Theta_2 = \{1, 2, 3\}$)**.** *Continue from Proposition 4. $\tilde{R}$ of the* selfish schemes *is tightly characterized by* (P1) *to* (P6) *plus an additional inequality:*

$$\tilde{R} \geq \frac{4F}{3} - \frac{M_1}{6} - \frac{M_2}{3}. \tag{P7}$$

The relationship of the selfish capacity $\tilde{R}$ versus $(M_1, M_2)$ is illustrated in Fig. 4. Note that one can prove that if $\tilde{R}$ satisfies inequality (P7), then it automatically satisfies (P3) and (P4). That is why in Fig. 4 there are only 5 subregions and the regions of (P3) and (P4) no longer appear.

When viewed separately, Propositions 4 and 5 describe the fundamental limits of unselfish and selfish coded caching when two users, with arbitrary cache sizes $(M_1, M_2)$, share concentrated[3], similar, but not identical interests, which alone are of important analytical value. Jointly, they provide the first proof that selfish coded caching is strictly suboptimal, e.g., the two points $v_{10}$ and $v_{11}$ in Fig. 3 can only be achieved by an unselfish design.

It is worth pointing out that the insufficiency of selfish coded caching is not due to the use of the average rate $\tilde{R}$ as the performance metric. Even when using the worst-case rate $R^*$ in (5), selfish designs are still insufficient.

**Corollary 1.** *Continue from Proposition 4. When $(M_1, M_2) = (1.5F, 1.5F)$, i.e., $v_{11}$ in Fig. 3, the worst-case capacity $R^*$ of the unselfish and selfish schemes are $0.5F$ and $\frac{7}{12}F$, respectively.*

By Propositions 4 and 5, we quickly see that the average-rate capacity $\tilde{R}$ of the unselfish and selfish schemes are $0.5F$ and $\frac{7}{12}F$, respectively. The rest of the proof is to construct schemes for which their $R^* = \tilde{R}$ match the average capacity.

[3]We say a user is of *concentrated interest* if the corresponding FDS $\Theta_k$ is small, e.g., $|\Theta_1| = 2$ and $|\Theta_2| = 3$ in Propositions 4 and 5. This is usually a result of highly effective next-file prediction.

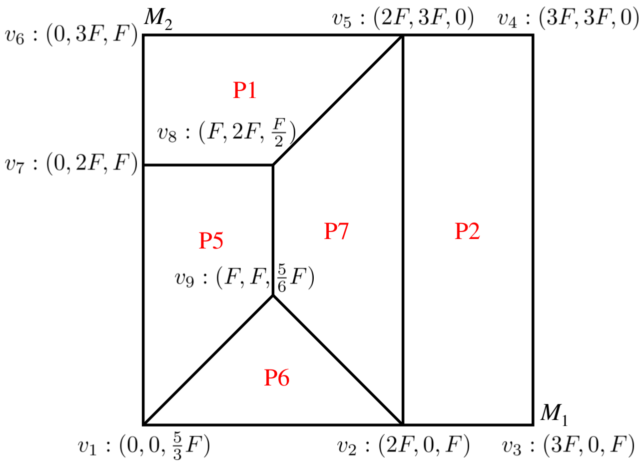


Fig. 4. The selfish capacity $\tilde{R}$ w. $\Theta_1=\{1,2\}$ and $\Theta_2=\{1,2,3\}$.

Corollary 1 shows that in this particular scenario, an optimal unselfish design further reduces the delivery rate by 14% when compared to an optimal selfish solution.

### *C. Other Miscellaneous Results*

For general $\Theta_k$, exact capacity characterization of $\tilde{R}$ remains an open problem. In the following, we provide some partial results that do not have matching converse and achievable rates.

**Proposition 6** (Converse w. $\Theta_1 \subsetneq \Theta_2$)**.** *Consider $K=2$ users and $\Theta_1=\{1,\cdots,N_1\}$ and $\Theta_2=\{1,\cdots,N_2\}$ satisfying*

$$3 \le 1.5N_1 \le N_2 \le 2N_1 \text{ and } N_1 \text{ and } N_2 \text{ being even.} \quad (8)$$

*The average-rate $\tilde{R}$ of an unselfish scheme must satisfy*

$$\tilde{R} \ge F-(M_1/N_1) \quad \text{(P1+)}$$
$$\tilde{R} \ge F-(M_2/N_2) \quad \text{(P2+)}$$
$$\tilde{R} \ge \frac{N_1+N_2}{2N_1}F-\frac{M_1+M_2}{2N_1} \quad \text{(P3+)}$$
$$\tilde{R} \ge \frac{3}{2}F-\frac{M_1+M_2}{N_2} \quad \text{(P4+)}$$
$$\tilde{R} \ge \frac{N_1+N_2}{N_2}F-\frac{M_1}{N_1}-\frac{M_2}{N_2} \quad \text{(P5+)}$$
$$\tilde{R} \ge \frac{N_1+N_2}{N_2}F-\frac{M_1}{N_2}-\frac{3M_2}{2N_2} \quad \text{(P6+)}$$

Fig. 5 illustrates the converse (P1+) to (P6+).

