

Article

Optimal Linear Error Correcting Delivery Schemes for Two Optimal Coded Caching Schemes [†]

Nujoom Sageer Karat ¹, Anoop Thomas ² and Balaji Sundar Rajan ^{1,*} 

¹ Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India; nujoom@iisc.ac.in

² School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Odisha 752050, India; anoopthomas@iitbbs.ac.in

* Correspondence: bsrajan@iisc.ac.in; Tel.: +91-9845988753

[†] This paper is an extended version of our paper published in IEEE International Symposium on Information Theory (ISIT) 2018, Vail, Colorado, USA.

Received: 25 April 2020; Accepted: 10 July 2020; Published: 13 July 2020

Abstract: For coded caching problems with small buffer sizes and the number of users no less than the amount of files in the server, an optimal delivery scheme was proposed by Chen, Fan, and Letaief in 2016. This scheme is referred to as the CFL scheme. In this paper, an extension to the coded caching problem where the link between the server and the users is error prone, is considered. The closed form expressions for average rate and peak rate of error correcting delivery scheme are found for the CFL prefetching scheme using techniques from index coding. Using results from error correcting index coding, an optimal linear error correcting delivery scheme for caching problems employing the CFL prefetching is proposed. Another scheme that has lower sub-packetization requirement as compared to CFL scheme for the same cache memory size was considered by J. Gomez-Vilardebo in 2018. An optimal linear error correcting delivery scheme is also proposed for this scheme.

Keywords: coded caching; index coding; error correcting delivery schemes;

1. Introduction

The problem of coded caching introduced in [1] plays a crucial role in reducing peak hour traffic in networks. In a coded caching scheme, a part of the content is made available in local cache of users, so that traffic can be reduced at peak hours. Coded caching scheme involves two phases: a placement phase and a delivery phase. In the placement phase or the prefetching phase, which is performed during off-peak times, the entire database is made available to each user. Users fill their cache with the available data. Delivery phase is carried out once the demands are revealed by the users. During prefetching, some parts of files have to be judiciously cached at each user in such a way that the rate of transmission is reduced during the delivery phase. The prefetching can be done with or without coding. If during prefetching, no coding of parts of files is done, the prefetching scheme is referred to as uncoded prefetching [1,2]. If coding is done during prefetching stage, then the prefetching scheme is referred to as coded prefetching [3–7].

The seminal work in [1] shows that apart from the local caching gains obtained by placing contents at user caches before the demands are revealed, a global caching gain can be obtained by coded transmissions. This scheme is extended to decentralized scheme in [8]. The caching problem is extended to the case of non-uniform demands [9], online coded caching [10], hierarchical caching [11], and device-to-device caching [12]. Cache management scheme which incorporates decoding information in making cache replacement decisions was studied in [13]. A decentralized secure coded caching approach was proposed in [14], in which nodes only transmit coded files to avoid eavesdropper wiretapping and protect the user contents.

An error correcting delivery scheme is required if the shared bottleneck link between the server and the users is error-prone. The minimum average rate and minimum peak rate of error correcting delivery schemes is characterized in [15]. The placement phase is assumed to be error-free. This assumption can be justified as during the placement phase there is no bandwidth constraint and any number of re-transmissions can be done to make the placement error-free. A similar model in which the delivery phase takes place over a packet erasure broadcast channel was considered in [16–19].

In this paper, we consider the caching schemes considered in [3,4]. The prefetching scheme employed in these papers is coded, i.e., parts of files are coded and placed in the user caches. Optimal linear error correcting delivery schemes are proposed for these prefetching schemes. The optimal error correcting delivery scheme refers to an error correcting delivery scheme, which uses the minimum number of transmissions for a given placement. The main contributions of this paper are as follows.

- The prefetching scheme proposed in [3] is considered for the case where the number of users is the same as the number of files. The minimum number of transmissions required for correcting finite number of transmission errors is obtained for this case (Section 3.1).
- When the number of users is greater than the number of files, a different prefetching scheme is employed in [3]. For this prefetching strategy, the minimum number of transmissions required for correcting finite number of transmission errors is obtained (Section 3.2).
- A linear error correcting delivery scheme for coded caching problem with coded prefetching for small buffer sizes is proposed. The expressions for average rate and peak rate of this error correcting delivery scheme are found (Section 4).
- The caching scheme in [4] is considered for the same cache memory region. The advantage here is that the sub-packetization requirement is low. Sub-packetization level refers to number of subfiles that each file is split into. An optimal linear error correcting delivery scheme is found for the prefetching scheme employed in [4] (Section 5).

Parts of the content of this manuscript are present in [20]. The additional content in this manuscript is the construction of optimal linear error correcting delivery scheme for the prefetching scheme employed in [4] with reduced sub-packetization (Section 5). The proofs of Theorem 1 and Theorem 2 are also not present in [20].

A work that is closely related to this work is present in [15] and its extended version [21]. In [21], error correction is considered for a particular class of uncoded prefetching, namely symmetric batch prefetching. The major distinctions of this paper with the work in [21] are as follows. This is the first paper in the literature that considers error correction for coded caching schemes with coded prefetching. On the other hand, the work in [21] involves uncoded prefetching. For a fixed demand, the delivery scheme of a coded caching problem with uncoded prefetching is an index coding problem [22]. When the prefetching is coded, it corresponds to a generalized index coding problem [23]. To derive bounds in [21], the concepts of index coding are used and, to derive the bounds in this paper, we use techniques from generalized index coding. This paper is the first to use the concepts of generalized index coding to derive bounds for coded caching problems.

In this paper, $\mathbb{F}_q$ denotes the finite field with q elements, where q is prime power, and $\mathbb{F}_q^*$ denotes the set of all non-zero elements of $\mathbb{F}_q$. For any positive integer K , $[K]$ denotes the set $\{1, 2, \dots, K\}$. For a $K \times N$ matrix $\mathbf{L}$, $\mathbf{l}_i$ denotes its i th row. For vector spaces $\mathcal{U}, \mathcal{V}$, $\mathcal{U} < \mathcal{V}$ denotes that $\mathcal{U}$ is a subspace of $\mathcal{V}$. A linear $[n, k, d]_q$ code $\mathcal{C}$ over $\mathbb{F}_q$ is a k -dimensional subspace of $\mathbb{F}_q^n$ with minimum Hamming distance d . The vectors that belong to the subspace $\mathcal{C}$ are called codewords. A matrix $\mathbf{G}$ of size $k \times n$ whose rows are linearly independent codewords of $\mathcal{C}$ is called a generator matrix of $\mathcal{C}$. Thus, a linear $[n, k, d]_q$ code $\mathcal{C}$ can be represented using its generator matrix $\mathbf{G}$ as, $\mathcal{C} = \{\mathbf{y}\mathbf{G} : \mathbf{y} \in \mathbb{F}_q^k\}$. $N_q[k, d]$ denotes the length of the shortest linear code over $\mathbb{F}_q$, which has dimension k and minimum distance d .

2. Preliminaries and Background

Results from error correction for index coding with coded side information are used to obtain error correcting delivery schemes for the caching problem. In this section we recall results from error correction for index coding with coded side information introduced in [24]. We also review the coded caching scheme with coded prefetching proposed in [3,4], for which optimal error correcting delivery schemes are presented in the following sections (Section 3, Section 4, and Section 5).

2.1. Generalized Index Coding Problem and Error Correction

The index coding (IC) problem with side-information was introduced by Birk and Kol [22]. A source broadcasts messages through a noiseless shared channel to multiple receivers, each demanding certain messages and knowing some other messages a priori as side-information. The source needs to meet the demands of each receiver in a minimum number of transmissions. The minimum number of transmissions required to meet the demands of all the receivers is termed as the optimal length of the index coding problem. In [23] and [25], a generalization of the index coding problem was discussed, where the demands of the receivers and the side-information are linear combinations of the messages. In [25], the authors refer to this class of problems as Generalized Index Coding (GIC) problems.

An instance, $\mathcal{I}$ of a GIC problem is described formally, as follows. There is a message vector $\mathbf{x}^T = (x_1, x_2, \dots, x_n)^T \in \mathbb{F}_q^{n \times 1}$ and there are m receivers. The i th receiver demands a linear combination of the messages $\mathbf{r}_i \mathbf{x}^T$, for some $\mathbf{r}_i \in \mathbb{F}_q^{1 \times n}$, where $\mathbf{r}_i$ is the request vector and $\mathbf{r}_i \mathbf{x}^T$ is the request packet of the i th receiver. The side-information is represented by a matrix $\mathbf{V}^{(i)} \in \mathbb{F}_q^{s_i \times n}$, where s_i is the number of packets possessed as side-information by the i th receiver. The packets that are known as side-information possessed by the i th receiver are $\mathbf{V}^{(i)} \mathbf{x}^T$. Let $\mathcal{X}^{(i)}$ denote the row space of $\mathbf{V}^{(i)}$. Receiver i can also compute the linear combination of messages $\mathbf{v} \mathbf{x}^T \forall \mathbf{v} \in \mathcal{X}^{(i)}$. Let $\mathbf{R}$ be an $m \times n$ matrix over $\mathbb{F}_q$ having $\mathbf{r}_i$ as its i th row. The matrix $\mathbf{R}$ represents the demands of all the m receivers. In the definition of GIC problem in [25] the source is assumed to possess only certain linear combinations of messages. In our work, it is assumed that all of the messages are independent and the source possesses all of them.

The min-rank of an instance $\mathcal{I}$ of the GIC problem over $\mathbb{F}_q$ is defined as

$$\kappa(\mathcal{I}) = \min\{\text{rank}(\mathbf{A} + \mathbf{R}) : \mathbf{A} \in \mathbb{F}_q^{m \times n}, \mathbf{a}_i \in \mathcal{X}^{(i)}, i \in [m]\}.$$

This is motivated by the fact that any encoding matrix for a GIC problem is of the form $\mathbf{A} + \mathbf{R}$. It is shown in [24] that the min-rank is the optimal length of linear generalized index code. Intuitively, min-rank can be viewed as the minimum communication load among all of the delivery schemes with linear decoding functions, where each $\mathbf{a}_i$ corresponds to the coefficients for the locally cached contents.

