IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 19, 2020, accepted December 1, 2020, date of publication December 7, 2020, date of current version December 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042849

# Improving Placement Delivery Array Coded Caching Schemes With Coded Placement

ZHANG MINGMING [1], CHENG MINQUAN [1], WANG JINYU [1,2], ZHONG XI [1], AND YISHAN CHEN [3]

[1]Guangxi Key Lab of Multi-Source Information Mining & Security, Guangxi Normal University, Guilin 541004, China
[2]College of Mathematics and Statistics, Guangxi Normal University, Guilin 541004, China
[3]Business College, Guilin Tourism University, Guilin 541004, China

Corresponding author: Cheng Minquan (chengqinshi@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62061004, in part by the Guangxi Higher Institutions Program of Introducing 100 High-Level Overseas Talents, Research Fund of the Guangxi Key Lab of Multi-Source Information Mining & Security under Grant 16-B-01, in part by the Guangxi Scientific Research Project of the Higher Education Department under Grant RZ2000000777, in part by the Guangxi Collaborative Innovation Center of Multi-Source Information Integration and Intelligent Processing, in part by the Guangxi Bagui Scholar Teams for Innovation and Research Project, and in part by the Guangxi Talent Highland Project of Big Data Intelligence and Application.

**ABSTRACT** Coded caching schemes with low subpacketization and small transmission rate are desirable in practice due to the requirement of low implementation complexity and efficiency of transmission. Placement delivery arrays (PDA in short) can be used to generate coded caching schemes. However, many known coded caching schemes, which have low subpacketizations, realized by PDAs do not fully use the users' caching content to create multicasting opportunities. So we propose a method to overcome this drawback. As an application, we obtain a new scheme, which has significantly advantages on the tradeoff between memory ratio and transmission rate.

**INDEX TERMS** Coded caching scheme, placement delivery array, memory ratio, subpacketization.

## I. INTRODUCTION

The wireless networks have been imposed tremendous pressure on the data transmission during the peak traffic times due to the explosive increasing of mobile services, especially the video streaming. A coded caching scheme, which has been recognized as an efficient solution to reduce this tremendous pressure, was proposed in [1] and has been rapidly used to in various settings such as cache-aided combination networks [2], D2D networks [3], hierarchical networks [4], insecure channel [5], among others.

### A. SYSTEMS MODEL

In a centralized $(K, M, N)$ coded 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, $N \geq K$. Denote the $N$ files by $\mathcal{W} = \{W_0, \ldots, W_{N-1}\}$ and $K$ users by $\mathcal{K} = \{0, \ldots, K-1\}$. 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 packets, i.e., $W_i = \{W_{i,j} : j = 0, 1, \ldots, F-1\}$. 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.
- **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 $\mathbf{d} = (d_0, \cdots, d_{K-1})$, i.e., user $k$ requests the $d_k$-th file $W_{d_k}$, where $d_k \in \{0, 1, \ldots, N-1\}$ and $k \in \mathcal{K}$. The server broadcasts a coded signal (XOR of some required packets) of size $S_{\mathbf{d}}$ packets to users such that each user is able to recover its requested file with the help of its caching contents.

In this article, we focus on the worst-case scenario, i.e., all the users require different files. In this case, the transmission rate of a coded caching scheme is defined as the maximal

The associate editor coordinating the review of this manuscript and approving it for publication was Xijun Wang.

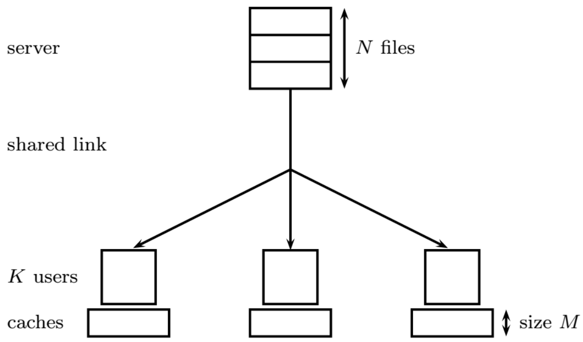


**FIGURE 1.** ($K, M, N$) coded caching system.

normalized transmission amount among all the requests in the delivery phase, i.e.

