



HAL
open science

Coded Caching Does Not Generally Benefit From Selfish Caching

Federico Brunero, Petros Elia

► **To cite this version:**

Federico Brunero, Petros Elia. Coded Caching Does Not Generally Benefit From Selfish Caching. ISIT 2022, IEEE International Symposium on Information Theory, IEEE, Jun 2022, Espoo, Finland. pp.1139-1144, 10.1109/ISIT50566.2022.9834646 . hal-04094994

HAL Id: hal-04094994

<https://hal.science/hal-04094994>

Submitted on 11 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coded Caching Does Not Generally Benefit From Selfish Caching

Federico Brunero and Petros Elia

Communication Systems Department, EURECOM, Sophia Antipolis, France

Email: {brunero, elia}@eurecom.fr

Abstract—In typical coded caching scenarios, the content of a central library is assumed to be of interest to all receiving users. However, in a realistic scenario the users may have diverging interests which may intersect to various degrees. What happens for example if each file is of potential interest to, say, 40% of the users and each user has potential interest in 40% of the library? What if then each user caches selfishly only from content of potential interest? In this work, we formulate the symmetric selfish coded caching problem, where each user naturally makes requests from a subset of the library, which defines its own file demand set (FDS), and caches selfishly only contents from its own FDS. For the scenario where the different FDSs symmetrically overlap to some extent, we propose a novel information-theoretic converse that reveals, for such general setting of symmetric FDS structures, that selfish coded caching yields a load performance which is strictly worse — in the non-trivial memory regime — than that in standard coded caching.

Index Terms—Coded caching, file popularity, index coding, information-theoretic converse, selfish caching.

I. INTRODUCTION

The explosion of network traffic in the recent years has sparked much interest in new communication techniques with the purpose of reducing the traffic load. In this context, caching has always played a key role in bringing contents closer to their destinations, thus reducing the volume of the communication problem during peak hours. Coded caching [1] has been proposed as a clever way to better exploit caching capabilities of receiving users, where such coded technique represents, in comparison to traditional prefetching, a key breakthrough in the way end-user caches are employed to change both the *volume* of the communication problem and the *structure* of the problem itself. Current research on coded caching encompasses many topics such as the role of multiple antennas in caching [2]–[4], the interplay between caching and file popularity [5]–[7], and a variety of other scenarios [8]–[10]. For a thorough review of the existing coded caching works, we strongly encourage the reader to refer to the longer version of this work [11].

A. Heterogeneous User Profiles and Selfish Coded Caching

If on one hand a key ingredient in standard prefetching systems has commonly been the exploitation of the fact that some contents are more popular than others, and thus are generally to be allocated more cache space, on the other we

are just beginning to explore the interplay between coded caching, heterogeneous preferences and selfish caching, where by selfish caching we refer to caching policies where each user caches only contents from its own set of individual preferences.

Recent works have sought to explore this interplay. For example, in the context of coded caching with users having heterogeneous content preferences, the work in [12] analyzed the peak load of three different coded caching schemes that account for the user preferences, revealing occasional performance gains that are similarly dependent on the structure of these preferences. Related analysis appears in [13], now for the average load of the same schemes in [12]. On the other hand, the work in [14] focused on finding instances where unselfish coded caching outperforms selfish designs. This work nicely considered the performance of selfish coded caching in the context of heterogeneous file demand sets (FDSs), cleverly employing bounds to show that, for the case of $K = 2$ users and $N = 3$ files, unselfish designs strictly outperform selfish designs in terms of communication load, albeit only by a factor of up to 14%. In addition, the notable work in [15] established, under the assumption of selfish and uncoded prefetching, the optimal average load for the case of $K = 2$ users and a variety of overlaps between the two users' profiles, also providing explicit prefetching schemes of a selfish nature. In a similar vein, the recent work in [16] considered the scenario where users are interested in a limited set of contents depending on their location. To the best of our knowledge, the above constitutes the majority of works on selfish coded caching.

B. Main Contributions

First, we propose a novel symmetric FDS structure which aims to regulate the selfishness effect and to encapsulate the aspect of overlapping of interests by calibrating the degree of separation between the interests of the users. Secondly, we develop an information-theoretic converse on the optimal worst-case load under the assumption of uncoded and selfish placement. The interesting outcome of such bound is that selfish coded caching *can* be rather detrimental. In [11] we also provide instances where this bound is found to be tight.

C. Paper Outline

The system model is presented in Section II. Section III presents the information-theoretic converse, whose main proof in Section III-A is followed by the clarifying example in Section III-B. Section IV concludes the paper.

This work was supported by the European Research Council (ERC) through the EU Horizon 2020 Research and Innovation Program under Grant 725929 (Project DUALITY). An extended version of this work [11] has been submitted to IEEE Transactions on Information Theory.