For each $i \in [m]$, the set $\mathcal{Z}^{(i)}$ is defined as

$$\mathcal{Z}^{(i)} \triangleq \{\mathbf{z} \in \mathbb{F}_q^{n \times 1} : \mathbf{V}^{(i)} \mathbf{z} = 0, \mathbf{r}_i \mathbf{z} \neq 0\}. \quad (1)$$

Thus, $\mathcal{Z}^{(i)}$ is a collection of vectors that belong to the null space of $\mathbf{V}^{(i)}$ and satisfy $\mathbf{r}_i \mathbf{z} \neq 0$. The set of vector spaces $\mathcal{J}(\mathcal{I})$ is defined as $\mathcal{J}(\mathcal{I}) \triangleq \{\mathcal{U} \subset \mathbb{F}_q^n : \mathcal{U} \setminus \{0\} \subset \cup_{i \in [m]} \mathcal{Z}^{(i)}\}$. The set $\mathcal{J}(\mathcal{I})$ is the collection of subspaces such that the set of non-zero vectors of any subspace in the collection is a subset of $\cup_{i \in [m]} \mathcal{Z}^{(i)}$. The maximum value of dimension among the dimensions of all the elements of $\mathcal{J}(\mathcal{I})$ is called the generalized independence number, denoted by $\alpha(\mathcal{I})$. Thus, the dimension of any subspace of $\mathbb{F}_q^n$ in $\cup_{i \in [m]} \mathcal{Z}^{(i)} \cup \{0\}$ serves as a lower bound for $\alpha(\mathcal{I})$. The generalized independence number can be viewed as a cut set type of argument when the message lives in a subspace where all of the cached contents are zero. It was shown in [24] that the min-rank serves as an upper bound for the generalized independence number, i.e.,

$$\alpha(\mathcal{I}) \leq \kappa(\mathcal{I}). \quad (2)$$

Generalized index coding problems were classified in [26]. In generalized index coding with coded side-information problems, the demand of every receiver is uncoded but the side-information is coded. In generalized index coding problems with coded demands, the side-information of every receiver is uncoded but the demand is coded. In our work the focus is on generalized index coding with coded side-information problems.

Error correcting index codes were introduced in [27] and later extended for generalized index coding problems in [24]. An Error Correcting Generalized Index Code (ECGIC) is a map that encodes the message vector $\mathbf{x}^T$, such that each user, given its side-information and received transmissions, can decode its requested packet $\mathbf{r}_i \mathbf{x}^T \in \mathbb{F}_q$, even in the presence of at most δ transmission errors. The smallest possible length of such a δ -error correcting index code is denoted by $\mathcal{N}_q[\mathcal{I}, \delta]$. An optimal linear $(\mathcal{I}, \delta)$ -ECGIC over $\mathbb{F}_q$ is a linear $(\mathcal{I}, \delta)$ -ECGIC over $\mathbb{F}_q$ of the smallest possible length. The optimal length of a linear error correcting index code is lower bounded by α -bound and upper bounded by κ -bound as

$$\underbrace{\mathcal{N}_q[\alpha(\mathcal{I}), 2\delta + 1]}_{\alpha\text{-bound}} \leq \mathcal{N}_q[\mathcal{I}, \delta] \leq \underbrace{\mathcal{N}_q[\kappa(\mathcal{I}), 2\delta + 1]}_{\kappa\text{-bound}}. \quad (3)$$

where $\mathcal{N}_q[k, d]$ is the length of an optimal linear classical error-correcting code of dimension k and minimum distance d over $\mathbb{F}_q$ [24,27].

If for some index coding problem, $\alpha(\mathcal{I}) = \kappa(\mathcal{I})$, then bounds in (3) meet with equality. Thus, for such problems,

$$\mathcal{N}_q[\alpha(\mathcal{I}), 2\delta + 1] = \mathcal{N}_q[\mathcal{I}, \delta] = \mathcal{N}_q[\kappa(\mathcal{I}), 2\delta + 1].$$

In general, the κ -bound is obtained by concatenating an optimal linear classical error correcting code and an optimal linear index code. When $\alpha(\mathcal{I}) = \kappa(\mathcal{I})$, as the optimal length of the ECGIC is same as κ -bound, for such problems, concatenation scheme would give optimal linear error correcting index codes [26,28–30].

2.2. Error Correcting Coded Caching Scheme

The problem of error correcting coded caching scheme was proposed in [15]. The server is connected to K users through a shared link, which is error prone. The server has access to N files $X_1, X_2, \dots, X_N$ each of size F bits. Every user has an isolated cache with memory MF bits, where $M \in [0, N]$. There are many ways in which users can fill their cache contents. A prefetching scheme is denoted by $\mathcal{M}$ and it specifies the way in which user caches are filled. Each user demands one of the N files. Let the demand vector be $\mathbf{d} = (d_1, \dots, d_K)$, where d_i is the index of the file demanded by user i . The number of distinct files requested in $\mathbf{d}$ is denoted by $N_e(\mathbf{d})$. The set of all possible demands is denoted by $\mathcal{D} = \{1, 2, \dots, N\}^K$. During the delivery phase, the server informed of the demand $\mathbf{d}$, transmits a function of $X_1, X_2, \dots, X_N$ over a shared link. A δ -error correcting coded caching scheme should be in such a way that using the cache contents and the transmitted data, each user i needs to decode the requested file X_{d_i} , even if δ transmissions are in error.

For a δ -error correcting coded caching problem, a communication rate $R(\delta)$ is achievable for demand $\mathbf{d}$ if and only if there exists a transmission of $R(\delta)F$ bits such that every user i is able to decode its desired file X_{d_i} even after at most δ transmissions are in error. Rate $R^*(\mathbf{d}, \mathcal{M}, \delta)$ is the minimum achievable rate for a given $\mathbf{d}$, $\mathcal{M}$ and δ . The average rate $R^*(\mathcal{M}, \delta)$ is defined as the expected minimum average rate for a given $\mathcal{M}$ and δ under uniformly random demand. Thus, $R^*(\mathcal{M}, \delta) = \mathbb{E}_{\mathbf{d}}[R^*(\mathbf{d}, \mathcal{M}, \delta)]$. Another quantity of interest is the peak rate, denoted by $R_{\text{worst}}^*(\mathcal{M}, \delta)$, which is defined as $R_{\text{worst}}^*(\mathcal{M}, \delta) = \max_{\mathbf{d}} R^*(\mathbf{d}, \mathcal{M}, \delta)$.

2.3. Coded Caching Scheme with Coded Prefetching

An optimal coded caching scheme for small cache sizes involving coded prefetching was given in [3] by Chen, Fan, and Letaief. We refer this scheme as the Chen Fan Letaief (CFL) scheme.

The prefetching scheme here is denoted by $\mathcal{M}_{\text{CFL}}$. For this scheme, the cache memory size of each user is $M = 1/K$. The prefetching strategy differs depending on the values of N and K . Thus, $\mathcal{M}_{\text{CFL}}$ involves two types of prefetching strategies, which are as given below:

- Consider the case when $N = K$ and $M = 1/K$. Each file is split into N subfiles, i.e., $X_i = (X_{i,1}, X_{i,2}, \dots, X_{i,N})$. During prefetching, the cache of user j is filled as $Y_j = X_{1,j} \oplus X_{2,j} \oplus \dots \oplus X_{N,j}$, an XORed version of subfiles. It is shown in [3] that $R(\delta = 0) = N_e(\mathbf{d})$ for $N_e(\mathbf{d}) \leq N - 1$ and $R(\delta = 0) = N - 1$ for $N_e(\mathbf{d}) = N$ are achievable. Furthermore, if $M \in [0, 1/N]$, $R^*(M, \delta = 0) = N(1 - M)$ is achievable by memory sharing.
- Consider $K > N$ and $M = 1/K$. Each file is split into NK subfiles, i.e., $X_i = (X_{i,1}, X_{i,2}, \dots, X_{i,NK})$. The cache of user i is given by $Y_i = X_{1,N(i-1)+j} \oplus \dots \oplus X_{N,N(i-1)+j}$, for $j = 1, 2, \dots, N$. For the number of distinct demands $N_e(\mathbf{d}) \leq N - 1$ files, it is shown in [3] that $R(\delta = 0) = N_e(\mathbf{d})$ is achievable. For $N_e(\mathbf{d}) = N$, the rate $R(\delta = 0) = N - N/K$ is achievable. Furthermore, if $M \in [0, 1/K]$, $R^*(M, \delta = 0) = N(1 - M)$ is achievable by memory sharing.

2.4. Coded Caching Scheme with Coded Prefetching with Low Sub-Packetization

A scheme for $N \leq K$ and cache capacities $1/K \leq M \leq N/K$ was proposed in [4] by Jesus Vilardebo. We refer to this scheme as the Jesus Vilardebo (JV) scheme. The prefetching scheme of the JV scheme is denoted by $\mathcal{M}_{\text{JV}}$. We consider this scheme only for $M = 1/K$. The advantage of JV scheme over CFL scheme is that the sub-packetization requirement is low. The scheme in [4] uses a low sub-packetization level as compared to the scheme in [3]. A low sub-packetization level is always preferred, because any practical scheme will require each of the subfiles to have some header information that allows for decoding at the end users. When there are a large number of subfiles, the header overhead may be non-negligible. The prefetching scheme, $\mathcal{M}_{\text{JV}}$, is described, as follows: each file X_i is partitioned into K parts $X_{i,1}, X_{i,2}, \dots, X_{i,K}$ for $i \in [N]$. Afterwards, the cache content for user i is populated as $Y_i : X_{1,i} \oplus X_{2,i} \oplus \dots \oplus X_{N,i}$ for $i \in [K]$. The rate achieved is same as the CFL scheme.

