

A framework of constructing placement delivery arrays for centralized coded caching

Minquan Cheng, Jinyu Wang, Xi Zhong, Qiang Wang

Abstract

In caching system, it is desirable to design a coded caching scheme with the transmission load R and subpacketization F as small as possible, in order to improve efficiency of transmission in the peak traffic times and to decrease implementation complexity. Yan et al. reformulated the centralized coded caching scheme as designing a corresponding $F \times K$ array called placement delivery array (PDA), where F is the subpacketization and K is the number of users. Motivated by several constructions of PDAs, we introduce a framework for constructing PDAs, where each row is indexed by a row vector of some matrix called row index matrix and each column's index is labelled by an element of a direct product set. Using this framework, a new scheme is obtained, which can be regarded as a generalization of some previously known schemes. When K is equal to $\binom{m}{t}q^t$ for positive integers m , t with $t < m$ and $q \geq 2$, we show that the row index matrix must be an orthogonal array if all the users have the same memory size. Furthermore, the row index matrix must be a covering array if the coded gain is $\binom{m}{t}$, which is the maximal coded gain under our framework. Consequently the lower bounds on the transmission load and subpacketization of the schemes are derived under our framework. Finally, using orthogonal arrays as the row index matrix, we obtain two more explicit classes of schemes which have significant advantages on the subpacketization while the transmission load is equal or close to that of the schemes constructed by Shangguan et al. (IEEE Trans. Inf. Theory, 64, 5755-5766, 2018) for the same number of users and memory size.

Index Terms

Coded caching scheme, placement delivery array, transmission load, subpacketization, orthogonal array, MDS code.

I. INTRODUCTION

THE wireless network has been imposed a tremendous pressure on the data transmission during the peak traffic times due to the explosive increasing mobile services, especially the video streaming. Caching system has been recognized as an efficient solution to reduce such pressure. In a caching system, some contents are proactively placed into the users' memory during the off peak traffic times. Then the traffic amount could be reduced when the cached content is required by users during the peak traffic times. So traditionally almost all of studies on caching system focused on exploiting the history or statistics of the user demands for an appropriate caching strategy. Maddah-Ali and Niesen [1] showed that coded caching can further reduce the traffic amount during the peak-traffic times by exploiting caches to create multicast opportunities.

A. Centralized caching system

In a centralized (K, M, N) caching system (see Figure 1), a single server containing N independent files with the same length connects to K users over a shared link and each user has a cache memory of size M files with $N \geq K$ and $N \geq M$. Denote the N files by $\mathcal{W} = \{W_0, \dots, W_{N-1}\}$ and K users by $\mathcal{K} = \{0, 1, \dots, K-1\}$. For such a system, an F -division (K, M, N) coded caching scheme consists of two phases as follows [1]:

- **Placement phase:** During the off peak traffic times, each file is divided into F equal packets¹, i.e., $W_i = \{W_{i,j} : j = 0, 1, \dots, F-1\}$. Then each user caches some packets (or linear combinations of packets) from the server. If packets are cached directly, it is called uncoded placement; if linear combinations of packets are cached, we call it coded placement. Denote \mathcal{Z}_k the contents cached by user k . In this phase we assume that the server does not know the users' requests in the following phase. For simplicity, most known results are proposed under the assumption of *identical uncoded caching policy*, i.e., if some user caches the i -th packet of some file, then it caches the i -th packet of all files.
- **Delivery phase:** During the peak traffic times, each user randomly requests one file from the files set \mathcal{W} independently. The request vector is denoted by $\mathbf{d} = (d_0, \dots, d_{K-1})$, i.e., user k requests the d_k -th file W_{d_k} , where $d_k \in [0, N)$ and $k \in \mathcal{K}$. The server broadcasts a coded signal (XOR of some required packets) of size S_d packets to users such that each user is able to recover its requested file with the help of its cache contents.

¹Memory sharing technique may lead to non equally divided packets [1], in this paper, we will not discuss this case.

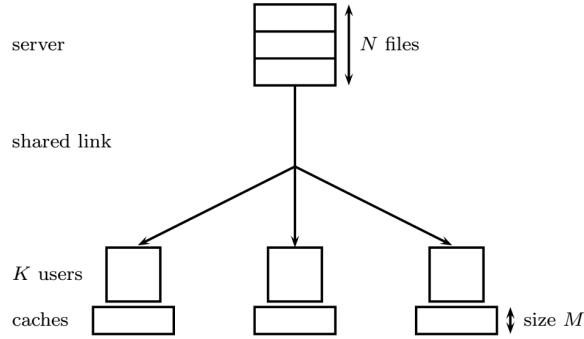


Fig. 1: Centralized (K, M, N) caching system

In this paper, we focus on the worst-case demand pattern, i.e., all the users require different files. In this case, the transmission load of a coded caching scheme is defined as the maximal normalized transmission amount among all the requests in the delivery phase, i.e.

$$R = \max_{\mathbf{d} \in \{0, \dots, N-1\}^K} \left\{ \frac{S_{\mathbf{d}}}{F} \right\}.$$

Since the implementation complexity of a coded caching scheme increases along with its subpacketization F , it is desirable to design a scheme with the transmission load and the subpacketization as small as possible.

B. Prior work

In this paper, we are interested in coded caching schemes with uncoded placement for a centralized caching system. There are many related studies such as [1], [5]–[21]. These results focus on either the transmission load or the subpacketization.

If each file in the library is large enough and $\frac{KM}{N}$ is an integer, Maddah-Ali and Niesen [1] proposed the first coded caching scheme with subpacketization $F = \binom{K}{KM/N}$ and obtained a lower bound on the transmission load. This scheme is referred to as the MN scheme, which is listed in the second row of Table I. In [6], an improved lower bound of the transmission load was derived by Ghasemi and Ramamoorthy using a combinatorial problem of optimally labeling the leaves of a directed tree. Wan et al. [7] showed that the MN scheme has minimum transmission load by the graph theory when $K < N$; Yu et al. [8] obtained the transmission load for various demand patterns by modifying the delivery phase of the MN scheme. There are some studies focusing on the transmission load under other conditions. For example, Jin, et al. [9] derived the average transmission load for conditions of non-uniform file popularity and various demand patterns respectively, using an optimization framework under a specific caching rule. When $K < N$ and all files have the same popularity, they also showed that the minimum transmission load is exactly that of the MN scheme.

It is well known that there is a tradeoff between R and F . To the best of our knowledge, most known results focusing on the subpacketization are proposed under identical uncoded caching policy. When K is large, the first tradeoff between F and R was derived by Shanmugam et al. in [12] using probabilistic arguments. For an arbitrary positive integer K , it is still hard to derive the tradeoff between F and R . Hence most previously known constructions focus on reducing the subpacketization. In [13] Cheng et al. showed that, given the minimum transmission load, the minimum subpacketization is $F = \binom{K}{KM/N}$, i.e., the subpacketization of the MN scheme. It is easy to see that the subpacketization F of the MN scheme increases very quickly when the number of users K increases. The first scheme with lower subpacketization compared with the MN scheme was proposed by Shanmugam et al. in [12] by a grouping method. Yan et al. in [14] generated two classes of schemes by means of a combinatorial structure called placement delivery array (PDA), which is an $F \times K$ array and could be used to characterize the placement and delivery phases for all possible requests. There are some other coded caching schemes with lower subpacketization levels obtained by increasing the transmission load R , for instance, schemes by means of linear block codes [15], the special $(6, 3)$ -free hypergraphs [16], the (r, t) Ruzsa-Szemerédi graphs [17], the strong edge coloring of bipartite graphs [18], the projective space over \mathbb{F}_p [19], combinatorial design [20] and so on. In [22] the authors generalized the constructions in [15] and proposed two classes of schemes with flexible memory size based on the constructions in [14], [16] using the generalized grouping method in [21]. By means of the MN scheme, the authors in [13] obtained two variant MN schemes where F is minimal for the fixed R and memory size M . Some of these schemes are listed in Table I, which was summarized by Krishnan in [19].

TABLE I: Summary of some known results

Schemes and parameters	User number K	Caching ratio $\frac{M}{N}$	Load R	Subpacketization F
MN scheme in [1], any $k, t \in \mathbb{Z}^+$ with $t < k$	k	$\frac{t}{k}$	$\frac{k-t}{1+t}$	$\binom{k}{t}$
Scheme in [14], any m and $q \in \mathbb{Z}^+$	$(m+1)q$	$\frac{1}{q}$	$q-1$	q^m
Scheme in [16], any $m, t, q \in \mathbb{Z}^+$ with $t < m$	$\binom{m}{t}q^t$	$1 - \left(\frac{q-1}{q}\right)^t$	$(q-1)^t$	q^m
Scheme in [13], any $k, t \in \mathbb{Z}^+$ with $t < k$	$\binom{k}{t+1}$	$1 - \frac{t+1}{\binom{k}{t}}$	$\frac{k}{\binom{k}{t}}$	$\binom{k}{t}$
	$\binom{k}{t}$	$\frac{t}{k}$	$\frac{\binom{k+1}{t+1}}{k}$	k
Scheme in [18], any $m, a, b, \lambda \in \mathbb{Z}^+$ with $a, b < m$ and $\lambda < \min\{a, b\}$	$\binom{m}{a}$	$1 - \frac{\binom{m-a}{\lambda} \binom{m-a}{b-\lambda}}{\binom{m}{b}}$	$\frac{\binom{m}{a+b-2\lambda} \binom{a+b-2\lambda}{a-\lambda}}{\binom{m}{b}}$	$\binom{m}{b}$

C. Contribution and organization of this paper

In [23], Shanmugam et al. discovered that most of coded caching schemes can be recasted into a PDA when $K \leq N$. This implies that PDA is a useful tool for designing coded caching schemes. It is worth noting that PDA can also be used to study distributed computing [25], [26], privacy coded caching [27], [28], linear function retrieval [29] and so on. In this paper, we will consider the schemes by means of PDAs when $K \leq N$.