D. Notation

We denote by \mathbb{Z}^+ the set of positive integers. For $n \in \mathbb{Z}^+$, we define $[n] := \{1, 2, \dots, n\}$. If $a, b \in \mathbb{Z}^+$ such that $a < b$, then $[a : b] := \{a, a + 1, \dots, b - 1, b\}$. For sets we use calligraphic symbols, whereas for vectors we use bold symbols. Given a finite set \mathcal{A} , we denote by $|\mathcal{A}|$ its cardinality. We denote by $\binom{n}{k}$ the binomial coefficient and we let $\binom{n}{k} = 0$ whenever $n < 0$, $k < 0$ or $n < k$. We use $m \bmod n$ to denote the modulo operation on m with integer divisor n , letting $m \bmod n = n$ when n divides m . For $n \in \mathbb{Z}^+$, we denote by S_n the group of all permutations of $[n]$ and by H_n the group of circular¹ permutations of $[n]$. To simplify, we denote by $\pi^n := (\pi(1), \dots, \pi(n))$ the vector of elements from $[n]$ permuted according to $\pi \in S_n$ and we let π^{-1} denote the inverse function of π .

II. SYSTEM MODEL

We consider the centralized caching scenario in [1] where one central server has access to a library \mathcal{L} containing N files of B bits each. This server is connected to K users through a shared error-free broadcast channel, and each user is equipped with a cache of size M files or, equivalently, MB bits.

The system works in two phases, i.e., the placement phase and the delivery phase. During the placement phase, the end-user caches are filled by the server according to a caching policy without any knowledge of future requests. During the delivery phase, after the demands of the users are simultaneously revealed, the server sends coded messages over the shared link to deliver the missing information to each user. Assuming that in the delivery phase each user demands simultaneously one file, the worst-case communication load R is defined as the total number of transmitted bits, normalized by the file-size B , that can guarantee delivery of all requested files in the worst-case scenario. The optimal worst-case communication load R^* is then formally defined as

$$R^*(M) := \inf\{R : (M, R) \text{ is achievable}\} \quad (1)$$

where the tuple (M, R) is said to be *achievable* if there exists a caching-and-delivery scheme which guarantees, for any possible demand, a load R for a given memory value M .

To capture the interplay between coded caching, heterogeneous interests and selfish caching, we propose an FDS structure that allows to calibrate the degree of separation between the interests of the different users.

Definition 1 (The Symmetric (K, α, F) FDS Structure). Let $\alpha \in [K]$ and $F \in \mathbb{Z}^+$. The symmetric (K, α, F) FDS structure assumes an N -file library $\mathcal{L} = \{\mathcal{W}_{\mathcal{S}} : \mathcal{S} \subseteq [K], |\mathcal{S}| = \alpha\}$ to be a collection of disjoint file classes $\mathcal{W}_{\mathcal{S}} = \{W_{f,\mathcal{S}} : f \in [F]\}$, where each class $\mathcal{W}_{\mathcal{S}}$ consists of F different files and each of such files is of interest to the α users in the set \mathcal{S} . As a consequence, the FDS of each user $k \in [K]$, which describes the files this user is potentially interested in, is defined as

$$\mathcal{F}_k = \{\mathcal{W}_{\mathcal{S}} : \mathcal{S} \subseteq [K], |\mathcal{S}| = \alpha, k \in \mathcal{S}\}. \quad (2)$$

¹This means that no vector $(\pi(1), \dots, \pi(n))$ with $\pi \in H_n$ can be obtained as a rotation of $(\sigma(1), \dots, \sigma(n))$ for some $\sigma \in H_n$ with $\sigma \neq \pi$.

We can see from the above that the library is partitioned into $C = \binom{K}{\alpha}$ disjoint classes of files, each composed of F files, for a total of $N = FC$ files. Each user $k \in [K]$ is then interested in its own FDS, which is made of $F \binom{K-1}{\alpha-1}$ files for each $k \in [K]$. Each file class is identified by an α -tuple \mathcal{S} , where \mathcal{S} simply tells which α users are interested in the file class $\mathcal{W}_{\mathcal{S}}$. We focus on the regime $N \geq K$, which is equivalent to $F \geq \lceil K/C \rceil$ for any fixed K and α . The following example can help familiarize the reader with the notation.

Example 1 (The Symmetric $(4, 3, 2)$ FDS Structure). Consider the symmetric $(K, \alpha, F) = (4, 3, 2)$ FDS structure. There are $C = \binom{4}{3} = 4$ classes² $\mathcal{W}_{123}, \mathcal{W}_{124}, \mathcal{W}_{134}, \mathcal{W}_{234}$ and $N = 8$ files: $W_{1,123}$ and $W_{2,123}$ from class \mathcal{W}_{123} , then $W_{1,124}$ and $W_{2,124}$ from class \mathcal{W}_{124} , and so on. The K FDSs take the form

$$\mathcal{F}_1 = \{\mathcal{W}_{123}, \mathcal{W}_{124}, \mathcal{W}_{134}\} \quad \mathcal{F}_3 = \{\mathcal{W}_{123}, \mathcal{W}_{134}, \mathcal{W}_{234}\} \quad (3)$$

$$\mathcal{F}_2 = \{\mathcal{W}_{123}, \mathcal{W}_{124}, \mathcal{W}_{234}\} \quad \mathcal{F}_4 = \{\mathcal{W}_{124}, \mathcal{W}_{134}, \mathcal{W}_{234}\} \quad (4)$$

where each FDS consists of 6 files in total. For \mathcal{F}_k denoting the FDS of user $k \in [4]$, user 1 is interested in files in $\mathcal{F}_1 = \{W_{f,\mathcal{S}} : \mathcal{S} \subseteq [4], |\mathcal{S}| = 3, 1 \in \mathcal{S}, f \in [2]\}$, user 2 is interested in files in $\mathcal{F}_2 = \{W_{f,\mathcal{S}} : \mathcal{S} \subseteq [4], |\mathcal{S}| = 3, 2 \in \mathcal{S}, f \in [2]\}$, and so on.