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

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

### B. PRIOR WORK

Maddah-Ali and Niesen [1] introduced the first deterministic $F$-division $(K, M, N)$ coded caching scheme with $F = \binom{K}{KM/N}$ when $KM/N$ is an integer. Obviously, the subpacketization $F = \binom{K}{KM/N}$ increases rapidly as $K$ increases, which makes this scheme impractical when $K$ is large. It is well known that there exists a tradeoff between the transmission rate and the subpacketization for a fixed number of users and a fixed memory ratio. Indeed, there are many research papers focus on constructing the coded caching schemes with lower subpacketization while sacrificing some transmission rate, such as [6]–[14], [17], etc.

The first scheme with lower subpacketization compared with the MN scheme was proposed by Shanmugam *et al.* in [15] by grouping method. Yan *et al.* [6] proposed an $F \times K$ combinatoric structure, which is called a placement delivery array (PDA), to generate an $F$-division $(K, M, N)$ coded caching scheme. Consequently, the MN scheme is equivalent a special PDA. Furthermore, for any positive integer $q > 1$, the schemes with $\frac{M}{N} = \frac{1}{q}$ and $1 - \frac{1}{q}$ respectively were obtained by constructing PDAs. It is worth noting that the $F$ and $R$ of the scheme with $\frac{M}{N} = \frac{1}{q}$ achieves the tradeoff proposed in [15]. It implies that this scheme maybe the best and the tradeoff is tight for some parameters. There are many other constructions focusing on further reducing the subpacketization by increasing the transmission rate, such as by the special (6, 3)-free hypergraphs [12], the resolvable combinatorial design and linear block codes [14], the $(r, t)$ Ruzsa-Szeméredi graphs [7], [13], the strong edge coloring of bipartite graphs [9], projective space [16], etc. In [7], Shanmugam *et al.* discovered that all deterministic $F$-division coded caching schemes can be recasted into a PDA when $K \leq N$. By means of PDA, Cheng *et al.* [17] showed that the MN PDA has the minimum subpacketization with minimum transmission rate among PDAs. There are also many other results on PDA such as [8], [9], [17]. Very recently there are several schemes with low subpacketization such as [10], [11], which can be represented by PDAs.

In fact, a scheme realized by a PDA has uncoded placement phase, which implies that any two packets cached by a user are independent. As a result, many cached packets are not fully used in some well known coded caching schemes realized by PDAs, i.e., do not generate multicasting opportunities in the delivery phase. Intuitively, for a scheme realized by a PDA, when the subpacketization $F$ is reduced, this drawback always exists.

### C. CONTRIBUTIONS AND ORGANIZATIONS

In this article, the main idea of the proposed scheme is as follows. On the observation that some cached contents of each user are not leveraged in the existing PDA schemes, we propose to use Minimum Distance Separable (MDS) code in the cache placement phase such that each user does not cache these unused cached contents. As a result, the needed memory ratio is reduced. A similar idea to reduce the cache redundancy by using MDS coded cache placement was originally proposed in [2] for the cache-aided combination network problem.

In order to verify the effectiveness of our method, we take the strong edge coloring based scheme in [9] as an example, and obtain a new scheme, which has smaller subpacketization and memory ratio than the original scheme in [9]. Simulations show that under the same memory ratio, the new scheme has smaller transmission rate and more flexible memory ratio than the original scheme in [9].

The rest of this article is organized as follows. The relationship between a PDA and a coded caching scheme is explained in Section II. In Section III we introduce our research motivation and present the main method. As an application, we construct a new scheme in Section IV. Finally, we conclude the paper in Section V.

## II. CODED CACHING SCHEMES REALIZED BY PDAs

In this article, we use the following notations unless otherwise stated. We use bold capital letters and curlicue letters to denote arrays and sets respectively. For any positive integers $m$ and $t$ with $t < m$, let $[0, m) = \{0, \ldots, m-1\}$ 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 $[0, m)$.

Next we will give the definition of a PDA, and demonstrate the relationship between a PDA and a coded caching scheme.

*Definition 1 [6]: For positive integers $K, F, Z$ 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 "$*$" and $S$ integers in $[0, S)$, is called a $(K, F, Z, S)$ placement delivery array (PDA) if it satisfies the following conditions:*

C1. *The symbol "$*$" appears $Z$ times in each column;*
C2. *Each integer in $[0, S)$ occurs at least once in the array;*

C3. *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*

a. $j_1 \neq j_2$, $k_1 \neq k_2$, *i.e., they lie in distinct rows and distinct columns; and*

b. $p_{j_1,k_2} = p_{j_2,k_1} = *$, *i.e., the corresponding* $2 \times 2$ *subarray formed by rows* $j_1, j_2$ *and columns* $k_1, k_2$ *must be of the following form*