Motivated by several constructions of PDA, in this paper we introduce a new framework for constructing PDA, where each row index is denoted by a row vector of some matrix called row index matrix and each column index is denoted by an element of a direct product set called column index set. Then the problem of designing a PDA (i.e., coded caching scheme) can be transformed into a problem of choosing a row index matrix and a column index set appropriately. It is worth noting that some previously known schemes (e.g., the MN scheme [1], the first PDA in [14], PDAs in [16]) can be obtained under our framework. For $K = \binom{m}{t}q^t$ for positive integers m, t with $t < m$ and $q \geq 2$, we obtain the following results:

- If all the users have the same memory size, the row index matrix must be an orthogonal array with degree m, q levels and strength t .
- If the coded gain is $\binom{m}{t}$, which is the maximal coded gain under our framework, the row index matrix must be a covering array with degree m, q levels and strength $m-t$. Consequently a lower bound on the subpacketization of the schemes with the maximal coded gain $\binom{m}{t}$ is obtained, i.e., $F \geq q^{m-t}$.

Here orthogonal arrays and covering arrays are classic combinatorial structures in combinatorial design theory [24] and they will be introduced in detail in Section IV. The above results imply the following two statements:

- When $K = \binom{m}{t}q^t$, if we want to design a coded caching scheme with the maximal coded gain $g = \binom{m}{t}$, we only need to choose an appropriate row index matrix which is not only an orthogonal array with strength t but also a covering array with strength $m-t$.
- The scheme in [14] listed in the third row of Table I achieves our lower bound on the subpacketization.

In addition, we present three new constructions of PDA under our framework, which are listed in Table II. It is worth noting that compared with the scheme from [16] (see Table I), the second construction reduces the subpacketization by a factor of q without sacrificing the transmission load and the third construction reduces the subpacketization by a factor of q^t at a cost of transmission load increasing from $(q-1)^t$ to q^t-1 for the same number of users and the same memory size.

TABLE II: Some main schemes in this paper

Schemes and parameters	User number K	Caching ratio $\frac{M}{N}$	Load R	Packet number F
Theorem 2, any $m, s, t \in \mathbb{Z}^+$, $\omega \in \mathbb{Z}$ with $0 \leq \omega \leq t \leq s, s+t-2\omega \leq m$	$\binom{t}{\omega} \binom{m}{t}$	$1 - \frac{\binom{m-t}{s-\omega}}{\binom{m}{s}}$	$\binom{m}{s+t-2\omega} / \binom{m}{s}$	$\binom{m}{s}$
Theorem 6, any $m, s, t, q \in \mathbb{Z}^+$, $0 < t < m$	$\binom{m}{t}q^t$	$1 - \left(\frac{q-1}{q}\right)^t$	$(q-1)^t$	q^{m-1}
Theorem 7, any $m, t \in \mathbb{Z}^+$ with $2t \leq m$, and some prime power q	$\binom{m}{t}q^t$	$1 - \left(\frac{q-1}{q}\right)^t$	$q^t - 1$	q^{m-t}

The rest of this paper is organized as follows.

In Section II we introduce basic notations, the definition of PDA and the relationship between a PDA and a coded caching scheme. Then in Section III a framework for constructing PDA is proposed and a new scheme is obtained, which can be regarded as a generalization of several previously known schemes. In Section IV, the requirements for the row index matrix is discussed when each user has the same memory size or the coded gain is the largest. Another two new schemes based on orthogonal arrays are obtained in Section V. Finally the conclusion is drawn in Section VI.

II. PRELIMINARIES

In this paper, we will use the following notations unless otherwise stated.

- We use bold capital letters, bold lower case letters and curlicue letters to denote arrays, vectors and sets respectively.
- For any positive integers m and t with $t < m$, let $[0, m) = \{0, 1, \dots, m-1\}$ and $\binom{[0, m)}{t} = \{\mathcal{T} \mid \mathcal{T} \subseteq [0, m), |\mathcal{T}| = t\}$, i.e., $\binom{[0, m)}{t}$ is the collection of all t -sized subsets of $[0, m)$.
- Given an $F \times m$ matrix \mathbf{F} and a subset $\mathcal{T} \subseteq [0, m)$, $\mathbf{F}|_{\mathcal{T}}$ is a submatrix obtained by taking only the columns with subscript $j \in \mathcal{T}$ (i.e., deleting all the columns with subscript $j \in [0, m) \setminus \mathcal{T}$). Similarly, given a vector \mathbf{a} with length m , $\mathbf{a}|_{\mathcal{T}}$ is a vector with length $|\mathcal{T}|$ obtained by taking only the coordinates with subscript $j \in \mathcal{T}$.
- For any two vectors \mathbf{x} and \mathbf{y} with the same length, $d(\mathbf{x}, \mathbf{y})$ is the number of coordinates in which \mathbf{x} and \mathbf{y} differ. Let $wt(\mathbf{x})$ be the weight of \mathbf{x} , i.e., the number of nonzero coordinates of \mathbf{x} .

A. Placement delivery array

Yan et al. [14] proposed a combinatorial structure, called placement delivery array, which can characterize the placement phase and delivery phase simultaneously.

Definition 1: ([14]) For positive integers K , F , and S , an $F \times K$ array $\mathbf{P} = (p_{j,k})$, $j \in [0, F)$, $k \in [0, K)$, composed of a specific symbol “*” called star and S symbols $0, 1, \dots, S-1$, is called a (K, F, S) placement delivery array (PDA) if it satisfies the following condition C1:

- C1. For any two distinct entries p_{j_1, k_1} and p_{j_2, k_2} , $p_{j_1, k_1} = p_{j_2, k_2} = s$ is an integer only if
- $j_1 \neq j_2$, $k_1 \neq k_2$, i.e., they lie in distinct rows and distinct columns; and
 - $p_{j_1, k_2} = p_{j_2, k_1} = *$, i.e., the corresponding 2×2 subarray formed by rows j_1, j_2 and columns k_1, k_2 must be of the following form

$$\begin{pmatrix} s & * \\ * & s \end{pmatrix} \text{ or } \begin{pmatrix} * & s \\ s & * \end{pmatrix}.$$

In addition, for any positive integer $Z \leq F$, the array \mathbf{P} is denoted by (K, F, Z, S) PDA if the following condition is further satisfied:

- C2. each column has exactly Z stars.

For instance, it is easy to verify that the following array is a $(6, 4, 2, 4)$ PDA.

$$\mathbf{P} = \begin{pmatrix} * & * & * & 0 & 1 & 2 \\ * & 0 & 1 & * & * & 3 \\ 0 & * & 2 & * & 3 & * \\ 1 & 2 & * & 3 & * & * \end{pmatrix}. \quad (1)$$

In [14], Yan et al. showed that a PDA can generate a coded caching scheme as follows.

Theorem 1: ([14]) If there exists a (K, F, Z, S) PDA, then we can obtain an F -division (K, M, N) coded caching scheme with $\frac{M}{N} = \frac{Z}{F}$ and transmission load $R = \frac{S}{F}$.

In particular, Algorithm 1 was used to generate a coded caching scheme from a PDA. In fact, a PDA $\mathbf{P} = (p_{j,k})$ is an $F \times K$ array composed of a specific symbol “*” and S integers, where columns represent the user index and rows represent the packet index. If $p_{j,k} = *$, then user k has cached the j -th packet of all files, which satisfies identical uncoded caching policy. If $p_{j,k} = s$ is an integer, it means that the j -th packet of all files is not stored by user k . Then the server broadcasts a coded packet (i.e. the XOR of all the requested packets indicated by s) to the users at time slot s . The property C1 of PDA guarantees that each user can get its requested packet, since it has cached all the other packets in the coded packet except its requested one. The occurrence number of the integer s in \mathbf{P} , denoted by g_s , is the coded gain at time slot s , since the coded packet broadcasted at time slot s is useful for g_s users.

Example 1: Using the PDA in (1) and Algorithm 1, we can obtain a 4-division $(6, 3, 6)$ coded caching scheme as follows.

- **Placement Phase:** From Line 2 of Algorithm 1, we have $W_n = \{W_{n,0}, W_{n,1}, W_{n,2}, W_{n,3}\}$, $n \in [0, 6)$. Then by Lines 3-5, the contents in each user are

$$\begin{aligned} \mathcal{Z}_0 &= \{W_{n,0}, W_{n,1} \mid n \in [0, 6)\}, \mathcal{Z}_1 = \{W_{n,0}, W_{n,2} \mid n \in [0, 6)\}, \\ \mathcal{Z}_2 &= \{W_{n,0}, W_{n,3} \mid n \in [0, 6)\}, \mathcal{Z}_3 = \{W_{n,1}, W_{n,2} \mid n \in [0, 6)\}, \\ \mathcal{Z}_4 &= \{W_{n,1}, W_{n,3} \mid n \in [0, 6)\}, \mathcal{Z}_5 = \{W_{n,2}, W_{n,3} \mid n \in [0, 6)\}. \end{aligned}$$

Algorithm 1 Coded caching scheme based on PDA in [14]

```

1: procedure PLACEMENT( $\mathbf{P}, \mathcal{W}$ )
2:   Split each file  $W_n \in \mathcal{W}$  into  $F$  packets, i.e.,  $W_n = \{W_{n,j} \mid j \in [0, F)\}$ .
3:   for  $k \in \mathcal{K}$  do
4:      $\mathcal{Z}_k \leftarrow \{W_{n,j} \mid p_{j,k} = *, \forall n \in [0, N)\}$ 
5:   end for
6: end procedure
7: procedure DELIVERY( $\mathbf{P}, \mathcal{W}, \mathbf{d}$ )
8:   for  $s = 0, 1, \dots, S - 1$  do
9:     Server sends  $\bigoplus_{p_{j,k}=s, j \in [0, F), k \in [0, K)} W_{d_k, j}$ .
10:  end for
11: end procedure

```

- **Delivery Phase:** Assume that the request vector is $\mathbf{d} = (0, 1, 2, 3, 4, 5)$. By Lines 8-10, the signals sent by the server are listed in Table III.