Deviating from standard notation practices, we will use the double-index notation W_{f_k, \mathcal{D}_k} to denote the file requested by user k . Hence, any demand will be described by the tuple (\mathbf{d}, \mathbf{f}) with $\mathbf{d} = (\mathcal{D}_1, \dots, \mathcal{D}_K)$ and $\mathbf{f} = (f_1, \dots, f_K)$. Trivially, it holds that $W_{f_k, \mathcal{D}_k} \in \mathcal{F}_k$ for each $k \in [K]$.

Our goal is to provide a converse bound on the optimal worst-case load under the assumption of uncoded and *selfish* placement, whose definition is given in the following.

Definition 2 (Uncoded and Selfish Cache Placement). A cache placement is *uncoded* if the bits of the files are simply copied within the caches of the users and is *selfish* when it guarantees for each $k \in [K]$ that a subfile of $W_{f,\mathcal{S}}$ can be cached at user k only if $W_{f,\mathcal{S}} \in \mathcal{F}_k$, i.e., only if the file $W_{f,\mathcal{S}}$ is potentially of interest to user k .

We denote by $R_{u,s}^*$ the optimal worst-case load under uncoded and selfish placement, and by R_{MAN} the load of the Maddah-Ali and Niesen (MAN) scheme in [1].

III. MAIN RESULTS

The converse bound employs the index coding techniques of [17] that proved the optimality of the MAN scheme under uncoded prefetching. Our main challenge will be to adapt the approach in [17] to reflect the system model in Section II. The result is stated in the following theorem, whereas the proof is presented in Section III-A.

Theorem 1. *Under the assumption of uncoded and selfish cache placement, and given the symmetric (K, α, F) FDS structure, the optimal worst-case communication load $R_{u,s}^*$ is*

²For simplicity, we will omit braces and commas when indicating sets, such that for example $\mathcal{W}_{\{1,2,3\}}$ may be written as \mathcal{W}_{123} .

lower bounded by R_{LB} which is a piecewise linear curve with corner points

$$(M, R_{\text{LB}}) = \left(t \frac{N}{K}, \frac{\binom{\alpha}{t+1} + (K - \alpha) \binom{\alpha-1}{t}}{\binom{\alpha}{t}} \right), \quad \forall t \in [0 : \alpha]. \quad (5)$$

The converse bound presented here shows that adding the selfish cache placement constraint implies a higher optimal communication load compared to the unselfish scenario. This is nicely reflected in the following³ corollary.

Corollary 1.1. *Given the symmetric (K, α, F) FDS structure and $\alpha \in [2 : K - 1]$, the converse reveals that*

$$\frac{R_{\text{u,s}}^*(t)}{R_{\text{MAN}}(t)} \geq 1, \quad \forall t \in [0 : \alpha - 1] \quad (6)$$

$$\frac{R_{\text{u,s}}^*(t)}{R_{\text{MAN}}(t)} > 1, \quad \forall t \in [\alpha - 2]. \quad (7)$$

For the (K, α, F) FDS structure and in the non-trivial range $t \in [0 : \alpha - 1]$, selfish coded caching is not better than unselfish coded caching, while optimal unselfish coded caching, in the non-extremal points of t and under uncoded placement, strictly outperforms any implementation of selfish coded caching.

Proof. Due to lack of space, the proof is relegated to the longer version of this work [11]. \square

Notice that Corollary 1.1 holds under the considered (K, α, F) FDS structure. Hence, we cannot state that selfish caching is detrimental for any conceivable FDS structure and for any memory point. However, in light of the results in Theorem 1 and Corollary 1.1, our claim is that selfish caching can be rather detrimental.

A. Main Proof

The derivation of the converse makes extensive use of the connection between caching and index coding identified in [1] and successfully exploited in [17]. Briefly, we recall that an index coding problem [18] consists of a server wishing to deliver N' independent messages to K' users via a basic bottleneck link. Each user $k \in [K']$ has its own *desired message set* $\mathcal{M}_k \subseteq [N']$ and has knowledge of its own *side information set* $\mathcal{A}_k \subseteq [N']$. The index coding problem can be described by the *side information graph*, a directed graph where each vertex is a message M_i for $i \in [N']$ and where there is an edge from M_i to M_j if M_i is in the side information set of the user requesting M_j . The derivation of our converse will make use of the following result from [19, Corollary 1].

Lemma 1 ([19, Corollary 1]). *In an index coding problem with N' messages M_i for $i \in [N']$, the minimum number of transmitted bits ρ is lower bounded as*

$$\rho \geq \sum_{i \in \mathcal{J}} |M_i| \quad (8)$$

³Notice that we do not consider $\alpha \in \{1, K\}$ in the corollary. Indeed, when $\alpha = 1$ the comparison is trivial, whereas for $\alpha = K$ the converse expression naturally matches that of unselfish coded caching.

for any acyclic subgraph \mathcal{J} of the side information graph.