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

*Theorem 1 [6]: Using Algorithm 1, an* $F$*-division* $(K, M, N)$ *coded caching scheme with memory ratio* $\frac{M}{N} = \frac{Z}{F}$ *and transmission rate* $R = \frac{S}{F}$ *can be realized by a* $(K, F, Z, S)$ *PDA.*

---

**Algorithm 1** Caching Scheme Based on PDA in [6]

---

1: **procedure** PLACEMENT($\mathbf{P}$, $\mathcal{W}$)
2: Split each file $W_i \in \mathcal{W}$ 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($\mathbf{P}$, $\mathcal{W}$, $d$)
8: **for** $s = 0, 1, \cdots, S-1$ **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**

---

*Example 1: It is easy to verify that the following array is a* $(6, 6, 2, 12)$ *PDA.*

$$\mathbf{P} = \begin{pmatrix} * & 0 & 1 & 2 & 3 & * \\ 0 & * & 4 & 5 & * & 6 \\ 1 & 4 & * & * & 7 & 8 \\ 2 & 5 & * & * & 9 & 10 \\ 3 & * & 7 & 9 & * & 11 \\ * & 6 & 8 & 10 & 11 & * \end{pmatrix} \quad (1)$$

*Using Algorithm 1, we can obtain a* 6*-division* $(6, 2, 6)$ *coded caching scheme in the following way.*

- ***Placement Phase**: From Line 2 in Algorithm 1, we have* $W_n = \{W_{n,0}, W_{n,1}, W_{n,2}, W_{n,3}, W_{n,4}, W_{n,5}\}$, $n \in [0, 6)$. *Then by Lines 3-5, the contents cached by users are*

$$\mathcal{Z}_0 = \mathcal{Z}_5 = \{W_{n,0}, W_{n,5} \mid n \in [0, 6)\},$$
$$\mathcal{Z}_1 = \mathcal{Z}_4 = \{W_{n,1}, W_{n,4} \mid n \in [0, 6)\},$$
$$\mathcal{Z}_2 = \mathcal{Z}_3 = \{W_{n,2}, W_{n,3} \mid n \in [0, 6)\}.$$

- ***Delivery Phase**: Assume that the request vector is* $\mathbf{d} = (0, 1, 2, 3, 4, 5)$. *By Lines 8-10, the server sends the coded signals in Table 1 at time slots* $0 - 11$.

*Remark 1: In a* $(K, F, Z, S)$ *PDA* $\mathbf{P} = (p_{j,k})$, $j \in [0, F)$, $k \in [0, K)$, *each column represents one user's caching contents, i.e., if* $p_{j,k} = *$, *then user* $k$ *has cached the* $j$*-th packet of all the files. Clearly, it belongs to uncoded placement. The property C1 of Definition 1 implies that all the users have the same memory size and the memory ratio is* $\frac{M}{N} = \frac{Z}{F}$. *If* $p_{j,k} = s$ *is an integer, it means that the* $j$*-th packet of all the files is not stored by user* $k$. *Then the XOR of the requested packets indicated by* $s$ *is broadcasted by the server at time slot* $s$. *The property C2 of Definition 1 implies that the number of signals transmitted by the server is exactly* $S$, *so the transmission rate is* $R = \frac{S}{F}$. *And the property C3 of Definition 1 guarantees that each user can get the requested packet, since it has cached all the other packets in the signal except its requested one.*

**TABLE 1.** Delivery steps in Example 1.

| Time Slot | Transmitted Signal |
|---|---|
| 0 | $W_{0,1} \oplus W_{1,0}$ |
| 1 | $W_{0,2} \oplus W_{2,0}$ |
| 2 | $W_{0,3} \oplus W_{3,0}$ |
| 3 | $W_{0,4} \oplus W_{4,0}$ |
| 4 | $W_{1,2} \oplus W_{2,1}$ |
| 5 | $W_{1,3} \oplus W_{3,1}$ |
| 6 | $W_{1,5} \oplus W_{5,1}$ |
| 7 | $W_{2,4} \oplus W_{4,2}$ |
| 8 | $W_{2,5} \oplus W_{5,2}$ |
| 9 | $W_{3,4} \oplus W_{4,3}$ |
| 10 | $W_{3,5} \oplus W_{5,3}$ |
| 11 | $W_{4,5} \oplus W_{5,4}$ |

## III. NEW SCHEMES FROM PDAs

In this section, based on an appropriate PDA, an improved scheme with coded placement is proposed.