TABLE III: Delivery steps in Example 1

Time Slot	Transmitted Signal
0	$W_{0,2} \oplus W_{1,1} \oplus W_{3,0}$
1	$W_{0,3} \oplus W_{2,1} \oplus W_{4,0}$
2	$W_{1,3} \oplus W_{2,2} \oplus W_{5,0}$
3	$W_{3,3} \oplus W_{4,2} \oplus W_{5,1}$

III. A FRAMEWORK FOR CONSTRUCTING PLACEMENT DELIVERY ARRAY

Motivated by several constructions of PDAs, we propose a framework for constructing PDAs as follows.

Construction 1: For any positive integers q, m and t with $0 < t < m$, let

$$\mathcal{F} \subseteq [0, q)^m, \quad \mathcal{K} \subseteq \binom{[0, m)}{t} \times [0, q)^t,$$

then an $|\mathcal{F}| \times |\mathcal{K}|$ array $\mathbf{P} = (p_{\mathbf{f}, \mathbf{k}})$ with $\mathbf{f} = (f_0, f_1, \dots, f_{m-1}) \in \mathcal{F}$ and $\mathbf{k} = (\mathcal{T}, \mathbf{b}) = (\{\delta_0, \delta_1, \dots, \delta_{t-1}\}, (b_0, b_1, \dots, b_{t-1})) \in \mathcal{K}$ with $0 \leq \delta_0 < \dots < \delta_{t-1} < m$, can be defined in the following way

$$p_{\mathbf{f}, \mathbf{k}} = \begin{cases} (\mathbf{e}, n_{\mathbf{e}}) & \text{if } d(\mathbf{f}|_{\mathcal{T}}, \mathbf{b}) = t, \\ * & \text{otherwise,} \end{cases} \quad (2)$$

where $\mathbf{e} = (e_0, e_1, \dots, e_{m-1}) \in [0, q)^m$ such that

$$e_i = \begin{cases} b_h & \text{if } i = \delta_h, h \in [0, t), \\ f_i & \text{otherwise.} \end{cases} \quad (3)$$

and $n_{\mathbf{e}}$ is the occurrence order of vector \mathbf{e} that occurs in column \mathbf{k} and starts from 0. For convenience, the matrix \mathbf{F} , whose collection of row vectors is \mathcal{F} , is called the row index matrix. The set \mathcal{K} is called the column index set.

We will show that the array constructed by Construction 1 is a PDA later. It is worth noting that in [16] the authors constructed a PDA, where $\mathcal{F} = [0, q)^m$, $\mathcal{K} = \binom{[0, m)}{t} \times [0, q)^t$ and the non-star entries are represented by $(m+t)$ -dimensional vectors. Clearly the difference between the construction in [16] and our construction lies in the representations of the row index matrix, column index set and definition of non-star entries. In our construction, the row index matrix and column index set are more flexible and the rule of non-star entries is simpler compared with the construction in [16]. Moreover, schemes with better performance can be obtained by Construction 1.

Example 2: Assume that $q = 2, m = 3$ and $t = 2$. Let $\mathcal{F} = \{(0, 0, 0), (1, 0, 1), (0, 1, 1), (1, 1, 0)\}$, i.e.,

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad (4)$$

$\mathcal{K} = \binom{[0,3]}{2} \times [0,2]^2$. Then we have $F = |\mathcal{F}| = 4$ and $K = |\mathcal{K}| = \binom{3}{2} \times 4 = 12$. From Construction 1, the following 4×12 array can be obtained.

f \ (T, b)	$\mathcal{T} = \{0, 1\}$				$\mathcal{T} = \{0, 2\}$				$\mathcal{T} = \{1, 2\}$			
	00	10	01	11	00	10	01	11	00	10	01	11
000	*	*	*	1100	*	*	*	1010	*	*	*	0110
101	*	*	0110	*	0000	*	*	*	*	1100	*	*
011	*	1010	*	*	*	1100	*	*	0000	*	*	*
110	0000	*	*	*	*	*	0110	*	*	*	1010	*

(5)

Let us consider the column represented by $(\{0, 1\}, (0, 0))$. From (3), we have $p_{(0,0,0),(\{0,1\},(0,0))} = *$ because $d((0, 0, 0)|_{\{0,1\}}, (0, 0)) = d((0, 0), (0, 0)) = 0$. Then $p_{(1,1,0),(\{0,1\},(0,0))} = (\mathbf{e}, n_{\mathbf{e}})$ with $\mathbf{e} = (0, 0, 0)$ and $n_{\mathbf{e}} = 0$, since $d((1, 1, 0)|_{\{0,1\}}, (0, 0)) = d((1, 1), (0, 0)) = 2$ and it is the first occurrence of \mathbf{e} in column $(\{0, 1\}, (0, 0))$. It is easy to check that the above array is a $(12, 4, 3, 4)$ PDA.

If $(\mathbf{e}, n_{\mathbf{e}})$ occurs in \mathbf{P} , we say \mathbf{e} occurs in \mathbf{P} . For each vector \mathbf{e} occurring in \mathbf{P} , if the number of its occurrences in each column is at most once, we always omit the occurrence order $n_{\mathbf{e}}$. For instance, the PDA in (5) can be written as

f \ (T, b)	$\mathcal{T} = \{0, 1\}$				$\mathcal{T} = \{0, 2\}$				$\mathcal{T} = \{1, 2\}$			
	00	10	01	11	00	10	01	11	00	10	01	11
000	*	*	*	110	*	*	*	101	*	*	*	011
101	*	*	011	*	000	*	*	*	*	110	*	*
011	*	101	*	*	*	110	*	*	000	*	*	*
110	000	*	*	*	*	*	011	*	*	*	101	*

Next we will prove that the array \mathbf{P} generated by Construction 1 is indeed a (K, F, S) PDA. We will show that \mathbf{P} satisfies condition C1 of Definition 1.

Proposition 1: Let \mathbf{P} be the array generated by Construction 1. If there are two distinct entries being the same vector $(\mathbf{e}, n_{\mathbf{e}})$, say $p_{\mathbf{f},(\mathcal{T},\mathbf{b})} = p_{\mathbf{f}',(\mathcal{T}',\mathbf{b}')} = (\mathbf{e}, n_{\mathbf{e}})$, then the following two statements must hold.

- 1) $\mathcal{T} \neq \mathcal{T}'$ and $\mathbf{f} \neq \mathbf{f}'$, which implies that they lie in different columns and different rows;
- 2) The subarray formed by rows \mathbf{f}, \mathbf{f}' and columns $(\mathcal{T}, \mathbf{b}), (\mathcal{T}', \mathbf{b}')$ must be of the following form

$$\begin{pmatrix} (\mathbf{e}, n_{\mathbf{e}}) & * \\ * & (\mathbf{e}, n_{\mathbf{e}}) \end{pmatrix} \text{ or } \begin{pmatrix} * & (\mathbf{e}, n_{\mathbf{e}}) \\ (\mathbf{e}, n_{\mathbf{e}}) & * \end{pmatrix}.$$

Proof. Clearly, each vector entry $(\mathbf{e}, n_{\mathbf{e}}) = (e_0, e_1, \dots, e_{m-1}, n_{\mathbf{e}})$ occurs at most once in each column by (2), i.e., $(\mathcal{T}, \mathbf{b}) \neq (\mathcal{T}', \mathbf{b}')$. Let

$$\begin{aligned} \mathbf{f} &= (f_0, f_1, \dots, f_{m-1}), & \mathbf{f}' &= (f'_0, f'_1, \dots, f'_{m-1}), \\ \mathcal{T} &= \{\delta_0, \delta_1, \dots, \delta_{t-1}\}, & \mathcal{T}' &= \{\delta'_0, \delta'_1, \dots, \delta'_{t-1}\}, \\ \mathbf{b} &= (b_0, b_1, \dots, b_{t-1}), & \mathbf{b}' &= (b'_0, b'_1, \dots, b'_{t-1}). \end{aligned}$$

If $\mathcal{T} = \mathcal{T}'$, then $\mathbf{b} \neq \mathbf{b}'$ holds. Without loss of generality, assume that $b_0 \neq b'_0$. By (3) we have $e_{\delta_0} = b_0 = b'_0$, which contradicts to the hypothesis. So we have $\mathcal{T} \neq \mathcal{T}'$, which implies that there must exist two distinct integers, say $i, i' \in [0, m)$, satisfying

$$i \in \mathcal{T}, i \notin \mathcal{T}' \quad \text{and} \quad i' \in \mathcal{T}', i' \notin \mathcal{T}.$$

Without loss of generality, assume that $i = \delta_0 \in \mathcal{T}$ and $i' = \delta'_0 \in \mathcal{T}'$. From Construction 1, we have

$$e_{\delta_0} = b_0 = f'_{\delta_0}, \quad \text{and} \quad e_{\delta'_0} = b'_0 = f_{\delta'_0}.$$

Hence $p_{\mathbf{f},(\mathcal{T}',\mathbf{b}')} = p_{\mathbf{f}',(\mathcal{T},\mathbf{b})} = *$ because both $d(\mathbf{f}|_{\mathcal{T}'}, \mathbf{b}') < t$ and $d(\mathbf{f}'|_{\mathcal{T}}, \mathbf{b}) < t$ hold. This implies $\mathbf{f} \neq \mathbf{f}'$. The proof is complete. \square

Remark 1: The array \mathbf{P} generated by Construction 1 is a (K, F, S) PDA with $K = |\mathcal{K}|$ and $F = |\mathcal{F}|$ by Proposition 1. Furthermore, if the number of stars in each column is the same, such as Z , then \mathbf{P} is a (K, F, Z, S) PDA. It is worth noting that the number of occurrences of each vector $(\mathbf{e}, n_{\mathbf{e}})$ in \mathbf{P} , i.e. the coded gain, is at most $\binom{m}{t}$ by Proposition 1.

Under the framework of Construction 1, we only need to choose a row index matrix and a column index set appropriately in order to design a coded caching scheme. Next a new scheme by Construction 1 will be presented, which can be regarded as a generalization of the MN scheme [1] and the PDA constructed in [16].