We provide the proof for the non-trivial range⁴ $\alpha \in [2 : K - 1]$ and for the range $t \in [0 : \alpha]$. Indeed, for $t = \alpha$ we have $M = \alpha N / K = F \binom{K-1}{\alpha-1}$, so the point $(M, R) = (\alpha N / K, 0)$ is trivially achievable as a consequence of each user being able to store the entirety of its FDS.

Since each file is of interest to α users, the first step toward the converse consists of splitting each file in a generic manner into a maximum of 2^α disjoint subfiles as

$$W_{f,\mathcal{S}} = \{W_{f,\mathcal{S},\mathcal{T}} : \mathcal{T} \subseteq \mathcal{S}\} \quad (9)$$

for each $\mathcal{S} \subseteq [K]$ with $|\mathcal{S}| = \alpha$ and for each $f \in [F]$, where $W_{f,\mathcal{S},\mathcal{T}}$ is the subfile of $W_{f,\mathcal{S}}$ cached exactly and only by users in \mathcal{T} . Such generic splitting satisfies the uncoded and selfish placement constraint as defined in Definition 2.

1) *Constructing the Index Coding Bound:* Assuming that each user requests a distinct file, we can now identify the index coding problem with $K' = K$ users and $N' = K 2^{\alpha-1}$ messages, such that for any tuple (\mathbf{d}, \mathbf{f}) the desired message set and the side information set are respectively given by $\mathcal{M}_k = \{W_{f_k, \mathcal{D}_k, \mathcal{T}} : \mathcal{T} \subseteq \mathcal{D}_k, k \notin \mathcal{T}\}$ and $\mathcal{A}_k = \{W_{f, \mathcal{S}, \mathcal{T}} : f \in [F], \mathcal{S} \subseteq [K], |\mathcal{S}| = \alpha, \mathcal{T} \subseteq \mathcal{S}, k \in \mathcal{T}\}$ for each user $k \in [K]$. To apply Lemma 1, we are interested in acyclic sets of vertices \mathcal{J} in the side information graph of the problem. In the spirit of [17], we know that the set

$$\bigcup_{k \in [K]} \bigcup_{\mathcal{T} \subseteq ([K] \setminus \{u_1, \dots, u_k\}) \cap \mathcal{D}_{u_k}} \{W_{f_{u_k}, \mathcal{D}_{u_k}, \mathcal{T}}\} \quad (10)$$

does not contain any directed cycle for any demand (\mathbf{d}, \mathbf{f}) and any vector \mathbf{u} , where $\mathbf{u} = \boldsymbol{\pi}^K$ for some $\boldsymbol{\pi} \in S_K$. Consequently, applying Lemma 1 yields the following lower bound

$$BR_{\text{u,s}}^* \geq R(\mathbf{d}, \mathbf{f}, \mathbf{u}) \quad (11)$$

where $R(\mathbf{d}, \mathbf{f}, \mathbf{u})$ is defined as

$$R(\mathbf{d}, \mathbf{f}, \mathbf{u}) := \sum_{k \in [K]} \sum_{\mathcal{T} \subseteq ([K] \setminus \{u_1, \dots, u_k\}) \cap \mathcal{D}_{u_k}} |W_{f_{u_k}, \mathcal{D}_{u_k}, \mathcal{T}}|. \quad (12)$$

2) *Constructing the Optimization Problem:* Our goal is to create several bounds as the one in (11) and eventually average all of them to obtain a useful lower bound on the optimal worst-case load. We aim to create a bound for a set \mathcal{C} of properly selected demands and for a set $\mathcal{U}_{(\mathbf{d}, \mathbf{f})}$ of properly selected permutations for each demand $(\mathbf{d}, \mathbf{f}) \in \mathcal{C}$. Hence, we aim to simplify the expression given by

$$\sum_{(\mathbf{d}, \mathbf{f}) \in \mathcal{C}} \sum_{\mathbf{u} \in \mathcal{U}_{(\mathbf{d}, \mathbf{f})}} BR_{\text{u,s}}^* \geq \sum_{(\mathbf{d}, \mathbf{f}) \in \mathcal{C}} \sum_{\mathbf{u} \in \mathcal{U}_{(\mathbf{d}, \mathbf{f})}} R(\mathbf{d}, \mathbf{f}, \mathbf{u}). \quad (13)$$

Remark 1. The careful selection of the demand set \mathcal{C} and permutation set $\mathcal{U}_{(\mathbf{d}, \mathbf{f})}$ for each $(\mathbf{d}, \mathbf{f}) \in \mathcal{C}$ is a pivotal

⁴When $\alpha = 1$ the proof is trivial, since for such case we have only two integer points corresponding to $t \in \{0, 1\}$: when $t = 0$ the load is equal to K , and when $t = 1$ each user has enough memory to cache entirely its own FDS and the load is equal to 0. The case $\alpha = K$ corresponds to the standard MAN scenario already considered in [17].

difference with respect to the procedure in [17]. Indeed, if one develops a converse bound following the exact same procedure in [17], then the resulting bound is looser than the one presented in this paper. Finding the good set of demands and permutations becomes the key to obtaining a tighter bound while keeping the problem analytically tractable.