2.5. Index Coding and Coded Caching

For a fixed prefetching $\mathcal{M}$ and for a fixed demand $\mathbf{d}$, the delivery phase of a coded caching problem is an index coding problem [1]. We denote such an index coding problem as $\mathcal{I}(\mathcal{M}, \mathbf{d})$. In fact, for fixed prefetching, a coded caching scheme consists of N^K parallel index coding problems one for each of the N^K possible user demands. Thus, finding the minimum achievable rate for a given demand $\mathbf{d}$ is equivalent to finding the min-rank of the equivalent index coding problem induced by the demand $\mathbf{d}$. Because the generalized independence number and min-rank of $\mathcal{I}(\mathcal{M}, \mathbf{d})$ depend on the caching scheme $\mathcal{M}$ and demand $\mathbf{d}$, we denote them as $\alpha(\mathcal{M}, \mathbf{d})$ and $\kappa(\mathcal{M}, \mathbf{d})$, respectively.

Consider the CFL prefetching scheme $\mathcal{M}_{\text{CFL}}$. The index coding problem that is induced by the demand $\mathbf{d}$ for the CFL prefetching is $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$. Each subfile $X_{i,j}$ corresponds to a message in the index coding problem. Because prefetching is coded, $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ represents a generalized index coding with coded side-information problem. Similarly, $\mathcal{I}(\mathcal{M}_{\text{JV}}, \mathbf{d})$ represents a generalized index coding problem that corresponds to the JV prefetching scheme.

There have been many works in the literature that make use of the link between index coding and coded caching. Some of them use index coding concepts to derive lower bounds on the rate [31,32]. In [33], multiple groupcast index coding is used to design coded caching delivery for multiple requests. A novel index coding scheme was introduced in [34], which, when applied to caching problem, is then shown to match an outer bound under the assumption of uncoded cache prefetching. A multihop index coding technique is proposed in [35] to code the cached contents in helpers to achieve order-optimal capacity gains. Decentralized caching schemes were proposed for two-layer networks in [36] which exploit index coding in the delivery phase and leverages multicast opportunities. None of these assume the channel to be error prone during the delivery phase. In our work, we use the error correction aspects of index coding to design optimal error correcting coded caching schemes.

We call a caching scheme optimal if both the placement and delivery schemes are designed in such a way to achieve an optimal rate memory pair. An optimal error correcting delivery scheme refers to an error correcting delivery scheme, which uses the minimum number of transmissions for a given placement. Unless the delivery scheme is carefully designed, one may end up with non-optimal error correcting delivery scheme for a given placement. This is clear from Example 4.8 of [27]. In that example, an index coding problem with five messages and five receivers are considered. The side information sets are given as $\mathcal{X}_1 = \{2, 5\}$, $\mathcal{X}_2 = \{1, 3\}$, $\mathcal{X}_3 = \{2, 4\}$, $\mathcal{X}_4 = \{3, 5\}$, and $\mathcal{X}_5 = \{1, 4\}$. For this problem, it can be calculated that $\alpha(\mathcal{I}) = 2$. Additionally, for this problem, $\min\text{-rank}$, $\kappa(\mathcal{I}) = 3$ over binary field. From code tables in [37], we have $N_2[2, 5] = 8$ and $N_2[3, 5] = 10$. Hence, $8 \leq N_2[\mathcal{I}, \delta] \leq 10$. Using a computer search, the authors of [27] have found that the optimal length $N_2[\mathcal{I}, 2] = 9$. Here the optimal length of the error correcting index code lies strictly between the α -bound and the κ -bound. Thus, for this problem, the construction of optimal linear error correcting index code by concatenation is not optimal. Similarly, in general for coded caching problems with an arbitrary placement and demand, the optimal error correcting delivery scheme cannot be constructed by concatenation unless we prove that the α and κ bounds meet for the corresponding index coding problem. In the schemes which we consider in this paper, we explicitly prove that the α and κ bounds meet for all of the demand cases for the given placement. Hence, the optimality of concatenation is guaranteed.

3. Generalized Independence Number for $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$

In this section we find a closed form expression for the generalized independence number $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ of the index coding problem $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$. There are two different prefetching schemes employed in [3], depending on the relationship between the number of messages and the number of receivers. For all the index coding problems corresponding to both these prefetching schemes, the generalized independence number is shown to be equal to the min-rank.

3.1. Number of Files Equal to the Number of Users ($N = K$)

In the CFL prefetching scheme, each file is split into N subfiles. Hence, the number of messages in $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ is N^2 . Each user is split into N receivers each demanding one message. Hence, there are a total of N^2 receivers. From the expressions of the achievable rates in [3], we get the min-rank $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ as

$$\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq \begin{cases} N_e(\mathbf{d})N & \text{if } N_e(\mathbf{d}) \leq N - 1 \\ N(N - 1) & \text{if } N_e(\mathbf{d}) = N \end{cases}. \quad (4)$$

We find the generalized independence number $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ for $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$. For different demands, a generalized independence number is calculated and it is shown to be equal to the min-rank of the corresponding generalized index coding problem. The technique of obtaining $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ is illustrated in the following example.

Example 1. Consider a coded caching problem with $N = K = 3$, $M = 1/3$ (see Figure 1). Because $M = 1/K$, the CFL scheme is used for solving the coded caching problem. Each file is split into $N = 3$ subfiles as $X_1 = (X_{1,1}, X_{1,2}, X_{1,3})$, $X_2 = (X_{2,1}, X_{2,2}, X_{2,3})$ and $X_3 = (X_{3,1}, X_{3,2}, X_{3,3})$. Let $\mathbf{X} = (X_{1,1}, X_{1,2}, X_{1,3}, X_{2,1}, X_{2,2}, X_{2,3}, X_{3,1}, X_{3,2}, X_{3,3})$ denote the vector obtained by concatenating X_1 , X_2 and X_3 . The cache contents of user i is $Y_i = (X_{1,i} \oplus X_{2,i} \oplus X_{3,i})$ for $i = 1, 2, 3$. Cache contents are depicted in Figure 1.

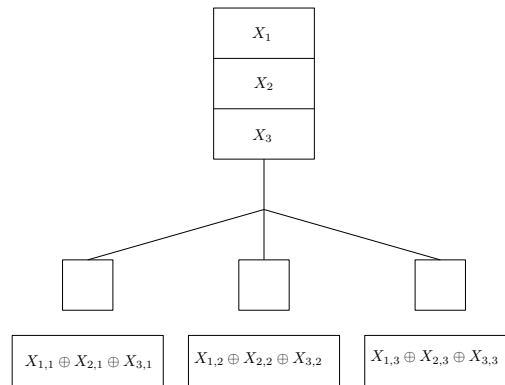


Figure 1. Coded Caching problem with $N = K = 3$, $M = 1/3$ and the hen Fan Letaief (CFL) placement.

First consider that all of the demands are distinct, i.e., $N_e(\mathbf{d}) = 3$. Without loss of generality, we can assume that the demand is $\mathbf{d} = (1, 2, 3)$. Consider the equations

$$\begin{aligned} e_1 : X_{1,1} \oplus X_{2,1} \oplus X_{3,1} &= 0, \\ e_2 : X_{1,2} \oplus X_{2,2} \oplus X_{3,2} &= 0, \text{ and} \\ e_3 : X_{1,3} \oplus X_{2,3} \oplus X_{3,3} &= 0. \end{aligned}$$

Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^9$, which consists of the vectors satisfying the equations e_1, e_2 and e_3 . From the rank-nullity theorem, $\dim(\mathcal{S}) \geq 6$. The induced generalized index coding problem $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ has 9 messages and 9 receivers. For this case, (1) can be rewritten as $\mathcal{Z}^{(i,j)} \triangleq \{\mathbf{z} \in \mathbb{F}_q^9 : e_i, X_{d_{i,j}} \neq 0\}$. Let $\mathcal{A}(\mathbf{d}) = \cup_{i,j \in [3]} \mathcal{Z}^{(i,j)} \cup \{\mathbf{0}\}$. The generalized independence number is the maximum among the dimensions of all the subspaces of $\mathbb{F}_q^9$ in $\mathcal{A}(\mathbf{d})$. We claim that all the vectors of $\mathcal{S}$ belong to the set $\mathcal{A}(\mathbf{d})$. Thus, $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq 6$. From the definition of $\mathcal{A}(\mathbf{d})$, it is clear that the all zero vectors $\mathbf{0}$ belonging to $\mathcal{S}$ also belong to $\mathcal{A}(\mathbf{d})$. Any other vector in $\mathcal{S}$ will have at least one non-zero coordinate $X_{i,j}$. The vector belonging to $\mathcal{S}$, having $X_{i,j} \neq 0$ belongs to the set $\mathcal{Z}^{(i,j)}$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$ and $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq 6$. From (4), we get $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq 6$. Hence, by (2), we have $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = \kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = 6$.

Finally, assume $N_e(\mathbf{d}) = 1$ and let $\mathbf{d} = (1, 1, 1)$. In addition to e_1, e_2 and e_3 , consider the following set of equations $e_4 : X_{2,1} = 0, e_5 : X_{2,2} = 0$, and $e_6 : X_{2,3} = 0$. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^9$, which consists of the vectors that satisfy the set of equations $e_1, e_2, \dots, e_6$. We follow the similar argument as above to show that all the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. By definition, $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. All of the vectors in $\mathcal{S}$ with $X_{1,j} \neq 0$ for $j = 1, 2, 3$ are present in $\mathcal{A}(\mathbf{d})$. By e_4, e_5 and e_6 , all the vectors in $\mathcal{S}$ have $X_{2,j} = 0$. The condition $X_{3,j} \neq 0$ and the set of equations $e_1, e_2, \dots, e_6$ force $X_{1,j} \neq 0$. Hence, all of the vectors in $\mathcal{S}$ with $X_{3,j} \neq 0$ are present in $\mathcal{Z}^{(1,j)}$. Thus, all of the vectors in $\mathcal{S}$ are present in $\mathcal{A}(\mathbf{d})$. Moreover, $\dim(\mathcal{S}) \geq 9 - 6 = 3$. Therefore $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq 3$. From (4), $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq 3$. Thus, by (2), $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = \kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = 3$. For other possible demands also, it can be verified from Table 1 that $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = \kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d})$.