Theorem 2: For any positive integers m, s, t and ω with $\omega \leq t \leq s$ and $s+t-2\omega \leq m$, there exists a $(\binom{t}{\omega} \binom{m}{t}, \binom{m}{s}, \binom{m}{s} - \binom{m-t}{s-\omega}, \binom{m}{s+t-2\omega})$ PDA which can realize a $(\binom{t}{\omega} \binom{m}{t}, M, N)$ coded caching scheme with $\frac{M}{N} = 1 - \binom{m-t}{s-\omega} / \binom{m}{s}$, subpacketization $F = \binom{m}{s}$ and transmission load $R = \binom{m}{s+t-2\omega} / \binom{m}{s}$.

Proof. Let $q = 2$. For positive integers m, s, ω and t with $\omega \leq t \leq s$ and $s + t - 2\omega \leq m$, let \mathcal{F} be a collection of all m -vectors with weight s , i.e.,

$$\mathcal{F} = \{\mathbf{f} \mid wt(\mathbf{f}) = s, \mathbf{f} \in [0, 2]^m\}. \quad (6)$$

Let

$$\mathcal{K} = \mathfrak{T} \times \mathcal{B}, \quad \mathfrak{T} = \binom{[0, m]}{t}, \quad \mathcal{B} = \{\mathbf{b} \mid wt(\mathbf{b}) = t - \omega, \mathbf{b} \in [0, 2]^t\}. \quad (7)$$

Then $K = |\mathcal{K}| = \binom{t}{\omega}$ and $F = |\mathcal{F}| = \binom{m}{s}$. From Construction 1, an $\binom{m}{s} \times \binom{t}{\omega} \binom{m}{t}$ array \mathbf{P} is obtained. For each column $(\mathcal{T}, \mathbf{b}) \in \mathcal{K}$, if there exists a row $\mathbf{f} \in \mathcal{F}$ satisfying that $p_{\mathbf{f}, (\mathcal{T}, \mathbf{b})}$ is a non-star entry, then $wt(\mathbf{f}|_{\mathcal{T}}) = w$ and $wt(\mathbf{f}|_{[0, m] \setminus \mathcal{T}}) = s - w$ by (6), (7) and Construction 1. Moreover, $\mathbf{f}|_{\mathcal{T}}$ is determined by $(\mathcal{T}, \mathbf{b})$ since $q = 2$. So the number of non-star entries in each column is $\binom{m-t}{s-\omega}$. Then there are $\binom{m}{s} - \binom{m-t}{s-\omega}$ stars in each column, i.e., $Z = \binom{m}{s} - \binom{m-t}{s-\omega}$ and $\frac{M}{N} = 1 - \binom{m-t}{s-\omega} / \binom{m}{s}$.

For each vector \mathbf{e} occurring in \mathbf{P} , the number of its occurrences in each column is at most once. Otherwise if \mathbf{e} occurs in column $(\mathcal{T}, \mathbf{e}|_{\mathcal{T}})$ at least twice, say $P_{(\mathbf{f}, (\mathcal{T}, \mathbf{e}|_{\mathcal{T}}))} = P_{(\mathbf{f}', (\mathcal{T}, \mathbf{e}|_{\mathcal{T}}))} = (\mathbf{e}, n_{\mathbf{e}})$. Let $\mathbf{e} = (e_0, e_1, \dots, e_{m-1})$, $\mathbf{f} = (f_0, f_1, \dots, f_{m-1})$ and $\mathbf{f}' = (f'_0, f'_1, \dots, f'_{m-1})$, we have

$$\begin{aligned} f_j &= f'_j = e_j & j \notin \mathcal{T}, \\ f_{j'} &= f'_{j'} \neq e_{j'} & j' \in \mathcal{T}. \end{aligned}$$

This implies $\mathbf{f} = \mathbf{f}'$, which contradicts to the hypothesis. So the vector $(\mathbf{e}, n_{\mathbf{e}})$ can be written as \mathbf{e} and we only need to count the number of different vectors \mathbf{e} occurring in \mathbf{P} . By Construction 1 and equations (6), (7), each vector occurring in \mathbf{P} has weight $s + t - 2\omega$. In fact, each vector \mathbf{e} with wight $s + t - 2\omega$ must occur in \mathbf{P} . Without loss of generality, let us consider the vector $\mathbf{e} = (e_0, e_1, \dots, e_{m-1})$ such that $e_j = 1$ for $j \in [0, s + t - 2\omega]$ and $e_j = 0$ for $j \in [s + t - 2\omega, m)$. Clearly $wt(\mathbf{e}) = s + t - 2\omega$. Then the vector $\mathbf{f} = (f_0, f_1, \dots, f_{m-1})$ with

$$f_j = \begin{cases} 1 & \text{if } j \in [0, s - \omega) \cup [s + t - 2\omega, s + t - \omega) \\ 0 & \text{otherwise} \end{cases}$$

belongs to \mathcal{F} by (6) and the element $(\mathcal{T}, \mathbf{b})$ with $\mathcal{T} = [s - \omega, s - \omega + t - 1)$, $\mathbf{b} = (b_0, b_1, \dots, b_{t-1})$ and

$$b_j = \begin{cases} 1 & \text{if } j \in [0, t - \omega) \\ 0 & \text{otherwise} \end{cases}$$

belongs to \mathcal{K} by (7) since $wt(\mathbf{f}) = s$, $|\mathcal{T}| = t$ and $wt(\mathbf{b}) = t - \omega$ hold. Furthermore, we have $p_{\mathbf{f}, (\mathcal{T}, \mathbf{b})} = \mathbf{e}$ by (2). As a result, there are $\binom{m}{s+t-2\omega}$ different vectors occurring in \mathbf{P} . So we have $S = \binom{m}{s+t-2\omega}$ and the transmission load is $R = \frac{S}{F} = \binom{m}{s+t-2\omega} / \binom{m}{s}$. The proof is complete. \square

Example 3: When $m = 4$, $s = t = 2$ and $\omega = 1$, from Theorem 2 we have $K = \binom{t}{\omega} \binom{m}{t} = 12$, $F = \binom{m}{s} = 6$, $Z = \binom{m}{s} - \binom{m-t}{s-\omega} = 4$ and $S = \binom{m}{s+t-2\omega} = 6$. The following array, which is generated by Construction 1 using \mathcal{F} in (6) and \mathcal{K} in (7), is a $(12, 6, 4, 6)$ PDA. Here we omit the occurrence order $n_{\mathbf{e}}$ since each vector \mathbf{e} occurs at most once in each column.

\mathbf{f}	$(\mathcal{T}, \mathbf{b})$		$\mathcal{T} = \{0, 1\}$		$\mathcal{T} = \{0, 2\}$		$\mathcal{T} = \{1, 2\}$		$\mathcal{T} = \{0, 3\}$		$\mathcal{T} = \{1, 3\}$		$\mathcal{T} = \{2, 3\}$	
	10	01	10	01	10	01	10	01	10	01	10	01	10	01
1100	*	*	*	0110	*	1010	*	0101	*	1001	*	*	*	*
1010	*	0110	*	*	1100	*	*	0011	*	*	*	*	1001	*
1001	*	0101	*	0011	*	*	*	*	*	1100	*	1010	*	*
0110	1010	*	1100	*	*	*	*	*	*	*	0011	*	0101	*
0101	1001	*	*	*	*	0011	1100	*	*	*	*	0110	*	*
0011	*	*	1001	*	0101	*	1010	*	0110	*	*	*	*	*

- Remark 2:**
- When $\omega = 0$, from Theorem 2, we have a $(\binom{m}{t}, \binom{m}{s}, \binom{m}{s} - \binom{m-t}{s}, \binom{m}{s+t})$ PDA, which is exactly the first PDA constructed in [16]. When $t = 1$, $K = m$, $\omega = 0$ and $s = KM/N$, from Theorem 2 we have a $(K, \binom{K}{KM/N}, \binom{K}{KM/N} - \binom{K-1}{KM/N-1}, \binom{K}{KM/N+1})$ PDA, which realizes the MN scheme in [1].
 - When $\omega \neq 0$, the scheme from Theorem 2 may have better performance than that of the scheme when $\omega = 0$. For example, in Table IV, compared with Scheme 2 with $\omega = 0$, Scheme 1 with $\omega = 2$ has more users, less memory size and subpacketization but has larger coded gain; Compared with Scheme 3 with $\omega = 0$, Scheme 4 with $\omega = 1$ has more users, the same memory size and subpacketization but has larger coded gain.

TABLE IV: Comparisons of the schemes with $\omega = 0$ and $\omega > 0$

Schemes	K	$\frac{M}{N}$	F	R	coded gain	ω
Scheme 1	360	0.9	210	0.57143	63	2
Scheme 2	120	0.91667	252	0.17857	56	0
Scheme 3	120	0.83333	210	0.57143	35	0
Scheme 4	360	0.83333	210	1.2	50	1

IV. NECESSARY CONDITIONS FOR THE ROW INDEX MATRIX

In this section, we consider the case that the column index set is $\mathcal{K} = \binom{[0, m]}{t} \times [0, q]^t$ which implies $K = \binom{m}{t} q^t$. The following definition will be used in the rest of this paper.

Definition 2 ([24]): Let \mathbf{F} be an $F \times m$ matrix over $[0, q]$ for positive integers $F, m, q \geq 2$, and $s \leq m$.

- \mathbf{F} is an orthogonal array (OA) with strength s , denoted by $\text{OA}_\lambda(F, m, q, s)$, if every $1 \times s$ row vector appears exactly λ times in $\mathbf{F}|_{\mathcal{S}}$ for each $\mathcal{S} \in \binom{[0, m]}{s}$.
- \mathbf{F} is a covering array (CA) with strength s , denoted by $\text{CA}_\lambda(F, m, q, s)$, if every $1 \times s$ row vector appears at least λ times in $\mathbf{F}|_{\mathcal{S}}$ for each $\mathcal{S} \in \binom{[0, m]}{s}$.