**Proposition 7** (Achievability w. $\Theta_1 \subsetneq \Theta_2$)**.** *Continue from Proposition 6. The lower bounds form 12 vertices in Fig. 5. Among them, 7 vertices $w_1$ to $w_7$ are achievable.*

Comparing Proposition 6 and 7, we have characterized the exact $\tilde{R}$ for the 3 subregions, each surrounded by the vertex sets $\{w_4,w_5,w_6\}$, $\{w_4,w_6,w_7\}$, and $\{w_1,w_2,w_3,w_4,w_7\}$, respectively. I.e., when either $M_1$ or $M_2$ is sufficiently large. Note that when $(N_1,N_2)=(2,3)$, then (P1+) to (P6+) collapse to (P1) to (P6) in Proposition 4, and they thus tightly characterize the capacity for all $(M_1,M_2)$.

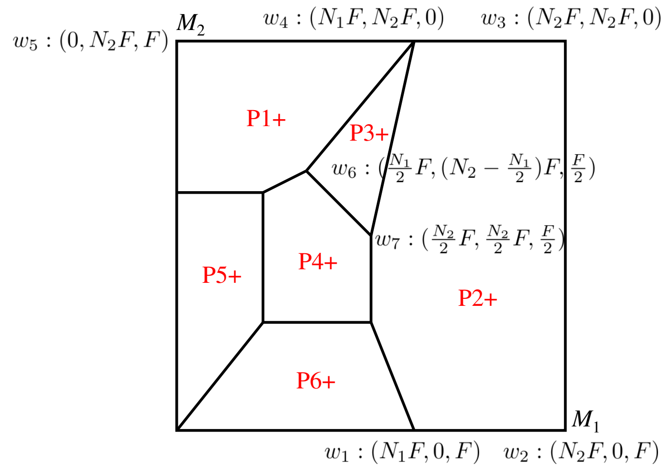


Fig. 5. The converse rate lower bounds of (P1+) to (P6+) w. $\Theta_1=\{1,\cdots,N_1\}$ and $\Theta_2=\{1,\cdots,N_2\}$ that satisfy (8). There are 12 vertices and we only label the 7 vertices for which we have a matching achievable rate.

## IV. Sketch Of The Proofs

Due to the space limits, we provide proofs only for some of the more difficult/interesting results while providing the details of all the proofs can be found in [15]. Specifically, we will prove inequalities (Q3) (and thus its symmetric version (Q4)), (P3), (P4), and (P7) and the achievability of vertex $v_{10}$.

*Proof of* (Q3)*:* By repeating the same arguments in [2], [3] but keeping track[4] of the request patterns $R_{(i,j)}$ being used, we can derive

$$NM_1+(2N-3)M_2+\sum_{i,j\in[N]:i\neq j} R_{(i,j)} \ge 2N(N-1)F \quad (9)$$

Then we notice that the following cut-set bound in [1] clearly holds

$$M_2+\sum_{i\in[N]} R_{(i,i)} \ge NF \quad (10)$$

Summing (9) and (10) and using $\tilde{R}=\frac{1}{N^2}\sum_{i,j\in[N]} R_{(i,j)}$, we have proven (Q3).

*Proof of* (P3)*:* For any $i,j\in[3]$ satisfying $\max(i,j)=3$ ($i$ and $j$ can both be 3) and any $k,l\in\{1,2\}$, we have

$$M_1+M_2+R_{(1,i)}+R_{(2,j)}+R_{(k,1)}+R_{(l,2)} \quad (11)$$
$$\ge H(X_{(1,i)},X_{(2,j)},Z_1)+H(X_{(k,1)},X_{(l,2)},Z_2) \quad (12)$$
$$=H(X_{(1,i)},X_{(2,j)},Z_1,W_1,W_2)$$
$$+H(X_{(k,1)},X_{(l,2)},Z_2,W_1,W_2) \quad (13)$$
$$\ge H(X_{(1,i)},X_{(2,j)},X_{(k,1)},X_{(l,2)},Z_2,W_1,W_2)$$
$$+H(W_1,W_2) \quad (14)$$
$$=H(X_{(1,i)},X_{(2,j)},X_{(k,1)},X_{(l,2)},Z_2,W_1,W_2,W_3)$$
$$+H(W_1,W_2) \quad (15)$$
$$=H(W_1,W_2,W_3)+H(W_1,W_2)=5F \quad (16)$$