Table 1. Generalized independence number and min-rank of $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ for different $\mathbf{d}$ in Example 1.

$\mathbf{d}$	$\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d})$	$\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d})$	$\mathbf{d}$	$\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d})$	$\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d})$	$\mathbf{d}$	$\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d})$	$\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d})$
(1, 2, 3)	6	6	(1, 3, 3)	6	6	(2, 2, 3)	6	6
(1, 3, 2)	6	6	(3, 1, 3)	6	6	(2, 3, 2)	6	6
(2, 1, 3)	6	6	(3, 3, 1)	6	6	(3, 2, 2)	6	6
(2, 3, 1)	6	6	(1, 1, 2)	6	6	(2, 3, 3)	6	6
(3, 1, 2)	6	6	(1, 2, 1)	6	6	(3, 2, 3)	6	6
(3, 2, 1)	6	6	(2, 1, 1)	6	6	(3, 3, 2)	6	6
(1, 2, 2)	6	6	(1, 1, 3)	6	6	(1, 1, 1)	3	3
(2, 1, 2)	6	6	(1, 3, 1)	6	6	(2, 2, 2)	3	3
(2, 2, 1)	6	6	(3, 1, 1)	6	6	(3, 3, 3)	3	3

In Example 1, the generalized independence number of the index coding problem $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ is equal to its min-rank. For different demands, the generalized index coding problem changes and, for all those problems, min-rank and generalized independence number are shown to be equal. This can be shown for all values of N , as given in the theorem below.

Theorem 1. For $N = K$ and $M = 1/K$,

$$\begin{aligned} \alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) &= \kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \\ &= \begin{cases} N_e(\mathbf{d})N & \text{if } N_e(\mathbf{d}) \leq N-1 \\ N(N-1) & \text{if } N_e(\mathbf{d}) = N \end{cases}, \end{aligned}$$

$\forall \mathbf{d} \in \mathcal{D}$, where $N_e(\mathbf{d})$ is the number of distinct demands.

Proof. In the CFL prefetching scheme $\mathcal{M}_{\text{CFL}}$, each file $X_i, i \in [N]$ is split into N subfiles $X_{i,1}, X_{i,2}, \dots, X_{i,N}$. User $i, i \in [N]$ caches $Y_i = (X_{1,i} \oplus X_{2,i} \oplus \dots \oplus X_{N,i})$. Let $\mathbf{X} = (X_{1,1}, \dots, X_{1,N}, X_{2,1}, \dots, X_{2,N}, \dots, X_{N,1}, \dots, X_{N,N})$ be the vector obtained by concatenation of vectors $X_i, i \in [N]$.

For a given demand $\mathbf{d}$, the delivery phase of the coded caching problem becomes a generalized index coding problem $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ with N^2 messages and N^2 receivers.

First, consider that all of the demands are distinct, i.e., $N_e(\mathbf{d}) = N$. Let the demand of the i th user be X_{d_i} . Thus $\mathbf{d} = (d_1, d_2, \dots, d_N)$. Consider the set of N equations denoted by $e_1, e_2, \dots, e_N$, where

$$e_i : (X_{1,i} \oplus X_{2,i} \oplus \dots \oplus X_{N,i}) = 0.$$

Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{N^2}$, which consists of the vectors that satisfy the set of equations $e_1, e_2, \dots, e_N$. From the rank-nullity theorem, we have $\dim(\mathcal{S}) \geq N^2 - N$.

For $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$, from (1) we have, $\mathcal{Z}^{(i,j)} \triangleq \{\mathbf{z} \in \mathbb{F}_q^{N^2} : e_i, X_{d_i,j} \neq 0\}$. Let $\mathcal{A}(\mathbf{d}) = \cup_{i,j \in [N]} \mathcal{Z}^{(i,j)} \cup \{\mathbf{0}\}$. The generalized independence number is the maximum of the dimensions of all subspaces of $\mathbb{F}_q^{N^2}$ in $\mathcal{A}(\mathbf{d})$. We show that $\mathcal{S}$ is such a subspace. For this, we need to show that all the vectors of $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. By the definition of $\mathcal{A}(\mathbf{d})$, the all zero vector $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. Any other vector in $\mathcal{S}$ will have at least one non-zero coordinate. The vectors that belong to $\mathcal{S}$ having $X_{d_i,j} \neq 0$ belongs to the set $\mathcal{Z}^{(i,j)}$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. The generalized independence number $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq N^2 - N$. From (4), we get $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq N^2 - N$. Hence, by (2), we have $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = \kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = N^2 - N$.

Consider the case where $N_e(\mathbf{d}) \leq N-1$. Without a loss of generality, we can assume that the first $N_e(\mathbf{d})$ users have distinct demands and that the i th user demands the file X_{d_i} for $i \in [N_e(\mathbf{d})]$. Also, without loss of generality, we can assume that the set of indices of the files that are not demanded are $N_e(\mathbf{d}) + 1, N_e(\mathbf{d}) + 2, \dots, N$. There are $U = N - N_e(\mathbf{d})$ files that are not demanded. In addition to $e_1, e_2, \dots, e_N$, consider the following set of equations $X_{N_e(\mathbf{d})+i,j} = 0$, for $i \in [U-1], j \in [N]$. The number of these equations is thus $N + N(U-1) = NU = N(N - N_e(\mathbf{d}))$. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{N^2}$, which consists of vectors that satisfy these equations. Hence, $\dim(\mathcal{S}) \geq N^2 - N(N - N_e(\mathbf{d})) = N_e(\mathbf{d})N$. By definition, $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. Any vector with the coordinate $X_{d_i,j} \neq 0$ for $i \in [N_e(\mathbf{d})]$ lies in $\mathcal{Z}^{(i,j)}$. The set of equations force all $X_{i,j} = 0$ for $i \in \{N_e(\mathbf{d}), \dots, N-1\}$. Moreover if $X_{N,j} \neq 0$ the set of equations force some $X_{d_i,j} \neq 0$ for some $i \in [N_e(\mathbf{d})]$. Hence any vector with $X_{N,j} \neq 0$ lies in some $\mathcal{Z}^{(i,j)}$ for $i \in [N_e(\mathbf{d})]$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. Therefore, $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq N_e(\mathbf{d})N$. Applying (4) and (2), $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = \kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = N_e(\mathbf{d})N$. $\square$

3.2. Number of Users more than the Number of Files ($N < K$)

In the CFL prefetching scheme for $N < K$, each file is split into NK subfiles. Hence, the number of messages in $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ is N^2K . Each user is split into NK receivers in $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$, each demanding

a single message. Thus, there are a total of NK^2 receivers. From the expressions for achievable rates in [3], we obtain the min-rank $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ as

$$\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq \begin{cases} NKN_e(\mathbf{d}) & \text{if } N_e(\mathbf{d}) \leq N - 1 \\ N^2(K - 1) & \text{if } N_e(\mathbf{d}) = N \end{cases}. \quad (5)$$

We find the generalized independence number $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ for $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$. The technique of obtaining $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ is illustrated in the following example.

Example 2. Consider a coded caching problem with $N = 3$, $K = 4$ and $M = 1/4$ (see Figure 2). According to the CFL scheme each file is split into $NK = 12$ subfiles as $X_1 = (X_{1,1}, X_{1,2}, \dots, X_{1,12})$, $X_2 = (X_{2,1}, X_{2,2}, \dots, X_{2,12})$, and $X_3 = (X_{3,1}, X_{3,2}, \dots, X_{3,12})$. Let $X = (X_{1,1}, \dots, X_{1,12}, \dots, X_{3,1}, \dots, X_{3,12})$ denote the vector obtained by concatenating X_1 , X_2 and X_3 . The cache of the i th user contains three coded packets $Y_i = (X_{1,3(i-1)+j} \oplus X_{2,3(i-1)+j} \oplus X_{3,3(i-1)+j})$ for $j = 1, 2, 3$. The cache contents are given in Figure 2.

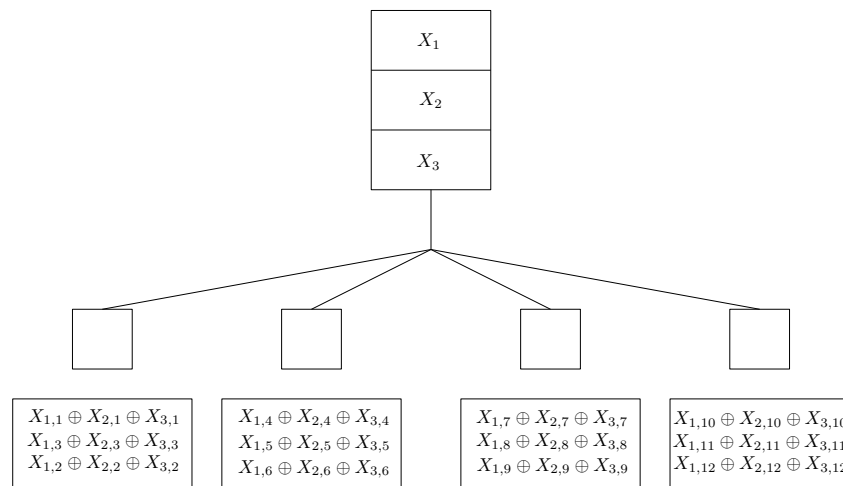


Figure 2. Coded Caching problem with $N = 3$, $K = 4$, $M = 1/4$ and CFL placement.

For a given demand $\mathbf{d}$, this problem becomes a generalized index coding problem $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$, having 36 messages and 48 receivers.