### A. RESEARCH MOTIVATIONS

Let $\mathbf{P}$ be a $(K, F, Z, S)$ PDA. For any integer $s \in [0, S)$, assume that the occurrence number of $s$ is $r_s$, say $p_{j_u,k_u} = s$, $u \in [0, r_s)$, $j_u \in [0, F)$ and $k_u \in [0, K)$. Consider the subarray formed by rows $j_0, \cdots, j_{r_s-1}$ and columns $k_0, \cdots, k_{r_s-1}$, which is of order $r_s \times r_s$ since we have $j_u \neq j_v$ and $k_u \neq k_v$ for all $0 \leq u \neq v < r_s$ from the property C3 of Definition 1. Furthermore, we have $p_{j_u,k_v} = *$ for all $0 \leq u \neq v < r_s$ from the property C3. Then this subarray is equivalent to the following $r_s \times r_s$ array

$$\mathbf{P}^{(s)} = \begin{pmatrix} s & * & \cdots & * & * \\ * & s & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & \cdots & s & * \\ * & * & \cdots & * & s \end{pmatrix} \quad (2)$$

with respect to row/column permutation. According to Line 9 in Algorithm 1, the signal $\bigoplus_{u\in[0,r_s)} W_{d_{k_u},j_u}$ is transmitted by the server at time slot $s$, which is simultaneously useful for $r_s$ users. So the occurrence number $r_s$ is called the coded caching gain at time slot $s$. Clearly we prefer to design a scheme with the coded caching gain at each time slot as large as possible. Furthermore, the stars in $\mathbf{P}^{(s)}$ guarantee that each user can recover its requested packet from the transmitted signal, so they are useful for generating the coded caching gain. In other words, for each star in $\mathbf{P}^{(s)}$, say $p_{j_u,k_v} = *$ where $0 \le u \neq v < r_s$, the packet $W_{d_{k_u},j_u}$, which is cached by user $k_v$ and simultaneously requested by user $k_u$, is helpful for user $k_v$ to recover its requested packet $W_{d_{k_v},j_v}$, so it generates a multicasting opportunity. Therefore, for a star entry $p_{j,k} = *$, we call it useful if it occurs in $\mathbf{P}^{(s)}$ for some integer $s \in [0, S)$, otherwise we call it useless.

In fact, given a PDA, if there exist some useless stars, we can delete these useless stars and then further reduce the subpacketization and the memory ratio without reducing the coded caching gain at each time slot. Now let us see the array $\mathbf{P}$ in (1) again. It is easy to check that the stars at $p_{j,5-j}$ $(j \in [0, 6))$ of $\mathbf{P}$ are useless. We delete all the useless stars and obtain the following array.

$$\mathbf{P}' = \left(\begin{array}{c|c|c|c|c|c} * & 0 & 1 & 2 & 3 & \\ \hline 0 & * & 4 & 5 & & 6 \\ \hline 1 & 4 & * & & 7 & 8 \\ \hline 2 & 5 & & * & 9 & 10 \\ \hline 3 & & 7 & 9 & * & 11 \\ \hline & 6 & 8 & 10 & 11 & * \end{array}\right) \quad (3)$$

According to $\mathbf{P}'$ in (3), we can modify the scheme in Example 1 as follows.

- **Placement Phase**: Each file $W_n$, $n \in [0, 6)$, is divided into 5 packets, say $(W_{n,0}, W_{n,1}, W_{n,2}, W_{n,3}, W_{n,4})$. Let

$$W_{n,5} = W_{n,0} + W_{n,1} + W_{n,2} + W_{n,3} + W_{n,4}.$$

Using the caching strategy in Lines 3-5 in Algorithm 1, all the users cache the following contents.

$$\begin{aligned} \mathcal{Z}_0 &= \{W_{n,0} \mid n \in [0, 6)\} \\ \mathcal{Z}_1 &= \{W_{n,1} \mid n \in [0, 6)\} \\ \mathcal{Z}_2 &= \{W_{n,2} \mid n \in [0, 6)\} \\ \mathcal{Z}_3 &= \{W_{n,3} \mid n \in [0, 6)\} \\ \mathcal{Z}_4 &= \{W_{n,4} \mid n \in [0, 6)\} \\ \mathcal{Z}_5 &= \{W_{n,5} \mid n \in [0, 6)\} \end{aligned}$$

Clearly the memory ratio of each user is $\frac{1}{5}$, which is smaller than the memory ratio $\frac{2}{6}$ in Example 1.