It is well known that $F = \lambda q^s$ for any $\text{OA}_\lambda(F, m, q, s)$ and thus $\text{OA}_\lambda(F, m, q, s)$ is sometimes written as $\text{OA}_\lambda(m, q, s)$ for short [24]. The parameter λ is the index of the orthogonal array. If λ is omitted, then it is understood to be 1. Similarly, the index of the covering array can be omitted when it is 1. Clearly for any $\text{CA}(F, m, q, s)$, we must have $F \geq q^s$ because any $\text{OA}(m, q, s)$ is also a $\text{CA}(q^s, m, q, s)$.

When $m = 3, q = 3$ and $s = 2$, let us consider the array in (4). For each $\mathcal{S} = \{0, 1\}, \{0, 2\}$ and $\{1, 2\}$, we have

$$\mathbf{F}|_{\{0,1\}} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \quad \mathbf{F}|_{\{0,2\}} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \mathbf{F}|_{\{1,2\}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$$

Clearly \mathbf{F} in (4) satisfies Definition 2. So \mathbf{F} is an $\text{OA}(3, 2, 2)$ and it is also a $\text{CA}(4, 3, 2, 2)$.

Next the necessary condition for the row index matrix will be discussed when each user has the same memory size or the coded gain is the largest.

A. Necessary conditions for the row index matrix when each user has the same memory size

Theorem 3: For any positive integers q, m and t with $t \leq m$, let $K = \binom{m}{t} q^t$. Then the array \mathbf{P} generated by Construction 1 is a (K, F, Z, S) PDA if and only if the row index matrix \mathbf{F} is an $\text{OA}_\lambda(m, q, t)$ with $\lambda = \frac{F-Z}{(q-1)^t}$, i.e., every $1 \times t$ row vector appears exactly λ times in $\mathbf{F}|_{\mathcal{T}}$ for any t -subset $\mathcal{T} \subseteq [0, m]$.

Proof. First we assume that \mathbf{P} is a (K, F, Z, S) PDA with $K = \binom{m}{t} q^t$. By Construction 1, the column index set is $\mathcal{K} = \binom{[0, m]}{t} \times [0, q]^t$ and the row index matrix is an $F \times m$ matrix \mathbf{F} . For each $\mathcal{T} \in \binom{[0, m]}{t}$ and for each $\mathbf{b} \in [0, q]^t$, let $h_{\mathcal{T}, \mathbf{b}}$ denote the number of occurrences of \mathbf{b} in $\mathbf{F}|_{\mathcal{T}}$.

When $t = 1$, for each $\mathcal{T} \in \binom{[0, m]}{1}$, we build a binary table (see Table V), where the entry at row $\{\mathcal{T}, \mathbf{b}\}$ and column $\mathbf{f}|_{\mathcal{T}}$ is 0 if $p_{\mathbf{f}, (\mathcal{T}, \mathbf{b})} = *$, otherwise it is 1. According to Table V, a $q \times q$ binary matrix Φ_1 is defined in (8).

TABLE V: The binary table for $t = 1$

$\{\mathcal{T}, \mathbf{b}\} \backslash \mathbf{f} _{\mathcal{T}}$	(0)	(1)	...	($q-1$)
$(\mathcal{T}, (0))$	0	1	...	1
$(\mathcal{T}, (1))$	1	0	...	1
\vdots	\vdots	\vdots	\ddots	\vdots
$(\mathcal{T}, (q-1))$	1	1	...	0

$$\Phi_1 = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{pmatrix} \quad (8)$$

When $t > 1$, Table VI can be obtained similarly, where the row and column labels are arranged in lexicographic order. Then a $q^t \times q^t$ binary matrix Φ_t is recursively defined in (9).

TABLE VI: The binary table for any $t > 1$

$\mathbf{f} \mathcal{T}$	$(0, b_1, \dots, b_{t-1}),$ $b_h \in [0, q), h \in [1, t)$	$(1, b_1, \dots, b_{t-1}),$ $b_h \in [0, q), h \in [1, t)$	\dots	$(q-1, b_1, \dots, b_{t-1}),$ $b_h \in [0, q), h \in [1, t)$
$(\mathcal{T}, (0, b_1, \dots, b_{t-1})),$ $b_h \in [0, q), h \in [1, t)$	0	Φ_{t-1}	\dots	Φ_{t-1}
$(\mathcal{T}, (1, b_1, \dots, b_{t-1})),$ $b_h \in [0, q), h \in [1, t)$	Φ_{t-1}	0	\dots	Φ_{t-1}
\vdots	\vdots	\vdots	\ddots	\vdots
$(\mathcal{T}, (q-1, b_1, \dots, b_{t-1})),$ $b_h \in [0, q), h \in [1, t)$	Φ_{t-1}	Φ_{t-1}	\dots	0

$$\Phi_t = \begin{pmatrix} 0 & \Phi_{t-1} & \dots & \Phi_{t-1} \\ \Phi_{t-1} & 0 & \dots & \Phi_{t-1} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{t-1} & \Phi_{t-1} & \dots & 0 \end{pmatrix} \quad (9)$$

Let

$$\mathbf{h}_{\mathcal{T},t} = (h_{\mathcal{T},(0,0,\dots,0)}, h_{\mathcal{T},(0,0,\dots,1)}, \dots, h_{\mathcal{T},(q-1,q-1,\dots,q-1)})^\top.$$

Since the number of non-star entries in each column of \mathbf{P} is $F - Z$, from Table VI, we have

$$\Phi_t \mathbf{h}_{\mathcal{T},t} = (F - Z) \cdot \mathbf{1}_{q^t \times 1}, \quad (10)$$

where $\mathbf{1}_{q^t \times 1}$ is the $q^t \times 1$ vector with all entries equal to 1. Obviously, Φ_1 in (8) is a circulant matrix and it is easy to check that it is invertible. Using elementary row operations the determinant of Φ_t can be obtained as $|\Phi_t| = (q-1)(-1)^{q-1}|\Phi_{t-1}|^q$. Hence Φ_t is also invertible for $t > 1$. Furthermore, the number of 1 in each row of Φ_t is $(q-1)^t$. Consequently the unique solution of (10) is

$$h_{\mathcal{T},(0,0,\dots,0)} = h_{\mathcal{T},(0,0,\dots,1)} = \dots = h_{\mathcal{T},(q-1,q-1,\dots,q-1)} = \frac{F - Z}{(q-1)^t}.$$

So the row index matrix \mathbf{F} is an $\text{OA}_\lambda(m, q, t)$ with $\lambda = \frac{F-Z}{(q-1)^t}$.

Conversely, we assume that the row index matrix \mathbf{F} is an $\text{OA}_\lambda(m, q, t)$. Then we have $F = \lambda q^t$ and there are $\lambda(q-1)^t$ non-star entries in each column of \mathbf{P} by Construction 1. This implies that there are $\lambda q^t - \lambda(q-1)^t$ stars in each column, i.e. $Z = \lambda q^t - \lambda(q-1)^t$. So \mathbf{P} is a (K, F, Z, S) PDA with $K = \binom{m}{t} q^t$ and $\frac{Z}{F} = 1 - (\frac{q-1}{q})^t$ by Proposition 1. Hence the proof is complete. \square

From the above proof, the following result is obvious.

Remark 3: If the row index matrix \mathbf{F} is an $\text{OA}_\lambda(m, q, t)$ and the column index set $\mathcal{K} = \binom{[0,m)}{t} \times [0, q)^t$, the array \mathbf{P} generated by Construction 1 is a (K, F, Z, S) PDA with $K = \binom{m}{t} q^t$ and $\frac{Z}{F} = 1 - (\frac{q-1}{q})^t$.

B. Necessary conditions for the row index matrix when the coded gain is the maximum

From Remark 1, the number of occurrences (i.e. the coded gain) of each vector in the array \mathbf{P} is at most $\binom{m}{t}$, i.e., the largest coded gain is $\binom{m}{t}$. In practice, for the fixed number of users, subpacketization and memory ratio, we prefer the coded gain as large as possible.

Theorem 4: For any positive integers q, m and t with $t \leq m$, let the array \mathbf{P} generated by Construction 1 be a (K, F, Z, S) PDA with $K = \binom{m}{t} q^t$. If each vector $(\mathbf{e}, n_{\mathbf{e}})$ in \mathbf{P} occurs exactly $\binom{m}{t}$ times, i.e. the coded gain is $\binom{m}{t}$, then we have

- 1) the row index matrix \mathbf{F} is a $\text{CA}(F, m, q, m-t)$, i.e., every $1 \times (m-t)$ row vector appears at least once in $\mathbf{F}|_{\mathcal{S}}$ for any $(m-t)$ -subset $\mathcal{S} \subseteq [0, m)$;
- 2) the subpacketization $F \geq q^{m-t}$.

Proof. Since the array \mathbf{P} generated by Construction 1 is a (K, F, Z, S) PDA with $K = \binom{m}{t} q^t$, we have the column index set $\mathcal{K} = \binom{[0,m)}{t} \times [0, q)^t$ by Construction 1. Moreover, the row index matrix \mathbf{F} is an $\text{OA}_\lambda(m, q, t)$ by Theorem 3.

When $t \geq m-t$, the result holds because an $\text{OA}_\lambda(m, q, t)$ is also an $\text{OA}_{\lambda'}(m, q, m-t)$ with $\lambda' = \lambda q^{2t-m}$. When $t < m-t$, without loss of generality, let $\overline{\mathcal{T}} = \{0, 1, \dots, m-t-1\}$, we need to check that any $(m-t)$ -dimensional row vector, without loss of generality, say $(0, 0, \dots, 0)$, is a row vector of $\mathbf{F}|_{\overline{\mathcal{T}}}$.