First consider that $N_e(\mathbf{d}) = N = 3$ and $\mathbf{d} = (1, 2, 3, 1)$. Consider the nine equations given by $e_{i,j} : (X_{1,3(i-1)+j} \oplus X_{2,3(i-1)+j} \oplus X_{3,3(i-1)+j}) = 0$ for $i = 1, 2, 3$ and $j = 1, 2, 3$. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{36}$ satisfying these nine equations. From the rank-nullity theorem, we get $\dim(\mathcal{S}) \geq 36 - 9 = 27$. For this case, (1) can be rewritten as $\mathcal{Z}^{(i,j)} \triangleq \{\mathbf{z} \in \mathbb{F}_q^{36} : e_{i,1}, e_{i,2}, e_{i,3}, X_{d,i,j} \neq 0\}$ for $i \in [4]$. Let $\mathcal{A}(\mathbf{d}) = \cup_{i \in [4], j \in [12]} \mathcal{Z}^{(i,j)} \cup \{\mathbf{0}\}$. The generalized independence number is the maximum among the dimensions of all the subspaces of $\mathbb{F}_q^{36}$ in $\mathcal{A}(\mathbf{d})$. We claim that $\mathcal{S}$ is such a subspace. This would mean that $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq 27$. For this, we need to show that all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. By the definition of $\mathcal{A}(\mathbf{d})$, the all zero vector $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. Any other vector in $\mathcal{S}$ will have at least one non-zero coordinate. All of the vectors in $\mathcal{S}$, having $X_{d,i,j} \neq 0$ belongs to $\mathcal{Z}^{(i,j)}$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$ and $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq 27$. From (5), we get $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq 3^2(4 - 1) = 27$. Hence, by (2), we have $\alpha(\mathcal{M}_{\text{CFL}}) = \kappa(\mathcal{M}_{\text{CFL}}) = 27$.

Consider now that $N_e(\mathbf{d}) = 2$ and $\mathbf{d} = (1, 2, 1, 2)$. In addition to the nine equations $e_{i,j}$ for $i = 1, 2, 3$ and $j = 1, 2, 3$, consider three more equations $e_{4,j} : (X_{1,9+j} \oplus X_{2,9+j} \oplus X_{3,9+j}) = 0$ for $j = 1, 2, 3$. Thus, we consider a set of twelve equations given by $\mathcal{E} = \{e_{i,j} : i \in [4], j \in [3]\}$. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{36}$ consisting of vectors that satisfy the equations in $\mathcal{E}$. Hence, from the rank-nullity theorem, we have $\dim(\mathcal{S}) \geq 36 - 12 = 24$. By definition, $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. Any non-zero vector in $\mathcal{S}$ with $X_{d,i,j} \neq 0$ for $i = 1, 2$ lies in the corresponding $\mathcal{Z}^{(i,j)}$. By $\mathcal{E}$, any $X_{3,j} \neq 0$ forces some $X_{i,j} \neq 0$ for $i = 1, 2$ and, hence, such vectors also lie in $\mathcal{A}(\mathbf{d})$. Thus all of the

vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. Therefore $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq 24$. From (5), we get $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq 12(2) = 24$. Hence, by (2), we have $\alpha(\mathcal{M}_{\text{CFL}}) = \kappa(\mathcal{M}_{\text{CFL}}) = 24$.

Finally, consider $N_e(\mathbf{d}) = 1$ and $\mathbf{d} = (1, 1, 1, 1)$. The files X_2 and X_3 are not demanded by any user. In addition to the equations in $\mathcal{E}$, here we consider a set of equations $X_{2,j} = 0$ for $j \in [12]$. Thus there are 24 equations in total. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{36}$ which satisfy these equations. By the rank-nullity theorem, the dimension of $\mathcal{S}$ is given by $\dim(\mathcal{S}) \geq 36 - 24 = 12$. The next step is to show that all the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. The all zero vector $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$ by definition. Any non-zero vector in $\mathcal{S}$ with $X_{1,j} \neq 0$ for $j \in [12]$ lies in the corresponding $\mathcal{Z}^{(i,j)}$. The 24 equations considered involve equations of the form $X_{2,j} = 0$ for $j \in [12]$. Hence, by E, any $X_{3,j} \neq 0$ forces $X_{1,j} \neq 0$ for $j \in [12]$ and hence such vectors also lie in $\mathcal{A}(\mathbf{d})$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. Therefore, $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq 12$. From (5), we get $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq 12(1) = 12$. Hence, by (2), we have $\alpha(\mathcal{M}_{\text{CFL}}) = \kappa(\mathcal{M}_{\text{CFL}}) = 12$.

The theorem below gives the general expression for $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d})$, when $N < K$.

Theorem 2. For $N < K$ and $M = 1/K$,

$$\begin{aligned} \alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) &= \kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \\ &= \begin{cases} NKN_e(\mathbf{d}) & \text{if } N_e(\mathbf{d}) \leq N - 1 \\ N^2(K - 1) & \text{if } N_e(\mathbf{d}) = N, \end{cases} \end{aligned}$$

$\forall \mathbf{d} \in \mathcal{D}$, where $N_e(\mathbf{d})$ is the number of distinct demands.

Proof. For $N < K$ and $M = 1/K$, the CFL prefetching scheme $\mathcal{M}_{\text{CFL}}$ is as follows. Each file is split into NK subfiles $X_i = (X_{i,1}, X_{i,2}, \dots, X_{i,NK})$. User $i, i \in [K]$ caches N coded packets given by $Y_i = X_{1,N(i-1)+j} \oplus \dots \oplus X_{N,N(i-1)+j}$, for $j \in [N]$. Let $X = (X_{1,1}, \dots, X_{1,NK}, \dots, X_{N,1}, \dots, X_{N,NK})$ be the vector obtained by the concatenation of vectors $X_i, i \in [N]$. For a given demand $\mathbf{d}$, this problem becomes a generalized index coding problem $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ with N^2K messages and NK^2 receivers.

First consider that all of the demands are distinct, i.e., $N_e(\mathbf{d}) = N$. Without loss of generality we can assume that the first N users demand distinct files, such that the i th user demands X_{d_i} for $i = 1, 2, \dots, N$. Thus, $\mathbf{d} = (d_1, d_2, \dots, d_K)$, such that $d_i \neq d_j$ for $i, j \in [N]$. Let $\mathcal{E} = \{e_{i,j} : i \in [K], j \in [N]\}$ represent a set of NK equations, where $e_{i,j} : (X_{1,N(i-1)+j} \oplus X_{2,N(i-1)+j} \oplus \dots \oplus X_{N,N(i-1)+j}) = 0$. We consider a subset of the equations in $\mathcal{E}$ of the form $e_{i,j}$ for $i, j \in [N]$. There are N^2 such equations. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{N^2K}$ consisting of vectors that satisfy these equations. From the rank-nullity theorem, we have $\dim(\mathcal{S}) \geq N^2K - N^2 = N^2(K - 1)$.

For $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$, (1) can be rewritten as $\mathcal{Z}^{(i,j)} \triangleq \{\mathbf{z} \in \mathbb{F}_q^{N^2K} : e_{i,k} \text{ for } k \in [N], X_{d_i,j} \neq 0\}$ for $i \in [K]$ and $j \in [NK]$. Let $\mathcal{A}(\mathbf{d}) = \cup_{i \in [K], j \in [NK]} \mathcal{Z}^{(i,j)} \cup \{\mathbf{0}\}$. The generalized independence number is the maximum among the dimensions of all the subspaces of $\mathbb{F}_q^{N^2K}$ in $\mathcal{A}(\mathbf{d})$. We show that $\mathcal{S}$ is such a subspace. For this, we need to show that all of the vectors of $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. By the definition of $\mathcal{A}(\mathbf{d})$, the all zero vector $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. The vectors that belong to $\mathcal{S}$ having $X_{d_i,j} \neq 0$ belong to the set $\mathcal{Z}^{(i,j)}$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$ and $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq N^2(K - 1)$. From (5), we get $\kappa(\mathcal{M}_{\text{CFL}}) \leq N^2(K - 1)$. Hence, by (2), we have $\alpha(\mathcal{M}_{\text{CFL}}) = \kappa(\mathcal{M}_{\text{CFL}}) = N^2(K - 1)$.

Consider the case where $N_e(\mathbf{d}) \leq N - 1$. Let the first $N_e(\mathbf{d})$ demands be distinct and the i th user demands X_{d_i} for $i \in [N_e(\mathbf{d})]$. Without loss of generality we can assume that the indices of the files that are not demanded are $N_e(\mathbf{d}) + 1, \dots, N$. There are $U = N - N_e(\mathbf{d})$ files that are not demanded. In addition to the NK equations that are presented in $\mathcal{E}$, consider the following equations $X_{N_e(\mathbf{d})+i,j} = 0$, for $i \in [U - 1]$ and $j \in [NK]$. The number of these equations is thus $NK + NK(U - 1) = NKU = NK(N - N_e(\mathbf{d}))$. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{N^2K}$ which consists of the vectors satisfying these equations. By the rank-nullity theorem, $\dim(\mathcal{S}) \geq N^2K - NK(N - N_e(\mathbf{d})) = NK(N_e(\mathbf{d}))$. By definition, $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. Any vector in $\mathcal{S}$ with the coordinate $X_{d_i,j} \neq 0$ for $i \in [N_e(\mathbf{d})]$ lies in $\mathcal{Z}^{(i,j)}$. The set of equations force all $X_{i,j} = 0$ for $i \in \{N_e(\mathbf{d}), \dots, N - 1\}$ and $j \in [NK]$. Moreover, by the set of equations

presented in $\mathcal{E}$, $X_{N,j} \neq 0$ would mean some other $X_{d,i,j} \neq 0$ for $i \in [N_e(\mathbf{d})]$. Hence, any vector with $X_{N,j} \neq 0$ lies in some $\mathcal{Z}^{(i,j)}$ for $i \in [N_e(\mathbf{d})]$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. Therefore $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq NKN_e(\mathbf{d})$. From (4), we have $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) \leq NKN_e(\mathbf{d})$. Hence from (2), $\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = \kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = NKN_e(\mathbf{d})$. $\square$

4. Optimal Linear Error Correcting Delivery Scheme for the CFL Prefetching Scheme

In this section we give expressions for the average rate and the worst case rate for a δ -error correcting delivery scheme for the CFL prefetching scheme. Also we propose a δ -error correcting delivery scheme for this case. From Theorem 1 and Theorem 2, we can conclude that for all the generalized index coding problems $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ induced from the CFL prefetching scheme,

$$\alpha(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = \kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}). \quad (6)$$

Hence, the α and κ bounds in (3) meet. Using this, the optimal linear error correcting delivery scheme can be constructed for the CFL prefetching scheme and hence the average rate can be calculated, as given in the following theorem.