[4]In [2], [3] each $R_{(i,j)}$ is replaced by $R^*$.

where (13) follows from that user 1 (resp. user 2) can decode $\{W_1, W_2\}$ based on $\{X_{(1,i)}, X_{(2,j)}, Z_1\}$ (resp. $\{X_{(k,1)}, X_{(l,2)}, Z_2\}$); (14) follows from the Shannon-type inequality; (15) follows from that user 2 can decode $W_3$ based on $\{X_{(1,i)}, X_{(2,j)}, Z_2\}$ since $\max(i, j) = 3$.

We then obtain three inequalities for $(i, j, k, l) = (3, 1, 1, 2)$, $(2, 3, 1, 2)$, and $(3, 3, 2, 1)$, respectively as follows.

$$M_1 + M_2 + R_{(1,3)} + R_{(2,1)} + R_{(1,1)} + R_{(2,2)} \geq 5F \quad (17)$$
$$M_1 + M_2 + R_{(1,2)} + R_{(2,3)} + R_{(1,1)} + R_{(2,2)} \geq 5F \quad (18)$$
$$M_1 + M_2 + R_{(1,3)} + R_{(2,3)} + R_{(2,1)} + R_{(1,2)} \geq 5F. \quad (19)$$

The summation of (17), (18), and (19) yields $3M_1 + 3M_2 + 12\tilde{R} \geq 15F$ or equivalently (P3).

*Proof of* (P4)*:* We first prove

$$M_1 + M_2 + R_{(2,3)} + R_{(1,2)} \quad (20)$$
$$\geq H(X_{(2,3)}, Z_1) + H(X_{(1,2)}, Z_2) \quad (21)$$
$$= H(X_{(2,3)}, Z_1, W_2) + H(X_{(1,2)}, Z_2, W_2) \quad (22)$$
$$\geq H(X_{(2,3)}, X_{(1,2)}, Z_1, Z_2, W_2) + H(W_2) \quad (23)$$
$$= H(X_{(2,3)}, X_{(1,2)}, Z_1, Z_2, W_1, W_2, W_3) + H(W_2) \quad (24)$$
$$= H(W_1, W_2, W_3) + H(W_2) = 4F. \quad (25)$$

where (22) to (24) follows similar arguments of (13) to (15), respectively. The summation of (17) and (25) yields $2M_1 + 2M_2 + 6\tilde{R} \geq 9F$ or equivalently (P4). Note that in the derivation of both (Q3) and (P4) we have to sum up two different types of inequalities to get the final results, i.e., (9) and (10) for (Q3), and (17) and (25) for (P4).

*Proof of* (P7)*:* Consider any $i \in \{1, 2\}$.

$$M_1 + M_2 + R_{(i,1)} + R_{(i,2)} + R_{(i,3)} \quad (26)$$
$$\geq H(X_{(i,3)}, Z_1) + H(X_{(i,1)}, X_{(i,2)}, Z_2) \quad (27)$$
$$= H(X_{(i,3)}, Z_1, W_i) + H(X_{(i,1)}, X_{(i,2)}, Z_2, W_1, W_2) \quad (28)$$
$$= H(X_{(i,3)}, Z_1, W_i) + H(X_{(i,1)}, X_{(i,2)}, Z_1, Z_2, W_1, W_2) \quad (29)$$
$$\geq H([X_{(i,j)}]_{j=1}^3, Z_1, Z_2, W_1, W_2) + H(Z_1, W_i) \quad (30)$$
$$= H([X_{(i,j)}]_{j=1}^3, Z_1, Z_2, W_1, W_2, W_3) + H(Z_1, W_i) \quad (31)$$
$$= H(W_1, W_2, W_3) + H(Z_1, W_i) = 3F + H(Z_1, W_i) \quad (32)$$

where (28) follows from that user 1 can decode $W_i$ based on $\{X_{(i,3)}, Z_1\}$ and user 2 can decode $\{W_1, W_2\}$ based on $\{X_{(i,1)}, X_{(i,2)}, Z_2\}$; (29) is one of the key steps and follows from the definition of selfish coded caching where $Z_1 = \phi_1(W_1, W_2)$; (30) follows from the assumption $i \in \{1, 2\}$ and the Shannon-type inequality; and (31) follows from user 2 can decode $W_3$ based on $X_{(i,3)}$ and $Z_2$.