For a permutation $\pi \in H_K$, we consider the demands (\mathbf{d}, \mathbf{f}) where $\mathcal{D}_k = \{k, \pi((\pi^{-1}(k) + 1) \bmod K), \dots, \pi((\pi^{-1}(k) + \alpha - 1) \bmod K)\}$ and $f_k \in [F]$ for each $k \in [K]$. There is a total of F^K such demands. Considering that the order of H_K is $(K-1)!$ and that we take F^K demands for each $\pi \in H_K$, we consider $(K-1)!F^K$ distinct demands in total, denoting by \mathcal{C} the set of such demands. Since the vector \mathbf{d} depends on some circular permutation $\pi \in H_K$, we will identify from now on each demand with $(\mathbf{d}_\pi, \mathbf{f}) \in \mathcal{C}$ to highlight such dependency. For a demand $(\mathbf{d}_\pi, \mathbf{f}) \in \mathcal{C}$, we let $\mathcal{U}_{(\mathbf{d}_\pi, \mathbf{f})}$ be the set containing the K circular shifts of π^K .

Toward simplifying the expression in (13), we count how many times each subfile $W_{f, \mathcal{S}, \mathcal{T}}$ — for any $f \in [F]$, $\mathcal{S} \subseteq [K]$ with $|\mathcal{S}| = \alpha$, $\mathcal{T} \subseteq \mathcal{S}$ and $|\mathcal{T}| = t'$ with $t' \in [0 : \alpha]$ — appears in (13). To this end, we need the following lemma.

Lemma 2. *Let $\pi \in S_K$ be a permutation of the set $[K]$. Consider $k_1, k_2 \in [K]$ such that $k_1 \neq k_2$. Consider*

$$\ell = (\pi^{-1}(k_2) - \pi^{-1}(k_1)) \bmod K. \quad (14)$$

Then, out of the K circular shifts of π^K , there is a total of $(K - \ell)$ of them such that k_1 appears before k_2 .

Proof. Due to lack of space, the proof is relegated to the longer version of this work [11]. \square

The counting argument proceeds as follows. First, we focus on some subfile $W_{f, \mathcal{S}, \mathcal{T}}$ and we assume that the file $W_{f, \mathcal{S}}$ is requested by some user $k \in (\mathcal{S} \setminus \mathcal{T})$. Next, we count the number of demands $(\mathbf{d}_\pi, \mathbf{f}) \in \mathcal{C}$ for which k and right-most element included from \mathcal{T} are at distance ℓ in π^K . Then, we evaluate how many times the subfile appears in the index coding bounds associated to such demands. After that, noticing that $\ell \in [t' : \alpha - 1]$, we repeat the procedure for each value of ℓ . Finally, since the same procedure can be repeated for each $k \in (\mathcal{S} \setminus \mathcal{T})$, we multiply the end result by $|\mathcal{S} \setminus \mathcal{T}| = (\alpha - t')$.

Let us focus on the subfile $W_{f, \mathcal{S}, \mathcal{T}}$ for some $f \in [F]$, $\mathcal{S} \subseteq [K]$ with $|\mathcal{S}| = \alpha$, $\mathcal{T} \subseteq \mathcal{S}$ and $|\mathcal{T}| = t'$ for some $t' \in [0 : \alpha]$. For some user $k \in (\mathcal{S} \setminus \mathcal{T})$ and for $\ell \in [t' : \alpha - 1]$, there is a total of $a_\ell := t'!(\alpha - 1 - t')!(K - \alpha)! \binom{\ell - 1}{t' - 1} F^{K-1}$ demands $(\mathbf{d}_\pi, \mathbf{f}) \in \mathcal{C}$ such that the file $W_{f, \mathcal{S}}$ is requested by user $k \in (\mathcal{S} \setminus \mathcal{T})$ and there are exactly ℓ elements in π^K between k and the right-most element included from \mathcal{T} . Let $\mathcal{C}_{k, \ell}$ be the set of such demands. Considering how the acyclic set of vertices in (10) is built, the subfile $W_{f, \mathcal{S}, \mathcal{T}}$ appears in the index coding bound induced by each $(\mathbf{d}_\pi, \mathbf{f}) \in \mathcal{C}_{k, \ell}$ whenever all the elements in \mathcal{T} appear after k in $\mathbf{u} \in \mathcal{U}_{(\mathbf{d}_\pi, \mathbf{f})}$. Since for each demand $(\mathbf{d}_\pi, \mathbf{f}) \in \mathcal{C}_{k, \ell}$ there are exactly ℓ elements in

⁵Letting $\pi \in H_K$ be a circular permutation ensures that no \mathbf{d} vector is repeated for any specific \mathbf{f} vector.