Theorem 3. For a coded caching problem with the CFL prefetching scheme for $M = 1/K$,

$$R^*(\mathcal{M}_{\text{CFL}}, \delta) = \mathbb{E}_{\mathbf{d}} \left[\frac{N_q[\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}), 2\delta + 1]}{n_{\text{CFL}}} \right],$$

where n_{CFL} is the number of subfiles into which each file is divided in the CFL scheme.

Proof. From (6) and (3), we can conclude that, for any generalized index coding problem induced from the coded caching problem with CFL prefetching, the α and κ bounds meet. Thus, the optimal linear error correcting delivery scheme would be the concatenation of the CFL delivery scheme with an optimal linear error correcting code. Thus, the optimal length for δ error corrections in those generalized index coding problems is $N_q[\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}), 2\delta + 1]$ and, hence, the statement of the theorem follows. $\square$

Corollary 1. For a coded caching problem with the CFL prefetching scheme for $M = 1/K$,

$$R_{\text{worst}}^*(\mathcal{M}_{\text{CFL}}, \delta) = \frac{N_q[\kappa_{\{\text{worst}\}}(\mathcal{M}_{\text{CFL}}, \mathbf{d}), 2\delta + 1]}{n_{\text{CFL}}},$$

where the value of $\kappa_{\{\text{worst}\}}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$ is obtained from (4) and (5) when $N_e(\mathbf{d}) = N$.

Proof. Worst case rate is required when the number of distinct demands is maximum. This happens when $N_e(\mathbf{d}) = N$. $\square$

Because the α and κ bounds become equal for $\mathcal{I}(\mathcal{M}_{\text{CFL}}, \mathbf{d})$, the optimal linear error correcting coded caching delivery scheme here would be the concatenation of the CFL delivery scheme with optimal classical error correcting scheme which corrects δ errors. Decoding can be done by syndrome decoding for error correcting generalized index codes proposed in [24,27].

In the remaining part of this section, few examples of the optimal linear error correcting delivery scheme for coded caching problems with the CFL prefetching are given.

Example 3. Consider the coded caching problem considered in Example 1 depicted in Figure 1. First consider that $N_e(\mathbf{d}) = 3$ and $\mathbf{d} = (1, 2, 3)$. We have shown that, for this case, $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = 6$. The transmissions in the CFL scheme are $T_1 : X_{2,1}$, $T_2 : X_{3,1}$, $T_3 : X_{1,2}$, $T_4 : X_{3,2}$, $T_5 : X_{1,3}$, and $T_6 : X_{2,3}$. If $\delta = 1$ transmission

error needs to be corrected, then, from [37], we have $N_2[6, 3] = 10$. A generator matrix that corresponds to $[10, 6, 3]_2$ code is

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

The optimal linear single error correcting delivery scheme is the concatenation of the CFL delivery scheme with the above code. Thus, single error correcting delivery scheme involves 10 transmissions. In addition to $T_1, \dots, T_6$ the following transmissions are required.

$$\begin{aligned} T_7 &: X_{2,1} \oplus X_{3,1} \oplus X_{1,2}, \\ T_8 &: X_{2,1} \oplus X_{3,2} \oplus X_{1,3}, \\ T_9 &: X_{3,1} \oplus X_{3,2} \oplus X_{2,3} \text{ and} \\ T_{10} &: X_{1,2} \oplus X_{1,3} \oplus X_{2,3}. \end{aligned}$$

Now, consider $N_e(\mathbf{d}) = 2$ and $\mathbf{d} = (1, 2, 1)$. For this case also, $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = 6$. The transmissions in the CFL scheme are $T_1 : X_{1,1}$, $T_2 : X_{1,2}$, $T_3 : X_{1,3}$, $T_4 : X_{2,1}$, $T_5 : X_{2,2}$ and $T_6 : X_{2,3}$. For single error correction, the concatenation is done with the same $[10, 6, 3]_2$ code. Considering the same generator matrix as before, the additional transmissions in the error correcting delivery scheme are

$$\begin{aligned} T_7 &: X_{1,1} \oplus X_{1,2} \oplus X_{1,3}, \\ T_8 &: X_{1,1} \oplus X_{2,1} \oplus X_{2,2}, \\ T_9 &: X_{1,2} \oplus X_{2,1} \oplus X_{2,3} \text{ and} \\ T_{10} &: X_{1,3} \oplus X_{2,2} \oplus X_{2,3}. \end{aligned}$$

Finally, consider $N_e(\mathbf{d}) = 1$ and $\mathbf{d} = (1, 1, 1)$. For this case, $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = 3$. The CFL transmission scheme involves the following three transmissions $T_1 : X_{1,1}$, $T_2 : X_{1,2}$, and $T_3 : X_{1,3}$. For single error correction, we have from [37] that $N_2[3, 3] = 6$. A generator matrix for the $[6, 3, 3]_2$ code is

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

Thus, the optimal linear single error correcting delivery scheme is the concatenation of the CFL delivery scheme with the above code. The additional transmissions required apart from T_1, T_2 and T_3 are

$$\begin{aligned} T_4 &: X_{1,1} \oplus X_{1,2}, \\ T_5 &: X_{1,1} \oplus X_{1,3} \text{ and} \\ T_6 &: X_{1,2} \oplus X_{1,3}. \end{aligned}$$

Decoding is done by syndrome decoding for generalized index codes [27] [24].

Example 4. Consider the coded caching problem considered in Example 2 depicted in Figure 2. Consider that $N_e(\mathbf{d}) = 3$ and $\mathbf{d} = (1, 2, 3, 1)$. We have shown that for this case $\kappa(\mathcal{M}_{\text{CFL}}, \mathbf{d}) = 27$. The transmissions in the CFL scheme are $T_1 : X_{2,1}$, $T_2 : X_{3,1}$, $T_3 : X_{2,2}$, $T_4 : X_{3,2}$, $T_5 : X_{2,3}$, $T_6 : X_{3,3}$, $T_7 : X_{1,4}$, $T_8 : X_{3,4}$, $T_9 : X_{1,5}$, $T_{10} : X_{3,5}$, $T_{11} : X_{1,6}$, $T_{12} : X_{3,6}$, $T_{13} : X_{1,7}$, $T_{14} : X_{2,7}$, $T_{15} : X_{1,8}$, $T_{16} : X_{2,8}$, $T_{17} : X_{1,9}$, $T_{18} : X_{2,9}$, $T_{19} : X_{2,10}$, $T_{20} : X_{3,10}$, $T_{21} : X_{2,11}$, $T_{22} : X_{3,11}$, $T_{23} : X_{2,12}$, $T_{24} : X_{3,12}$, $T_{25} : X_{1,1} \oplus X_{1,10}$, $T_{26} : X_{1,2} \oplus X_{1,11}$, and $T_{27} : X_{1,3} \oplus X_{1,12}$. If $\delta = 1$ transmission error needs to be corrected, then from [37], we have $N_2[27, 3] = 42$.

The optimal linear single error correcting delivery scheme involves concatenation of CFL delivery scheme with a generator matrix that corresponds to the $[42, 27, 3]_2$ code.

5. Optimal Linear Error Correcting Scheme for $M = 1/K$ with Reduced Sub-Packetization

In this section, we consider the scheme presented in [4], which we call the JV scheme for $M = 1/K$ and discuss the error correction. For the case $N = K$, the JV scheme is exactly same as the CFL scheme. For $K > N$, the JV scheme has an advantage in terms of sub-packetization. The JV scheme splits each file into K subfiles where as CFL scheme required NK subfiles. Here, we find a closed form expression for the generalized independence number $\alpha(\mathcal{M}_{JV}, \mathbf{d})$ of the index coding problem $\mathcal{I}(\mathcal{M}_{JV}, \mathbf{d})$. For all of the index coding problems that correspond to all possible demands, we show that the generalized independence number is equal to the min-rank. Hence in this case also, α and κ bounds meet and the optimal error correcting delivery scheme is obtained by concatenation scheme.

The expression for min-rank for this scheme can be obtained from the achievable rate expressions in [3,4], as

$$\kappa(\mathcal{M}_{JV}, \mathbf{d}) \leq \begin{cases} KN_e(\mathbf{d}) & \text{if } N_e(\mathbf{d}) \leq N - 1 \\ N(K - 1) & \text{if } N_e(\mathbf{d}) = N \end{cases}. \quad (7)$$

The calculation of generalized independence number and how it becomes equal to the min-rank are illustrated in the following example.

Example 5. Consider a coded caching problem with $N = 3$, $K = 6$ and $M = 1/6$. According to the JV scheme, each file is split into $K = 6$ subfiles as: $X_i = (X_{i,1}, X_{i,2}, X_{i,3}, X_{i,4}, X_{i,5}, X_{i,6})$ for $i \in \{1, 2, 3\}$. Let X be the vector obtained by concatenation of X_1 , X_2 and X_3 , given by $X = (X_{1,1}, \dots, X_{1,6}, X_{2,1}, \dots, X_{2,6}, X_{3,1}, \dots, X_{3,6})$. The cache contents of user i is $Y_i = X_{1,i} \oplus X_{2,i} \oplus X_{3,i}$ for $i \in \{1, 2, \dots, 6\}$. For a given demand $\mathbf{d}$, this problem becomes a generalized index coding problem $\mathcal{I}(\mathcal{M}_{JV}, \mathbf{d})$, having 18 messages and 36 receivers.