Since \mathbf{F} is an $\text{OA}_\lambda(m, q, t)$, there exists a row vector $\mathbf{f}_0 = (f_{0,0}, f_{0,1}, \dots, f_{0,m-1}) = (0, \dots, 0, f_{0,t}, \dots, f_{0,m-1})$ in \mathcal{F} . Let i_0 be the smallest coordinate satisfying $f_{0,i_0} \neq 0$. If $i_0 \geq m-t$, then the result holds. Otherwise if $i_0 < m-t$, then we choose $\mathcal{T}_0 = \{i_0, m-t+1, \dots, m-1\}$ and $\mathbf{b}_0 = (0, b_{0,1}, \dots, b_{0,t-1})$ such that $b_{0,j} \neq f_{0,m-t+j}$ for $1 \leq j < t$. By Construction 1, there is a vector $\mathbf{e}_0 = (e_{0,0}, e_{0,1}, \dots, e_{0,m-1}) = (0, \dots, 0, f_{0,i_0+1}, \dots, f_{0,m-t}, b_{0,1}, \dots, b_{0,t-1})$ occurring at row \mathbf{f}_0 and column $(\mathcal{T}_0, \mathbf{b}_0)$.

Since the coded gain is $\binom{m}{t}$, each vector occurring in \mathbf{P} must appear in exactly $\binom{m}{t}$ columns. Let $\mathcal{T}_1 = \{m-t, m-t+1, \dots, m-1\}$. Then \mathbf{e}_0 must appear at column $(\mathcal{T}_1, \mathbf{e}_0|_{\mathcal{T}_1})$ and row

$$\mathbf{f}_1 = (f_{1,0}, f_{1,1}, \dots, f_{1,m-1}) = (0, \dots, 0, e_{0,i_0+1}, \dots, e_{0,m-t-1}, f_{1,m-t}, \dots, f_{1,m-1})$$

with $f_{1,j} \neq e_{0,j}$ for $j \in [m-t, m)$. Let i_1 be the smallest coordinate satisfying $f_{1,i_1} \neq 0$, i.e., $\mathbf{f}_1 = (0, \dots, 0, f_{1,i_1}, \dots, f_{1,m-1})$, clearly $i_1 > i_0$. If $i_1 \geq m-t$, the result holds. Otherwise if $i_1 < m-t$, we can repeat the same process until we obtain a row $\mathbf{f}_u = (0, \dots, 0, f_{u,i_u}, \dots, f_{u,m-1})$ with $f_{u,i_u} \neq 0$ and $i_u \geq m-t$. This shows that $(0, 0, \dots, 0)$ is a row vector of $\mathbf{F}|_{\overline{\mathcal{T}}}$. Hence the row index matrix \mathbf{F} is a $\text{CA}(F, m, q, m-t)$, which implies $F \geq q^{m-t}$. The proof is complete. \square

In summary, the following result is obtained.

Theorem 5: For any positive integers q, m and t with $t \leq m$, let the array \mathbf{P} generated by Construction 1 be a (K, F, Z, S) PDA with $K = \binom{m}{t} q^t$. The coded caching scheme realized by \mathbf{P} has memory ratio $\frac{Z}{F} = 1 - \left(\frac{q-1}{q}\right)^t$ and transmission load $R \geq (q-1)^t$. Furthermore, if the transmission load achieves the lower bound, i.e. $R = (q-1)^t$, then the subpacketization $F \geq q^{m-t}$.

Proof. From Theorem 3 and Remark 3, the coded caching scheme realized by \mathbf{P} has the memory ratio $\frac{Z}{F} = 1 - \left(\frac{q-1}{q}\right)^t$. From Remark 1, each vector $(\mathbf{e}, n_{\mathbf{e}})$ in \mathbf{P} occurs at most $\binom{m}{t}$ times, which implies that the coded gain is at most $\binom{m}{t}$. Hence $S \geq \frac{K(F-Z)}{\binom{m}{t}} = F(q-1)^t$. Consequently we have $R = \frac{S}{F} \geq (q-1)^t$. If the transmission load achieves the lower bound, i.e., $R = (q-1)^t$, then the coded gain is $\binom{m}{t}$. Hence we have $F \geq q^{m-t}$ from Theorem 4. \square

V. NEW SCHEMES VIA ORTHOGONAL ARRAYS

In this section, we demonstrate our framework by presenting two more explicit new schemes that use different OAs as the row index matrix \mathbf{F} in Construction 1. These new schemes have significant advantages in subpacketization compared with the schemes in [16]. Again, the column index set \mathcal{K} is the set $\binom{[0, m]}{t} \times [0, q]^t$.

A. New scheme via $\text{OA}(m, q, m-1)$

In this subsection, we use a trivial orthogonal array $\text{OA}(m, q, m-1)$ as the row index matrix \mathbf{F} . Then the array \mathbf{P} generated by Construction 1 has the following properties.

Lemma 1: For any positive integers m, t with $t \leq m$ and $q \geq 2$, let the column index set \mathcal{K} be $\binom{[0, m]}{t} \times [0, q]^t$ and the row index matrix \mathbf{F} be an $\text{OA}(m, q, m-1)$. For each vector $(\mathbf{e}, n_{\mathbf{e}})$ in the array \mathbf{P} generated by Construction 1, we have

- 1) \mathbf{e} occurs in exactly $\binom{m}{t}$ columns;
- 2) The number of times in each column that \mathbf{e} occurs is the same.

Proof. Recall that if $(\mathbf{e}, n_{\mathbf{e}})$ occurs in \mathbf{P} , we say \mathbf{e} occurs in \mathbf{P} . For any vector $\mathbf{e} = (e_0, \dots, e_{m-1})$ occurring in \mathbf{P} , we assume that $p_{\mathbf{f}, (\mathcal{T}, \mathbf{e}|_{\mathcal{T}})} = (\mathbf{e}, n_{\mathbf{e}})$ with $\mathbf{f} = (f_0, \dots, f_{m-1})$. For any fixed column $(\mathcal{T}, \mathbf{e}|_{\mathcal{T}})$, the collection of rows in which \mathbf{e} occurs is denoted by $\mathcal{F}_{\mathcal{T}, \mathbf{e}}$, i.e.,

$$\mathcal{F}_{\mathcal{T}, \mathbf{e}} = \{\mathbf{f} \in \mathcal{F} \mid f_i = e_i \text{ for } i \in [0, m) \setminus \mathcal{T} \text{ and } f_i \neq e_i \text{ for } i \in \mathcal{T}\}$$

For any $\mathcal{T}' \in \binom{[0, m]}{t}$ with $|\mathcal{T} \cap \mathcal{T}'| = t-1$, assume that $\mathcal{T} \setminus \mathcal{T}' = \{w\}$ and $\mathcal{T}' \setminus \mathcal{T} = \{w'\}$. Clearly, we have $w \neq w'$. For the fixed column $(\mathcal{T}', \mathbf{e}|_{\mathcal{T}'})$, the collection of rows in which \mathbf{e} occurs is

$$\mathcal{F}_{\mathcal{T}', \mathbf{e}} = \{\mathbf{f} \in \mathcal{F} \mid f_i = e_i \text{ for } i \in [0, m) \setminus \mathcal{T}' \text{ and } f_i \neq e_i \text{ for } i \in \mathcal{T}'\}$$

Since \mathbf{F} is an $\text{OA}(m, q, m-1)$, we can define a mapping ψ from $\mathcal{F}_{\mathcal{T}, \mathbf{e}}$ to $\mathcal{F}_{\mathcal{T}', \mathbf{e}}$ as follows

$$\psi(\mathbf{f}) = \mathbf{f}', \quad \mathbf{f} \in \mathcal{F}_{\mathcal{T}, \mathbf{e}}, \quad \mathbf{f}' \in \mathcal{F}_{\mathcal{T}', \mathbf{e}},$$

where $f'_i = f_i$ for $i \in [0, m) \setminus \{w, w'\}$ and $f'_w = e_w$. Since every row vector in $[0, q)^{m-1}$ occurs exactly once in submatrix $\mathbf{F}|_{[0, m) \setminus \{w'\}}$, there is only one row vector $\mathbf{f}' \in \mathcal{F}$ corresponding to \mathbf{f} . Furthermore, we can show that $f'_{w'} \neq e_{w'}$, which implies $\mathbf{f}' \in \mathcal{F}_{\mathcal{T}', \mathbf{e}}$. Otherwise if $f'_{w'} = e_{w'}$, then the $(m-1)$ -dimensional vector $\mathbf{f}|_{[0, m) \setminus \{w\}}$ occurs at least twice in $\mathbf{F}|_{[0, m) \setminus \{w\}}$, which is contradictory to that \mathbf{F} is an $\text{OA}(m, q, m-1)$. So ψ is an injective mapping from $\mathcal{F}_{\mathcal{T}, \mathbf{e}}$ to $\mathcal{F}_{\mathcal{T}', \mathbf{e}}$. Then we have $|\mathcal{F}_{\mathcal{T}, \mathbf{e}}| \leq |\mathcal{F}_{\mathcal{T}', \mathbf{e}}|$. Similar to the above mapping ψ , we can also define another injective mapping from $\mathcal{F}_{\mathcal{T}', \mathbf{e}}$ to $\mathcal{F}_{\mathcal{T}, \mathbf{e}}$ as follows

$$\phi(\mathbf{f}') = \mathbf{f}, \quad \mathbf{f}' \in \mathcal{F}_{\mathcal{T}', \mathbf{e}}, \quad \mathbf{f} \in \mathcal{F}_{\mathcal{T}, \mathbf{e}},$$

where $f_i = f'_i$ for $i \in [0, m) \setminus \{w, w'\}$ and $f_{w'} = e_{w'}$. Similarly, we have $|\mathcal{F}_{\mathcal{T}, \mathbf{e}}| \geq |\mathcal{F}_{\mathcal{T}', \mathbf{e}}|$. Therefore $|\mathcal{F}_{\mathcal{T}, \mathbf{e}}| = |\mathcal{F}_{\mathcal{T}', \mathbf{e}}|$, which implies that \mathbf{e} occurs the same number of times in columns $(\mathcal{T}, \mathbf{e}|_{\mathcal{T}})$ and $(\mathcal{T}', \mathbf{e}|_{\mathcal{T}'})$, respectively. For any $\mathcal{T}'' \in \binom{[0, m)}{t}$, we can find a sequence $\mathcal{T}_0, \dots, \mathcal{T}_k \in \binom{[0, m)}{t}$ such that $\mathcal{T}_0 = \mathcal{T}$, $\mathcal{T}_k = \mathcal{T}''$, and $|\mathcal{T}_i \cap \mathcal{T}_{i+1}| = t-1$ for $0 \leq i \leq k-1$. Hence \mathbf{e} occurs the same number of times in columns $(\mathcal{T}, \mathbf{e}|_{\mathcal{T}})$ and $(\mathcal{T}'', \mathbf{e}|_{\mathcal{T}''})$ respectively. The proof is complete. \square

From Lemma 1, a new scheme can be obtained as follows.