π^K separating k and the right-most element from \mathcal{T} , we know from Lemma 2 that there are $(K - \ell)$ vectors $\mathbf{u} \in \mathcal{U}_{(\mathbf{d}_\pi, \mathbf{f})}$ where all the elements in \mathcal{T} appear after k . Observing that $\ell \in [t' : \alpha - 1]$ and that such reasoning applies whenever the file $W_{f, \mathcal{S}}$ is requested by any of the $(\alpha - t')$ users in $\mathcal{S} \setminus \mathcal{T}$, the specific subfile $W_{f, \mathcal{S}, \mathcal{T}}$ appears $(\alpha - t') \sum_{\ell=t'}^{\alpha-1} a_\ell (K - \ell)$ times in (13). Since the same reasoning applies to any other subfile, the expression in (13) can be rewritten as

$$R_{\mathbf{u}, \mathcal{S}}^* \geq \frac{1}{BK!F^K} \sum_{(\mathbf{d}_\pi, \mathbf{f}) \in \mathcal{C}} \sum_{\mathbf{u} \in \mathcal{U}_{(\mathbf{d}_\pi, \mathbf{f})}} R(\mathbf{d}_\pi, \mathbf{f}, \mathbf{u}) \quad (15)$$

$$= \sum_{t'=0}^{\alpha} f(t') x_{t'} \quad (16)$$

where $f(t')$ and $x_{t'}$ are defined as

$$f(t') := N \frac{(\alpha - t')}{F^K K!} \sum_{\ell=t'}^{\alpha-1} a_\ell (K - \ell) \quad (17)$$

$$0 \leq x_{t'} := \sum_{\mathcal{S} \subseteq [K]: |\mathcal{S}|=\alpha} \sum_{\mathcal{T} \subseteq \mathcal{S}: |\mathcal{T}|=t'} \sum_{f \in [F]} \frac{|W_{f, \mathcal{S}, \mathcal{T}}|}{NB}. \quad (18)$$

At this point, we seek to lower bound the minimum worst-case load $R_{\mathbf{u}, \mathcal{S}}^*$ by lower bounding the solution to the following optimization problem

$$\min_{\mathbf{x}} \sum_{t'=0}^{\alpha} f(t') x_{t'} \quad (19a)$$

$$\text{subject to } \sum_{t'=0}^{\alpha} x_{t'} = 1 \quad (19b)$$

$$\sum_{t'=0}^{\alpha} t' x_{t'} \leq \frac{KM}{N} \quad (19c)$$

where (19b) and (19c) correspond to the file-size constraint and the cumulative cache-size constraint, respectively.

3) *Bounding the Solution to the Optimization Problem:* The function $f(t')$ can be simplified as

$$f(t') = \frac{\binom{\alpha}{t'+1} + (K - \alpha) \binom{\alpha-1}{t'}}{\binom{\alpha}{t'}} \quad (20)$$

by means of the hockey-stick identity and after some algebraic manipulations. We encourage the reader to refer to [11] for the technical passages. Then, $x_{t'}$ can be considered as a probability mass function and so the optimization problem in (19) can be seen as the minimization of $\mathbb{E}[f(t')]$. Since we know from [11, Lemma 3] that $f(t')$ is convex and strictly decreasing for increasing t' , the sequence of inequalities $\mathbb{E}[f(t')] \geq f(\mathbb{E}[t']) \geq f(KM/N)$ holds by Jensen's inequality and the memory-size constraint in (19c). Finally, for $t := KM/N$ the converse bound is a piecewise linear curve with the corner points in (5), concluding the proof. \square

B. A Detailed Example for the Converse Bound

Consider the symmetric $(4, 3, 1)$ FDS structure which involves a file library $\mathcal{L} = \{W_{1, \mathcal{S}} : \mathcal{S} \subseteq [4], |\mathcal{S}| = 3\}$ consisting of $C = \binom{K}{\alpha} = \binom{4}{3} = 4$ classes of files. Since there is only

$F = 1$ file per class, there is a total of $N = FC = 4$ files, so in this example we make no distinction between files and classes of files. For simplicity, we will here refer to file $W_{1,S}$ directly as W_S , which means that each file is entirely described by a 3-tuple $\mathcal{S} \subseteq [4]$, and each demand instance is entirely defined by the $\mathbf{d} = (\mathcal{D}_1, \dots, \mathcal{D}_4)$ vector only. The FDS of each user $k \in [4]$ is given by $\mathcal{F}_k = \{W_S : \mathcal{S} \subseteq [4], |\mathcal{S}| = 3, k \in \mathcal{S}\}$ and it consists of $F \binom{K-1}{\alpha-1} = 3$ files. This example considers $t \in [0 : 3]$, simply because having $M = tN/K = 3$ implies that the point $(M, R) = (3, 0)$ is trivially achievable being each user able to preemptively cache the entirety of its FDS.

Assuming the most general uncoded and selfish cache placement, each file W_S is split into a total of $2^\alpha = 8$ disjoint subfiles as

$$W_S = \{W_{S,\mathcal{T}} : \mathcal{T} \subseteq \mathcal{S}\}, \quad \forall \mathcal{S} \subseteq [4] : |\mathcal{S}| = 3 \quad (21)$$

where again $W_{S,\mathcal{T}}$ is the subfile of W_S cached by users in \mathcal{T} .