First consider that $N_e(\mathbf{d}) = N = 3$ and let $\mathbf{d} = (1, 2, 3, 1, 2, 3)$. Consider the equations $e_i : X_{1,i} \oplus X_{2,i} \oplus X_{3,i} = 0$ for $i \in \{1, 2, \dots, 6\}$. Out of these, consider the first three equations e_1, e_2 and e_3 . Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{18}$ consisting of vectors which satisfy these three equations. Subsequently, from the rank-nullity theorem, we have $\dim(\mathcal{S}) \geq 18 - 3 = 15$. For this case, (1) can be rewritten as $\mathcal{Z}^{(i,j)} \triangleq \{\mathbf{z} \in \mathbb{F}_q^{18} : e_i, X_{d,i,j} \neq 0\}$ for $i, j \in [6]$. Let $\mathcal{A}(\mathbf{d}) = \cup_{i \in [6], j \in [6]} \mathcal{Z}^{(i,j)} \cup \{\mathbf{0}\}$. The generalized independence number is the maximum among the dimensions of all the subspaces of $\mathbb{F}_q^{18}$ in $\mathcal{A}(\mathbf{d})$. We claim that $\mathcal{S}$ is such a subspace. This would mean that $\alpha(\mathcal{M}_{JV}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq 15$. For this, we need to show that all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. By the definition of $\mathcal{A}(\mathbf{d})$, the all zero vector $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. Any other vector in $\mathcal{S}$ will have at least one non-zero coordinate. All of the vectors in $\mathcal{S}$, having $X_{d,i,j} \neq 0$ belongs to $\mathcal{Z}^{(i,j)}$. Because all files are demanded, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$ and $\alpha(\mathcal{M}_{JV}, \mathbf{d}) \geq 15$. From (7), we get $\kappa(\mathcal{M}_{JV}, \mathbf{d}) \leq 3(6 - 1) = 15$. Hence, by (2), we have $\alpha(\mathcal{M}_{JV}, \mathbf{d}) = \kappa(\mathcal{M}_{JV}, \mathbf{d}) = 15$.

Consider now that $N_e(\mathbf{d}) = 2$ and $\mathbf{d} = (1, 2, 1, 2, 1, 2)$. In this case, consider the six equations $e_1, e_2, \dots, e_6$. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{18}$ consisting of vectors which satisfy these equations. Hence from the rank-nullity theorem, we have $\dim(\mathcal{S}) \geq 18 - 6 = 12$. By definition, $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. Any non-zero vector in $\mathcal{S}$ with $X_{d,i,j} \neq 0$ for $i = 1, 2$ and $j = 1, 2, \dots, 6$ lies in the corresponding $\mathcal{Z}^{(i,j)}$. Now, we have to consider the vectors that satisfy these equations and $X_{3,j} \neq 0$ for $j = 1, 2, \dots, 6$. From the equations e_i , any $X_{3,j} \neq 0$ forces some $X_{i,j} \neq 0$ for $i = 1, 2$ and, hence, such vectors also lie in $\mathcal{A}(\mathbf{d})$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. Therefore, $\alpha(\mathcal{M}_{JV}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq 12$. From (7), we get $\kappa(\mathcal{M}_{JV}, \mathbf{d}) \leq 6(2) = 12$. Hence, by (2), we have $\alpha(\mathcal{M}_{JV}, \mathbf{d}) = \kappa(\mathcal{M}_{JV}, \mathbf{d}) = 12$.

Finally, consider $N_e(\mathbf{d}) = 1$ and $\mathbf{d} = (1, 1, 1, 1, 1, 1)$. The files X_2 and X_3 are not demanded by any user. Here, we consider the equation $e_1, e_2, \dots, e_6$. In addition to this, consider the equations $X_{2,1} = 0$, $X_{2,2} = 0$, $X_{2,3} = 0$, $X_{2,4} = 0$, $X_{2,5} = 0$ and $X_{2,6} = 0$. Thus, there are twelve equations in total. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{18}$ which satisfy these equations. By the rank-nullity theorem, the dimension of $\mathcal{S}$ is given by $\dim(\mathcal{S}) \geq 18 - 12 = 6$. The next step is to show that all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. The all zero vector $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$ by definition. Any non-zero vector in $\mathcal{S}$ with $X_{1,j} \neq 0$ for $j \in [6]$ lies in the corresponding $\mathcal{Z}^{(1,j)}$.

From the set of twelve equations, $X_{3,j} \neq 0$ force some $X_{1,j} \neq 0$ for $j \in [6]$ and, hence, such vectors also lie in $\mathcal{A}(\mathbf{d})$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. Therefore, $\alpha(\mathcal{M}_{JV}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq 6$. From (7), we get $\kappa(\mathcal{M}_{JV}, \mathbf{d}) \leq 6(1) = 6$. Hence, by (2), we have $\alpha(\mathcal{M}_{JV}, \mathbf{d}) = \kappa(\mathcal{M}_{JV}, \mathbf{d}) = 6$.

This can be generalized as in the theorem below.

Theorem 4. For $N < K$ and $M = 1/K$,

$$\alpha(\mathcal{M}_{JV}, \mathbf{d}) = \kappa(\mathcal{M}_{JV}, \mathbf{d}) = \begin{cases} KN_e(\mathbf{d}) & \text{if } N_e(\mathbf{d}) \leq N - 1 \\ N(K - 1) & \text{if } N_e(\mathbf{d}) = N, \end{cases}$$

$\forall \mathbf{d} \in \mathcal{D}$, where $N_e(\mathbf{d})$ is the number of distinct demands.

Proof. For $N < K$ and $M = 1/K$, the JV prefetching scheme $\mathcal{M}_{JV}$ is as follows. Each file is split into K subfiles $X_i = (X_{i,1}, X_{i,2}, \dots, X_{i,K})$. The cache content of user i is given as $Y_i = X_{1,i} \oplus X_{2,i} \oplus \dots \oplus X_{N,i}$. Let $X = (X_{1,1}, \dots, X_{1,K}, X_{2,1}, \dots, X_{2,K}, X_{3,1}, \dots, X_{N,1}, \dots, X_{N,K})$ be the vector obtained by the concatenation of vectors $X_i, i \in [N]$. For a given demand $\mathbf{d}$, this problem becomes a generalized index coding problem $\mathcal{I}(\mathcal{M}_{JV}, \mathbf{d})$ with NK messages and K^2 receivers.

First, consider that all of the demands are distinct, i.e., $N_e(\mathbf{d}) = N$. Without a loss of generality, we can assume that the first N users demand distinct files, such that $X_{d_i} = X_i$ for $i = 1, 2, \dots, N$. Thus $\mathbf{d} = (d_1, d_2, \dots, d_K)$, such that $d_i = i$ for $i, j \in [N]$. Consider the set of equations $e_i : X_{1,i} \oplus X_{2,i} \oplus \dots \oplus X_{N,i} = 0$ for $i \in [K]$. Out of these equations, consider a subset of N equations e_i for $i \in [N]$. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{NK}$ consisting of vectors satisfying these equations. From the rank-nullity theorem, we have $\dim(\mathcal{S}) \geq NK - N = N(K - 1)$.

For $\mathcal{I}(\mathcal{M}_{JV}, \mathbf{d})$, (1) can be rewritten as $\mathcal{Z}^{(i,j)} \triangleq \{\mathbf{z} \in \mathbb{F}_q^{NK} : e_i \text{ for } i \in [N], X_{d_i,j} \neq 0\} \text{ for } i, j \in [K]$. Let $\mathcal{A}(\mathbf{d}) = \cup_{i \in [K], j \in [K]} \mathcal{Z}^{(i,j)} \cup \{\mathbf{0}\}$. The generalized independence number is the maximum among the dimensions of all the subspaces of $\mathbb{F}_q^{NK}$ in $\mathcal{A}(\mathbf{d})$. We show that $\mathcal{S}$ is such a subspace. For this, we need to show that all the vectors of $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. By the definition of $\mathcal{A}(\mathbf{d})$, the all zero vector $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. The vectors that belong to $\mathcal{S}$ having $X_{d_i,j} \neq 0$ belong to the set $\mathcal{Z}^{(i,j)}$. Thus, all of the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$ and $\alpha(\mathcal{M}_{JV}, \mathbf{d}) \geq N(K - 1)$. From (7), we get $\kappa(\mathcal{M}_{JV}) \leq N(K - 1)$. Hence, by (2), we have $\alpha(\mathcal{M}_{JV}, \mathbf{d}) = \kappa(\mathcal{M}_{JV}, \mathbf{d}) = N(K - 1)$.

Consider the case where $N_e(\mathbf{d}) \leq N - 1$. Let the first $N_e(\mathbf{d})$ demands be distinct and without loss of generality assume that the i th user demands X_i for $i \in [N_e(\mathbf{d})]$. Thus, the indices of the files that are not demanded are $N_e(\mathbf{d}) + 1, \dots, N$. There are $U = N - N_e(\mathbf{d})$ files that are not demanded. In addition to the K equations e_i , consider the following equations $X_{i,j} = 0$, for $i \in \{N_e(\mathbf{d}) + 1, \dots, N - 1\}$ and $j \in [K]$. Thus, the number of these equations is $K + K(U - 1) = KU = K(N - N_e(\mathbf{d}))$. Let $\mathcal{S}$ be the subspace of $\mathbb{F}_q^{NK}$ which consists of the vectors satisfying these equations. By the rank-nullity theorem, $\dim(\mathcal{S}) \geq NK - K(N - N_e(\mathbf{d})) = KN_e(\mathbf{d})$. By definition, $\mathbf{0}$ lies in $\mathcal{A}(\mathbf{d})$. Any vector in $\mathcal{S}$ with the coordinate $X_{i,j} \neq 0$ for $i \in [N_e(\mathbf{d})]$ lies in $\mathcal{Z}^{(i,j)}$. The set of equations force all $X_{i,j} = 0$ for $i \in \{N_e(\mathbf{d}), \dots, N - 1\}$ and $j \in [K]$. Moreover by the same set of equations, $X_{N,j} \neq 0$ would mean some other $X_{i,j} \neq 0$ for $i \in [N_e(\mathbf{d})]$. Hence any vector with $X_{N,j} \neq 0$ lies in some $\mathcal{Z}^{(i,j)}$ for $i \in [N_e(\mathbf{d})]$. Thus all the vectors in $\mathcal{S}$ lie in $\mathcal{A}(\mathbf{d})$. Therefore, $\alpha(\mathcal{M}_{JV}, \mathbf{d}) \geq \dim(\mathcal{S}) \geq KN_e(\mathbf{d})$. From (7), we have $\kappa(\mathcal{M}_{JV}, \mathbf{d}) \leq KN_e(\mathbf{d})$. Hence, from (2), $\alpha(\mathcal{M}_{JV}, \mathbf{d}) = \kappa(\mathcal{M}_{JV}, \mathbf{d}) = NKN_e(\mathbf{d})$. $\square$

Because, for this scheme, also, α and κ bounds meet, the optimal coded caching delivery scheme here would be the concatenation of the JV delivery scheme with optimal classical error correcting scheme that corrects δ errors. Decoding can be done by syndrome decoding for error correcting generalized index codes proposed in [24,27]. This is illustrated while using an example below.