Varying $i \in \{1, 2\}$, we have

$$M_1 + M_2 + \sum_{j=1}^{3} R_{(1,j)} \geq 3F + H(Z_1, W_1) \quad (33)$$

$$M_1 + M_2 + \sum_{j=1}^{3} R_{(2,j)} \geq 3F + H(Z_1, W_2). \quad (34)$$

The summation of (33) and (34) leads to

$$2M_1 + 2M_2 + 6\tilde{R} \geq 6F + H(Z_1, W_1) + H(Z_1, W_2) \quad (35)$$
$$\geq 6F + H(Z_1, W_1, W_2) + H(Z_1) = 8F + M_1 \quad (36)$$

and hence the bound (P7), where (36) follows from the Shannon-type inequality.

*Proof of* $v_{10}$*:* We first divide the three files $A$, $B$, and $C$ into two halves $(A_1, A_2)$, $(B_1, B_2)$, and $(C_1, C_2)$, respectively. In the placement phase, user 1 caches $Z_1 = (A_1, B_1, C_2 \oplus A_2 \oplus B_2)$ of total cache size $\frac{3F}{2}$, and user 2 caches $Z_2 = (A_2, B_2)$ of total cache size $F$. In the delivery phase, for the 6 possible demand patterns the server transmits $X_{(1,1)} = A_1 \oplus A_2$, $X_{(1,2)} = A_2 \oplus B_1$, $X_{(1,3)} = (C_1, C_2 \oplus B_2)$, $X_{(2,1)} = A_1 \oplus B_2$, $X_{(2,2)} = B_1 \oplus B_2$, and $X_{(2,3)} = (C_1, C_2 \oplus A_2)$ corresponding to the rates $R_{(1,1)} = R_{(1,2)} = R_{(2,1)} = R_{(2,2)} = \frac{F}{2}$, and $R_{(1,3)} = R_{(2,3)} = F$. The average rate is therefore $\tilde{R} = \frac{2}{3}F$. Note that a basic alignment technique is needed for the construction.

## V. Conclusion

We consider coded caching with heterogeneous file popularity. Each user $k$ only desires the files in his/her file demand set (FDS) $\Theta_k$ with equal probability $\frac{1}{|\Theta_k|}$ and we investigate the average rate capacity in various scenarios. A byproduct is the first proof showing that in some simple setting, unselfish designs can strictly outperform selfish ones by 14%.

## References


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

[2] C. Tian, "Symmetry, demand types and outer bounds in caching systems," in *Proc. IEEE Int. Symp. Inform. Theory*, Jul. 2016, pp. 825–829.

[3] D. Cao, D. Zhang, P. Chen, N. Liu, W. Kang, and D. Gunduz, "Coded caching with heterogeneous cache sizes and link qualities: The two-user case," in *Proc. IEEE Int. Symp. Inform. Theory*, 2018, pp. 1545–1549.

[4] C. Li, "On rate region of caching problems with non-uniform file and cache sizes," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 238–241, 2017.

[5] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," *IEEE/ACM Trans. Netw.*, vol. 23, no. 4, pp. 1029–1040, Aug 2015.

[6] J. Zhang, X. Lin, C. Wang, and X. Wang, "Coded caching for files with distinct file sizes," in *Proc. IEEE Int. Symp. Inform. Theory*, June 2015, pp. 1686–1690.

[7] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "Characterizing the rate-memory tradeoff in cache networks within a factor of 2," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 647–663, Jan 2019.

[8] U. Niesen and M. A. Maddah-Ali, "Coded caching with nonuniform demands," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 1146–1158, 2017.

[9] J. Zhang, X. Lin, and X. Wang, "Coded caching under arbitrary popularity distributions," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 349–366, Jan. 2018.

[10] M. Ji, A. M. Tulino, J. Llorca, and G. Caire, "Order-optimal rate of caching and coded multicasting with random demands," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3923–3949, Jun. 2017.

[11] J. Hachem, N. Karamchandani, and S. N. Diggavi, "Coded caching for multi-level popularity and access," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3108–3141, May 2017.

[12] P. Quinton, S. Sahraei, and M. Gastpar, "A novel centralized strategy for coded caching with non-uniform demands," *arXiv:1801.10563*, 2018.

[13] A. M. Daniel and W. Yu, "Optimization of heterogeneous coded caching," *arXiv:1708.04322*, Aug. 2017.

[14] https://engineering.purdue.edu/~chihw/pub_pdf/ISIT2019_micro_cap_proof.pdf.

[15] https://engineering.purdue.edu/~chihw/pub_pdf/ISIT2019_FDS_selfish_proof.pdf.