Theorem 6: For any positive integers m, t with $t < m$, and $q \geq 2$, there always exists an $((\binom{m}{t}q^t, q^{m-1}, q^{m-1} - (q-1)^t q^{m-t-1}, (q-1)^t q^{m-1})$ PDA, which can realize a $((\binom{m}{t}q^t, M, N)$ coded caching scheme with $\frac{M}{N} = 1 - (\frac{q-1}{q})^t$, subpacketization $F = q^{m-1}$ and transmission load $R = (q-1)^t$.

Proof. Let the column index set $\mathcal{K} = \binom{[0, m)}{t} \times [0, q)^t$ and

$$\mathcal{F} = \{(f_0, f_1, \dots, f_{m-2}, \sum_{i=0}^{m-2} f_i) \mid f_0, f_1, \dots, f_{m-2} \in \mathbb{Z}_q\}. \quad (11)$$

Then $F = q^{m-1}$. It is easy to check that the corresponding row index matrix \mathbf{F} , whose collection of row vectors is \mathcal{F} , is an $\text{OA}(m, q, m-1)$, which is also an $\text{OA}_\lambda(m, q, t)$ with $\lambda = q^{m-1-t}$. By Remark 3, the array \mathbf{P} generated by Construction 1 is a (K, F, Z, S) PDA with memory ratio $\frac{M}{N} = \frac{Z}{F} = 1 - (\frac{q-1}{q})^t$. Furthermore, by Lemma 1, each vector $(\mathbf{e}, n_{\mathbf{e}})$ in \mathbf{P} occurs exactly $\binom{m}{t}$ times. Hence the coded gain is $\binom{m}{t}$. Consequently we obtain $S = \frac{K(F-Z)}{\binom{m}{t}} = F(q-1)^t$ and thus the transmission load is $R = \frac{S}{F} = (q-1)^t$. \square

Remark 4: • When $t = 1$, the scheme in Theorem 6 is exactly the one from [14], which is listed in the third row of Table I. Moreover, it achieves both the lower bounds on transmission load R and subpacketization F in Theorem 5.

- For any positive $1 < t < m$, the scheme in Theorem 6 has the same number of users $K = \binom{m}{t}q^t$, memory ratio $\frac{M}{N} = 1 - (\frac{q-1}{q})^t$ and transmission load $R = (q-1)^t$ as the scheme from [16], which is listed in the fourth row of Table I, while its subpacketization is just $\frac{1}{q}$ times as that of the scheme from [16].
- Let the column index set $\mathcal{K} = \binom{[0, m)}{t} \times [0, q)^t$ and $\mathcal{F} = [0, q)^m$. It is easy to check that the array generated by Construction 1 is a $((\binom{m}{t}q^t, q^m, q^m - (q-1)^t q^{m-t}, (q-1)^t q^m)$ PDA, which is exactly the second PDA in [16].

Remark 5: Given a PDA, clearly its subarray is also a PDA. However it is worth noting that it is not easy to find out a rule of deletion such that the resulting subarray of the PDA in [16] has the same performance as the PDA in Theorem 6. In fact the performance of a PDA depends heavily on how the entries are defined. For example, when $t = 1$, the scheme in [14] has much smaller subpacketization than that of the scheme in [16] for the same number of users, memory size and transmission load. Indeed, the construction of the scheme with $K = mq$ in [14] has to use two formulas, i.e., the first is for the users in $[0, (m-1)q)$ (i.e., formula (37) in [14]) and the second is for the users in $[(m-1)q, mq)$ (i.e., the formula (38) in [14]). In contrast, our construction provides a simple rule to define these entries, i.e., (3) and (2) in Construction 1.

B. New scheme via $\text{OA}(m, q, m-t)$

For any positive integers m and s with $s \leq m$ and for any prime power q , the following results will be used in this subsection.

- Let q be a prime power and \mathbb{F}_q^m denote the vector space of all m -tuples over the finite field \mathbb{F}_q . A vector set \mathcal{C} is called an $[m, s]_q$ linear code over \mathbb{F}_q if \mathcal{C} is a s -dimensional subspace of \mathbb{F}_q^m . Each vector in \mathcal{C} is called a codeword. An $[m, s]_q$ linear code \mathcal{C} has q^s codewords.
- The minimum distance of a linear code is the smallest distance between distinct codewords. An $[m, s, d]_q$ code is an $[m, s]_q$ linear code \mathcal{C} with minimum distance d . If q is omitted, then it is understood to be 2. An $[m, s, d]_q$ code is called an $[m, s]_q$ maximum distance separable (MDS) code if $d = m - s + 1$. For each codeword $\mathbf{c} \in \mathcal{C}$, $\mathcal{E}_{\mathcal{C}}(\mathbf{c}, m-s)$ is the sphere with radius $m-s$ and centre \mathbf{c} , i.e., $\mathcal{E}_{\mathcal{C}}(\mathbf{c}, m-s) = \{\mathbf{e} \in \mathbb{F}_q^m \mid d(\mathbf{e}, \mathbf{c}) \leq m-s\}$.

Lemma 2: For any positive integers m and s with $s \leq m$ and for any prime power q , let \mathcal{C} be an $[m, s]_q$ MDS code, then

$$\bigcup_{\mathbf{c} \in \mathcal{C}} \mathcal{E}_{\mathcal{C}}(\mathbf{c}, m-s) = \mathbb{F}_q^m.$$

A key property of an $[m, s]_q$ MDS code is that each codeword in \mathcal{C} is determined by its any s coordinates [30]. Based on this property, given an $[m, s]_q$ MDS code \mathcal{C} , an $\text{OA}(m, q, s)$ can be obtained by taking the codewords in \mathcal{C} as its row vectors [24].

For any positive integers m and t with $t < m$, there exists an $[m, m-t]_q$ MDS code \mathcal{C} for some prime power q . Let $\mathcal{F} = \mathcal{C}$ and the column index set $\mathcal{K} = \binom{[0, m]}{t} \times [0, q)^t$, then we have $F = q^{m-t}$ and the corresponding row index matrix \mathbf{F} , whose collection of row vectors is \mathcal{C} , is an $\text{OA}(m, q, m-t)$. If $m-t \geq t$, i.e., $t \leq \frac{m}{2}$, the row index matrix \mathbf{F} is also an $\text{OA}_\lambda(m, q, t)$ with $\lambda = q^{m-2t}$. By Remark 3, the array \mathbf{P} generated from Construction 1 is a (K, F, Z, S) PDA with $K = \binom{m}{t} q^t$ and $Z/F = 1 - (\frac{q-1}{q})^t$. Furthermore, we obtain the following result.

Theorem 7: Let m, t be positive integers with $m \geq 2t$ and $q \geq 2$ be a prime power. If there exists an $[m, m-t]_q$ MDS code \mathcal{C} , then there always exists an $(\binom{m}{t} q^t, q^{m-t}, q^{m-t} - (q-1)^t q^{m-2t}, q^m - q^{m-t})$ PDA which can realize a $(\binom{m}{t} q^t, M, N)$ coded caching scheme with $\frac{M}{N} = 1 - (\frac{q-1}{q})^t$, subpacketization $F = q^{m-t}$ and transmission load $R = q^t - 1$.

Proof. Let $\mathcal{F} = \mathcal{C}$ and $\mathcal{K} = \binom{[0, m]}{t} \times [0, q)^t$. From the above analysis, the array \mathbf{P} generated from Construction 1 is a (K, F, Z, S) PDA with $K = \binom{m}{t} q^t$, $F = q^{m-t}$ and $Z/F = 1 - (\frac{q-1}{q})^t$. The rest is to prove $R = q^t - 1$, i.e., $S = FR = q^m - q^{m-t}$.

First we prove that each vector in $\mathbb{F}_q^m \setminus \mathcal{C}$ appears in \mathbf{P} . By Construction 1, a vector \mathbf{e} appears in \mathbf{P} if and only if there exists a vector $\mathbf{f} \in \mathcal{C}$ satisfying $d(\mathbf{e}, \mathbf{f}) = t$. Since \mathcal{C} is an $[m, m-t]_q$ MDS code, by Lemma 2, we have $\bigcup_{\mathbf{f} \in \mathcal{C}} \mathcal{E}_{\mathcal{C}}(\mathbf{f}, t) = \mathbb{F}_q^m$. Hence, for any vector $\mathbf{e} \in \mathbb{F}_q^m \setminus \mathcal{C}$, there exists a vector $\mathbf{f} \in \mathcal{C}$ satisfying $0 < d(\mathbf{e}, \mathbf{f}) \leq t$. If $d(\mathbf{e}, \mathbf{f}) = t$, then \mathbf{e} appears in \mathbf{P} . If $0 < d(\mathbf{e}, \mathbf{f}) < t$, there exists a vector \mathbf{e}' satisfying 1) $d(\mathbf{e}', \mathbf{f}) = t$ and 2) \mathbf{e} is located on the line generated by \mathbf{f} and \mathbf{e}' . Consequently we have $\mathbf{e} = \alpha \mathbf{f} + \beta \mathbf{e}'$ for some $\alpha, \beta \in \mathbb{F}_q \setminus \{0\}$. Then we have $\mathbf{e} = (\alpha + \beta) \mathbf{f} + \beta(\mathbf{e}' - \mathbf{f})$, which implies $d(\mathbf{e}, (\alpha + \beta) \mathbf{f}) = wt(\beta(\mathbf{e}' - \mathbf{f})) = t$. Since $(\alpha + \beta) \mathbf{f} \in \mathcal{C}$, \mathbf{e} appears in \mathbf{P} .