Example 6. Consider the coded caching problem considered in Example 5. Consider that $N_e(\mathbf{d}) = 3$ and $\mathbf{d} = (1, 2, 3, 1, 2, 3)$. We have shown that for this case $\kappa(\mathcal{M}_{JV}, \mathbf{d}) = 15$. The transmissions in the JV scheme

are $T_1 : X_{1,2}, T_2 : X_{1,3}, T_3 : X_{1,5}, T_4 : X_{1,6}, T_5 : X_{1,4} \oplus X_{1,1}, T_6 : X_{2,1}, T_7 : X_{2,3}, T_8 : X_{2,4}, T_9 : X_{2,6}, T_{10} : X_{2,5} \oplus X_{2,2}, T_{11} : X_{3,1}, T_{12} : X_{3,2}, T_{13} : X_{3,4}, T_{14} : X_{3,5}, T_{15} : X_{3,6} \oplus X_{3,3}$. If $\delta = 1$ transmission error needs to be corrected, then from [37], we have $N_2[15, 3] = 20$. The optimal linear single error correcting delivery scheme involves a concatenation of JV delivery scheme with a generator matrix that corresponds to the $[20, 15, 3]_2$ code.

6. Conclusions

In this work, we obtained the minimum number of transmissions required for a linear δ -error correcting delivery scheme for coded caching problems with the CFL prefetching scheme. We proposed an optimal linear error correcting delivery scheme for the above case. We also found closed form expressions for the average rate and the peak rate for these problems. We considered the JV scheme, which uses low sub-packetization as compared to CFL scheme and proved that for this scheme also concatenation of JV delivery scheme with optimal error correcting code is optimal.

Author Contributions: Conceptualization, N.S.K., A.T. and B.S.R.; writing—original draft preparation, N.S.K. and A.T.; writing—review and editing, N.S.K., A.T. and B.S.R.; supervision, B.S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Engineering Research Board (SERB) of Department of Science and Technology (DST), Government of India, through J. C. Bose National Fellowship to B. Sundar Rajan.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Maddah-Ali, M.A.; Niesen, U. Fundamental limits of caching. *IEEE Trans. Inf. Theory* **2014**, *60*, 2856–2867.
2. Yu, Q.; Maddah-Ali, M.A.; Avestimehr, A.S. The exact rate-memory tradeoff for caching with uncoded prefetching. *IEEE Trans. Inf. Theory* **2018**, *64*, 1281–1296.
3. Chen, Z.; Fan, P.; Letaief, K.B. Fundamental limits of caching: improved bounds for users with small buffers. *IET Commun.* **2016**, *10*, 2315–2318.
4. Gomez-Vilardebo, J. Fundamental limits of caching: improved rate-memory tradeoff with coded prefetching. *IEEE Trans. Commun.* **2018**, *66*, 4488–4497.
5. Amiri, M.M.; Gündüz, D. Fundamental limits of coded caching: improved delivery rate-cache capacity tradeoff. *IEEE Trans. Commun.* **2017**, *65*, 806–815.
6. Tian, C.; Chen, J. Caching and delivery via interference elimination. *IEEE Trans. Inf. Theory* **2018**, *64*, 1548–1560.
7. Zhang, K.; Tian, C. Fundamental limits of coded caching: from uncoded prefetching to coded prefetching. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 1153–1164.
8. Maddah-Ali, M.A.; Niesen, U. Decentralized coded caching attains order-optimal memory-rate tradeoff. *IEEE/ACM Trans. Netw.* **2015**, *23*, 1029–1040.
9. Niesen, U.; Maddah-Ali, M.A. Coded caching with nonuniform demands. *IEEE Trans. Inf. Theory* **2017**, *63*, 1146–1158.
10. Pedarsani, R.; Maddah-Ali, M.A.; Niesen, U. Online coded caching. *IEEE/ACM Trans. Netw.* **2016**, *24*, 836–845.
11. Karamchandani, N.; Niesen, U.; Maddah-Ali, M.A.; Diggavi, S.N. Hierarchical coded caching. *IEEE Trans. Inf. Theory* **2016**, *62*, 3212–3229.
12. Ji, M.; Caire, G.; Molisch, A.F. Fundamental limits of caching in wireless D2D networks. *IEEE Trans. Inf. Theory* **2016**, *62*, 849–869.
13. Chen, J.; Lee, V.C.S.; Liu, K.; Li, J. Efficient cache management for network-coding-assisted data broadcast. *IEEE Trans. Veh. Technol.* **2017**, *66*, 3361–3375.
14. Kiskani, M.K.; Sadjadpour, H.R. A secure approach for caching contents in wireless ad hoc networks. *IEEE Trans. Veh. Technol.* **2017**, *66*, 10249–10258.

15. Karat, N.S.; Thomas, A.; Rajan, B.S. Optimal error correcting delivery scheme for coded caching with symmetric batch prefetching. In Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018.
16. Bidokhti, S.S.; Wigger, M.; Timo, R. Noisy broadcast networks with receiver caching. *IEEE Trans. Inf. Theory* **2018**, *64*, 6996–7016.
17. Timo, R.; Wigger, M. Joint cache-channel coding over erasure broadcast channels. In Proceedings of the IEEE International Symposium on Wireless Communications Systems (ISWCS), Bruxelles, Belgium, 25–28 August 2015.
18. Bidokhti, S.S.; Wigger, M.; Timo, R. Erasure broadcast networks with receiver caching. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016.
19. Bidokhti, S.S.; Wigger, M.; Timo, R. An upper bound on the capacity-memory tradeoff of degraded broadcast channels. In Proceedings of the International Symposium on Turbo Codes and Iterative Information Processing, Brest, France, 5–9 September 2016.
20. Karat, N.S.; Thomas, A.; Rajan, B.S. Optimal error correcting delivery scheme for an optimal coded caching scheme with small buffers. In Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018.
21. Karat, N.S.; Thomas, A.; Rajan, B.S. Error correction in coded caching with symmetric batch prefetching, *IEEE Trans. Commun.* **2019**, *67*, 5264–5274.
22. Birk, Y.; Kol, T. Coding-on-demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients. *IEEE Trans. Inf. Theory* **2006**, *52*, 2825–2830.
23. Dai, M.; Shum, K.W.; Sung, C.W. Data dissemination with side information and feedback. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 4708–4720.
24. Byrne, E.; Calderini, M. Error correction for index coding with coded side information. *IEEE Trans. Inf. Theory* **2017**, *63*, 3712–3728.
25. Shum, K.W.; Dai, M.; Sung, C.W. Broadcasting with coded side information. In Proceedings of the 2012 IEEE 23rd International Symposium Personal Indoor and Mobile Radio Communications (PIMRC), Sydney, NSW, Australia, 9–12 September 2012.
26. Karat, N.S.; Samuel, S.; Rajan, B.S. Optimal error correcting index codes for some generalized index coding problems. *IEEE Trans. Commun.* **2019**, *67*, 929–942.
27. Dau, S.H.; Skachek, V.; Chee, Y.M. Error correction for index coding with side information. *IEEE Trans. Inf. Theory* **2013**, *59*, 1517–1531.
28. Samuel, S.; Rajan, B.S. Optimal linear error-correcting index codes for single-prior index-coding with side information. In Proceedings of the 2017 IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, USA, 19–22 March 2017.
29. Karat, N.S.; Rajan, B.S. Optimal linear error correcting index codes for some index coding problems. In Proceedings of the 2017 IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, USA, 19–22 March 2017.
30. Samuel, S.; Karat, N.S.; Rajan, B.S. Optimal linear error correcting index codes for some generalized index-coding problems. In Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017.
31. Reddy, K.S.; Karamchandani, N. On the exact rate-memory trade-off for multi-access coded caching with uncoded placement," in Proceedings of the International Conference on Signal processing and Communications (SPCOM) 2018, Bangalore, India, 16–19 July 2018.
32. Parrinello, E.; Unsal, A.; Elia, P. Fundamental limits of coded caching with multiple antennas, shared caches and uncoded prefetching. *IEEE Trans. Inf. Theory* **2020**, *66*, 2252–2268.
33. Ji, M.; Tulino, A.M.; Llorca, J.; Caire, G. Caching and coded multicasting: multiple groupcast index coding. In Proceedings of the 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Atlanta, GA, USA, 3–5 December 2014.
34. Wan, K.; Tuninetti, D.; Piantanida, P. A novel index coding scheme and its application to coded caching. In Proceedings of the 2017 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 12–17 February 2017.
35. Kiskani, M.K.; Sadjadpour, H.R. Multihop caching-aided coded multicasting for the next generation of cellular networks. *IEEE Trans. Veh. Technol.* **2017**, *66*, 2576–2585.

36. Zhang, L.; Wang, Z.; Xiao, M.; Wu, G.; Liang, Y. Li, S. Decentralized caching schemes and performance limits in two-layer networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 12177–12192.
37. Grassl, M. Bounds on the Minimum Distance of Linear Codes and Quantum Codes. Available online: <http://www.codetables.de> (accessed on 11 July 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).