1) *Constructing the Index Coding Bound:* For any given demand $\mathbf{d} = (\mathcal{D}_1, \dots, \mathcal{D}_4)$ where the k -th user asks for a distinct file $W_{\mathcal{D}_k}$ for each $k \in [4]$, we consider the index coding problem with $K' = K = 4$ users and $N' = K2^{\alpha-1} = 16$ messages, where each user $k \in [4]$ has a desired message set $\mathcal{M}_k = \{W_{\mathcal{D}_k,\mathcal{T}} : \mathcal{T} \subseteq \mathcal{D}_k, k \notin \mathcal{T}\}$ and a side information set $\mathcal{A}_k = \{W_{S,\mathcal{T}} : \mathcal{S} \subseteq [4], |\mathcal{S}| = 3, \mathcal{T} \subseteq \mathcal{S}, k \in \mathcal{T}\}$.

If we identify again in the side information graph of the problem the acyclic set of vertices as in (10), applying Lemma 1 yields the following lower bound

$$BR_{u,s}^* \geq R(\mathbf{d}, \mathbf{u}) \quad (22)$$

where now we define

$$R(\mathbf{d}, \mathbf{u}) := \sum_{k \in [4]} \sum_{\mathcal{T} \subseteq ([4] \setminus \{u_1, \dots, u_k\}) \cap \mathcal{D}_{u_k}} |W_{\mathcal{D}_{u_k}, \mathcal{T}}|. \quad (23)$$

Recall that (22) holds for any demand \mathbf{d} and any $\mathbf{u} = \boldsymbol{\pi}^4$ for some $\pi \in S_4$.

2) *Constructing the Optimization Problem and Bounding its Solution:* As described in Section III-A, we let \mathcal{C} be the set of distinct demands where we have a \mathbf{d}_π vector for each circular permutation $\pi \in H_4$. The number of such permutations is $(K-1)! = 6$. For simplicity, we denote by π_i the i -th circular permutation for $i \in [6]$, where such circular permutations are given by

$$\boldsymbol{\pi}_1^4 = (1, 2, 3, 4) \quad \boldsymbol{\pi}_4^4 = (1, 3, 4, 2) \quad (24)$$

$$\boldsymbol{\pi}_2^4 = (1, 2, 4, 3) \quad \boldsymbol{\pi}_5^4 = (1, 4, 2, 3) \quad (25)$$

$$\boldsymbol{\pi}_3^4 = (1, 3, 2, 4) \quad \boldsymbol{\pi}_6^4 = (1, 4, 3, 2). \quad (26)$$

None of the above permutations can be obtained as a rotation of any of the others. Hence, this ensures to have in \mathcal{C} the following distinct demand vectors

$$\mathbf{d}_{\pi_1} = (123, 234, 134, 124) \quad \mathbf{d}_{\pi_4} = (134, 123, 234, 124) \quad (27)$$

$$\mathbf{d}_{\pi_2} = (124, 234, 123, 134) \quad \mathbf{d}_{\pi_5} = (124, 123, 134, 234) \quad (28)$$

$$\mathbf{d}_{\pi_3} = (123, 124, 234, 134) \quad \mathbf{d}_{\pi_6} = (134, 124, 123, 234) \quad (29)$$

with $\mathcal{U}_{\mathbf{d}_{\pi_i}}$ containing the 4 circular shifts of $\boldsymbol{\pi}_i^4$. For instance, it is $\mathcal{U}_{\mathbf{d}_{\pi_1}} = \{(1, 2, 3, 4), (2, 3, 4, 1), (3, 4, 1, 2), (4, 1, 2, 3)\}$.

Now, we create a bound as in (22) for each \mathbf{d}_{π_i} with $i \in [6]$ and for each $\mathbf{u} \in \mathcal{U}_{\mathbf{d}_{\pi_i}}$ for a given \mathbf{d}_{π_i} , we sum all such bounds and we obtain the expression given by

$$\sum_{i \in [6]} \sum_{\mathbf{u} \in \mathcal{U}_{\mathbf{d}_{\pi_i}}} BR_{u,s}^* \geq \sum_{i \in [6]} \sum_{\mathbf{u} \in \mathcal{U}_{\mathbf{d}_{\pi_i}}} R(\mathbf{d}_{\pi_i}, \mathbf{u}). \quad (30)$$

We simplify (30) by counting how many times each subfile $W_{S,\mathcal{T}}$ — for any $\mathcal{S} \subseteq [4]$ with $|\mathcal{S}| = 3$, $\mathcal{T} \subseteq \mathcal{S}$ and $|\mathcal{T}| = t'$ with $t' \in [0 : 3]$ — appears in (30).