Next, for each vector \mathbf{e} occurring in \mathbf{P} , we show that the number of its occurrences in each column is at most once. Assume that \mathbf{e} appears at $p_{\mathbf{f}_1, (\mathcal{T}, \mathbf{b})}$ and $p_{\mathbf{f}_2, (\mathcal{T}, \mathbf{b})}$, then we have $\mathbf{f}_1|_{[0, m] \setminus \mathcal{T}} = \mathbf{e}|_{[0, m] \setminus \mathcal{T}} = \mathbf{f}_2|_{[0, m] \setminus \mathcal{T}}$ by Construction 1. Consequently, we have $\mathbf{f}_1 = \mathbf{f}_2$, because \mathbf{f}_1 and \mathbf{f}_2 belong to the $[m, m-t]_q$ MDS code \mathcal{C} .

Therefore $S = q^m - q^{m-t}$. The proof is complete. \square

From Theorem 6 we have an $(\binom{m}{t} q^t, M, N)$ coded caching scheme, say Scheme1, with memory ratio $\frac{M}{N} = 1 - (\frac{q-1}{q})^t$, subpacketization $F_1 = q^{m-1}$ and transmission load $R_1 = (q-1)^t$. From Theorem 7 we have an $(\binom{m}{t} q^t, M, N)$ coded caching scheme, say Scheme2, with memory ratio $\frac{M}{N} = 1 - (\frac{q-1}{q})^t$, subpacketization $F_2 = q^{m-t}$ and transmission load $R_2 = q^t - 1$. Clearly Scheme1 has the same number of users K and memory ratio $\frac{M}{N}$ as Scheme2. Moreover, $\frac{F_1}{F_2} = q^{t-1}$ and $\frac{R_1}{R_2} = \frac{(q-1)^t}{q^t - 1}$. If q is large enough, then $\frac{R_1}{R_2}$ approximates 1 and $\frac{F_1}{F_2}$ approximates infinity. This implies that Scheme 2 has much smaller subpacketization and almost the same transmission load compared to Scheme1 for large q . For example, we compare them in Table VII when $t = 2$ and $m = 10, 20, 30$ and 40 . From Table VII, we can see that for fixed t , the memory ratio $\frac{M}{N}$ decreases as q increases. Furthermore, the value of $\frac{R_1}{R_2}$ approximates 1 and the value of $\frac{F_1}{F_2}$ increases fast as q increases.

TABLE VII: The comparisons of schemes from Theorem 6 and Theorem 7

m	q	K	$\frac{M}{N}$	Rate ratio $\frac{R_1}{R_2}$	subpacketization ratio $\frac{F_1}{F_2}$
10	11	5445	0.1736	0.9091	11
20	23	100510	0.0930	0.9167	23
30	31	418035	0.0635	0.9375	31
40	41	1311180	0.0482	0.9524	41

VI. CONCLUSION

In this paper, we introduced a framework to construct PDA and thus coded caching schemes. Based on this framework, we only need to choose the row index matrix and the column index set appropriately in order to design a coded caching scheme. We demonstrated our approach by providing a generalization of several previously known schemes constructed in [1], [16]. When the number of users is $K = \binom{m}{t} q^t$ for any positive integers m, q and t with $t \leq m$, we showed that the row index matrix must be an orthogonal array if each user has the same memory size. On the other hand, when the coded gain is $\binom{m}{t}$, which is the largest coded gain under this framework, the row index matrix must be a covering array. Consequently lower bounds on the transmission load R and the subpacketization F were derived. In particular, we showed that the scheme in [14] achieves these two lower bounds. Furthermore, We demonstrated our approach by producing new explicit schemes with smaller subpacketization than some previously known schemes. Using a trivial orthogonal array as the row index matrix, we constructed a new scheme, which has the same number of users, memory ratio and transmission load as the scheme in [16],

while its subpacketization is smaller. Similarly, based on a MDS code, we constructed another new scheme with the same number of users, memory ratio and almost the same transmission load (under certain conditions) as the scheme in [16], while its subpacketization is much smaller. Finally we remark that all the row vectors of \mathcal{F} considered in this paper are all distinct. However, we can also consider the case where some row vectors are repeated using our construction.

REFERENCES

- [1] M. A. Maddah-Ali and U. Niesen, Fundamental limits of caching, *IEEE Trans. Inform. Theory*, vol. 60, no. 5, pp. 2856-2867, 2014.
- [2] M. Ji, G. Caire, and A. F. Molisch, Fundamental limits of caching in wireless D2D networks, *IEEE Trans. Inform. Theory*, vol. 62, no. 2, pp. 849-869, 2016.
- [3] N. Karamchandani, U. Niesen, M. A. Maddah-Ali, and S. Diggavi, Hierarchical coded caching, in Proc. *IEEE ISIT*, Honolulu, HI, Jun. 2014, pp. 2142-2146.
- [4] A. Sengupta, R. Tandon, T. C. Clancy, Fundamental limits of caching with secure delivery, *IEEE Trans. Inf. Forensics and Security*, vol. 10, no. 2, pp.355-370, 2015.
- [5] M. M. Amiri and D. Gündüz, Fundamental limits of caching: Improved delivery rate-cache capacity trade-off, *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 806-815, 2016.
- [6] H. Ghasemi and A. Ramamoorthy, Improved lower bounds for coded caching, in Proc. *IEEE ISIT*, Hong Kong, Jun. 2015, pp. 1696-1700.
- [7] K. Wan, D. Tuninetti, and P. Piantanida, On the optimality of uncoded cache placement, in Proc. *IEEE ITW*, Cambridge, UK, Sept. 2016.
- [8] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, The exact rate-memory tradeoff for caching with uncoded prefetching, *IEEE Trans. Inform. Theory*, vol. 64, no. 2, pp. 1281-1296, 2018.
- [9] S. Jin, Y. Cui, H. Liu, and G. Caire, Uncoded placement optimization for coded delivery, in Proc. *IEEE WiOpt*, Shanghai, China, May. 2018.
- [10] A. Sengupta, R. Tandon, and T. C. Clancy, Improved approximation of storage-rate tradeoff for caching via new outer bounds, in Proc. *IEEE ISIT*, Hong Kong, Jun. 2015, pp. 1691-1695.
- [11] C. Tian and J. Chen, Caching and delivery via interference elimination, in Proc. *IEEE ISIT*, Barcelona, July 2016, pp. 830-834.
- [12] K. Shanmugam, M. Ji, A. M. Tulino, J. Llorca, and A. G. Dimakis, Finite-length analysis of caching-aided coded multicasting, *IEEE Trans. Inform. Theory*, vol. 62, no. 10, pp. 5524-5537, 2016.
- [13] M. Cheng, J. Jiang, X. Tang, and Q. Yan, Some variant of known coded caching schemes with good performance, *IEEE Trans. Commun.*, vol. 68, no.3, pp. 1370-1377, Mar., 2020.
- [14] Q. Yan, M. Cheng, X. Tang, and Q. Chen, On the placement delivery array design in centralized coded caching scheme, *IEEE Trans. Inform. Theory*, vol. 63, no. 9, pp. 5821-5833, 2017.
- [15] L. Tang, A. Ramamoorthy, Coded caching schemes with reduced subpacketization from linear block codes, *IEEE Trans. Inform. Theory*, vol. 64, no. 4, pp. 3099-3120, 2018.
- [16] C. Shanguan, Y. Zhang, and G. Ge, Centralized coded caching schemes: A hypergraph theoretical approach, *IEEE Trans. Inform. Theory*, vol. 64, no. 8, pp. 5755-5766, 2018.
- [17] K. Shanmugam, A. M. Tulino, and A. G. Dimakis, Coded caching with linear subpacketization is possible using Ruzsa-Szemerédi graphs, in Proc. *IEEE ISIT*, Aachen, Germany, Jun. 2017, pp. 1237-1241.
- [18] Q. Yan, X. Tang, Q. Chen, and M. Cheng, Placement delivery array design through strong edge coloring of bipartite graphs, *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 236-239, 2018.
- [19] P. Krishnan, Coded caching via line graphs of bipartite graphs, in Proc. *IEEE ITW*, Guangzhou, Nov. 2018.
- [20] S. Agrawal, K. V. S. Sree, P. Krishnan, Coded caching based on combinatorial designs, in Proc. *IEEE ISIT*, Paris, 7-12 July 2019, pp. 1227-1231.
- [21] M. Cheng, J. Jiang, Q. Wang, Y. Yao, A generalized grouping scheme in coded caching, *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3422-3430, May 2019.
- [22] M. Cheng, J. Jiang, Q. Yan, X. Tang, Coded caching schemes for flexible memory sizes, *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4166-4176, June. 2019.
- [23] K. Shanmugam, A. G. Dimakis, J. Llorca, and A. M. Tulino, A unified Ruzsa-Szemerédi framework for finite-length coded caching, in Proc. *The 51st ACSSC*, Pacific Grove, CA, 2017, pp. 631-635.
- [24] D.R. Stinson, *Combinatorial Designs: Construction and Analysis*, Springer, 2003, New York.
- [25] Q. Yan, S. Yang, and M. Wigger, Storage, computation, and communication: A fundamental tradeoff in distributed computing, in Proc. *IEEE ITW*, Guangzhou, 25-29 Nov. 2018.
- [26] Q. Yan, M. Wigger, S. Yang, and X. Tang, A fundamental storage-communication tradeoff in distributed computing with straggling nodes, *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7311 - 7327, June. 2020.
- [27] M. Cheng, D. Liang, and R. Wei, On secure coded caching via combinatorial method, arXiv:2005.01043v1 [cs.IT] 3 May 2020.
- [28] V R Aravind, P. Sarvepalli, and A. Thangaraj, Coded caching with demand privacy: Constructions for lower subpacketization and generalizations, arXiv:2007.07475v1 [cs.IT] 15 Jul 2020.
- [29] Q. Yan, and D. Tuninetti, Key superposition simultaneously achieves security and privacy in cache-aided linear function retrieval, arXiv:2009.06000v1 [cs.IT] 13 Sep 2020.
- [30] J. H. van Lint, *Introduction to Coding Theory*, third version, Springer, 1999, Printed in Germany.