For example, let us focus on the subfile $W_{123,2}$. First of all, we notice that such subfile may appear to the RHS of (22) whenever W_{123} is requested by any user $k \in (\mathcal{S} \setminus \mathcal{T}) = \{1, 3\}$. Assume that W_{123} is requested by user 1, which means we consider the bounds with demands \mathbf{d}_{π_1} and \mathbf{d}_{π_3} . Denoting by ℓ the distance⁶ between user 1 and user 2 in $\boldsymbol{\pi}_i^4$ for $i \in \{1, 3\}$, we notice that $\ell \in \{1, 2\}$. We see that there is only $\boldsymbol{\pi}_1^4$ for $\ell = 1$, so $W_{123,2}$ appears $(K - \ell) = 3$ times in the bounds with demand \mathbf{d}_{π_1} and permutations in $\mathcal{U}_{\mathbf{d}_{\pi_1}}$. Similarly, there is only $\boldsymbol{\pi}_3^4$ for $\ell = 2$, so $W_{123,2}$ appears $(K - \ell) = 2$ times in the bounds with demand \mathbf{d}_{π_3} and permutations in $\mathcal{U}_{\mathbf{d}_{\pi_3}}$. Thus, summing over the possible values of ℓ , the subfile $W_{123,2}$ appears $3 + 2 = 5$ times in the bounds with demands \mathbf{d}_{π_1} and \mathbf{d}_{π_3} , and their relative permutations of users. The same rationale follows for the bounds built with \mathbf{d}_{π_2} , \mathbf{d}_{π_6} and their relative permutations, i.e., the bounds where file W_{123} is requested by user 3. Hence, the subfile $W_{123,2}$ appears in (30) a total of $2 \times (3 + 2) = 10$ times. The same holds for any other subfile cached at only one user and a similar counting argument can be made for arbitrary $|\mathcal{T}| = t'$ with $t' \in [0 : 3]$, as described in Section III-A. Thus, we can formulate an optimization problem as in (19), bounding its solution by means of Jensen's inequality and convexity of $f(t')$. The resulting bound is a piecewise linear curve with corner points $(M, R_{LB}) = (t, (3 - t)/(1 + t) + 1 - t/3)$ for each $t \in [0 : 3]$.

IV. CONCLUSION

In this work, we investigated the effects that selfish caching can have on the optimal worst-case communication load in the coded caching framework. We proposed for a general FDS structure, which seeks to capture the degree of intersection between the interests of the different users, a novel information-theoretic converse on the minimum worst-case communication load under uncoded and selfish prefetching. Such bound definitively resolves the question of whether selfish caching is *generally* beneficial or not. Indeed, since for the proposed symmetric FDS structure Corollary 1.1 reveals that any non-zero load brought about by symmetrically selfish caching is always (with the exception of the extreme points of t) strictly worse than the optimal load guaranteed in the unselfish scenario, we can claim that selfish coded caching *can* cause serious unbounded⁷ performance deterioration in caching systems.

⁶We remind that such distance represents the number of elements which separate index 1 and index 2, including the latter.

⁷This is proved in the longer version of this work [11].

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] E. Parrinello, A. Ünsal, and P. Elia, "Fundamental limits of coded caching with multiple antennas, shared caches and uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2252–2268, Apr. 2020.
- [3] E. Lampiris, A. Bazco-Nogueras, and P. Elia, "Resolving the feedback bottleneck of multi-antenna coded caching," *IEEE Trans. Inf. Theory*, 2021, early access. doi: 10.1109/TIT.2021.3139013.
- [4] A. Tölli, S. P. Shariatpanahi, J. Kaleva, and B. H. Khalaj, "Multi-antenna interference management for coded caching," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2091–2106, Mar. 2020.
- [5] 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.
- [6] Z. Zhang and M. Tao, "Deep learning for wireless coded caching with unknown and time-variant content popularity," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1152–1163, Feb. 2021.
- [7] J. Hachem, N. Karamchandani, and S. Diggavi, "Coded caching for multi-level popularity and access," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3108–3141, May 2017.
- [8] E. Lampiris and P. Elia, "Full coded caching gains for cache-less users," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7635–7651, Dec. 2020.
- [9] H. Joudeh, E. Lampiris, P. Elia, and G. Caire, "Fundamental limits of wireless caching under mixed cacheable and uncacheable traffic," *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4747–4767, Jul. 2021.
- [10] H. Zhao, A. Bazco-Nogueras, and P. Elia, "Vector coded caching multiplicatively boosts the throughput of realistic downlink systems," Feb. 2022, arXiv: 2202.07047 [cs.IT].
- [11] F. Brunero and P. Elia, "Unselfish coded caching can yield unbounded gains over symmetrically selfish caching," Sep. 2021, arXiv: 2109.04807 [cs.IT].
- [12] S. Wang and B. Peleato, "Coded caching with heterogeneous user profiles," in *2019 IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 2619–2623.
- [13] C. Zhang and B. Peleato, "On the average rate for coded caching with heterogeneous user profiles," in *ICC 2020 - 2020 IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
- [14] C.-H. Chang and C.-C. Wang, "Coded caching with heterogeneous file demand sets — the insufficiency of selfish coded caching," in *2019 IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 1–5.
- [15] C.-H. Chang, C.-C. Wang, and B. Peleato, "On coded caching for two users with overlapping demand sets," in *ICC 2020 - 2020 IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
- [16] K. Wan, M. Cheng, M. Kobayashi, and G. Caire, "On the optimal load-memory tradeoff of coded caching for location-based content," *IEEE Trans. Commun.*, Mar. 2022, early access. doi: 10.1109/TCOMM.2022.3158326.
- [17] K. Wan, D. Tuninetti, and P. Piantanida, "An index coding approach to caching with uncoded cache placement," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1318–1332, Mar. 2020.
- [18] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1479–1494, Mar. 2011.
- [19] F. Arbabjolfaei, B. Bandemer, Y.-H. Kim, E. Şaçoğlu, and L. Wang, "On the capacity region for index coding," in *2013 IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 962–966.