- **Delivery Phase**: We also assume that the request vector is $\mathbf{d} = (0, 1, 2, 3, 4, 5)$. By Lines 8-10, the server also sends the coded signals listed in Table 1. Then each user can decode its requested file since each file $W_n$ can be recovered by any 5 packets out of $\{W_{n,j} | j \in [0, 6)\}$. The coded caching gain at each time slot is the same as that of the original scheme in Example 1. For instance, user 1 first decodes the required packets $W_{1,0}$, $W_{1,5}$, $W_{1,2}$ and $W_{1,3}$ from $W_{0,1} \oplus W_{1,0}$, $W_{1,5} \oplus W_{5,1}$, $W_{1,2} \oplus W_{2,1}$ and $W_{1,3} \oplus W_{3,1}$ respectively. Then it can get $W_{1,4} = W_{1,5} - W_{1,0} - W_{1,1} - W_{1,2} - W_{1,3}$.

### B. NEW SCHEMES

Following the proposal in subsection III-A, the following result can be obtained.

*Theorem 2: For any $(K, F, Z, S)$ PDA $\mathbf{P}$, if there exist $Z'$ useless stars in each column, then we can obtain an $(F - Z')$-division $(K, M, N)$ coded caching scheme with memory ratio $\frac{M}{N} = \frac{Z-Z'}{F-Z'}$ and transmission rate $R = \frac{S}{F-Z'}$, in which the coded caching gain at each time slot is the same as the original scheme realized by $\mathbf{P}$ and Algorithm 1.*

*Proof:* Assume that $\mathbf{P}$ is a $(K, F, Z, S)$ PDA where each column has $Z'$ useless stars. Deleting the $Z'$ useless stars in each column, we obtain a new array $\mathbf{P}' = (p'_{j,k})$, $j \in [0, F)$, $k \in [0, K)$. Clearly each column of $\mathbf{P}'$ has $Z'$ blanks, $Z - Z'$ stars and $F - Z$ integers.

Based on $\mathbf{P}'$, we modify the placement strategy in Algorithm 1 as follows: The server divides each file into $F - Z'$ equal-sized packets and then encodes them using an $(F, F - Z')$ maximum distance separable (MDS) code in an appropriate operation field [18]. Let the resulting encoded packets be denoted by $W_{n,0}, W_{n,1}, \ldots, W_{n,F-1}$ for each file $W_n$, $n \in [0, N)$. Using the caching strategy in Lines 3-5 in Algorithm 1, each user $k$ caches $\mathcal{Z}_k = \{W_{n,j} \mid p'_{j,k} = *, j \in [0, F), n \in [0, N)\}$. Clearly the memory ratio of each user is $\frac{M}{N} = \frac{Z-Z'}{F-Z'}$.

In the delivery phase, we also use the delivery strategy in Algorithm 1 as follows: For any request vector $d$, using Lines 7-11 of Algorithm 1, each user can get exactly $F - Z'$ required coded packets by the property C3 of Definition 1 and Remark 1. From the property of an $(F, F - Z')$ MDS code, each user can recover its requested file. So the transmission rate is $R = \frac{S}{F-Z'}$. Furthermore, the coded caching gain at each time slot is the same as that of the original scheme realized by $\mathbf{P}$ since the occurrence number of each integer is unchanged. □

Given an appropriate PDA $\mathbf{P}$, using Theorem 2, we can obtain a new scheme which has lower subpacketization and memory ratio while keeping the coded caching gain at each time slot unchanged.

## IV. NEW SCHEMES BASED ON THE PDAs IN [9]

In this section, we take the strong edge coloring based scheme in [9], [19] as an example. From the results on optimality in [20], one can check that the scheme in [9] has the smallest transmission rate among all the schemes with the same placement strategy. However, we will show that there are useless stars in the PDA corresponding to the scheme in [9], and use Theorem 2 to obtain an improved scheme.

### A. CONSTRUCTIONS FROM [9], [19]

For the readers' convenience, the construction of the PDA corresponding to the strong edge coloring based scheme in [19] is given here.

*Construction 1 ( [9], [19]): For any positive integers $H$, $b$, $r$, $\lambda$ satisfying $0 < r, b < H$, $\lambda < \min\{r, b\}$ and $r + b \le H + \lambda$, let $\mathcal{F} = \binom{[0,H)}{b}$, $\mathcal{K} = \binom{[0,H)}{r}$ and $\mathcal{I} = \binom{[0,H)}{\lambda}$. There exist*

- *an* $\left(\binom{H}{r}, \binom{H}{b}, \binom{H}{b} - \binom{r}{\lambda}\binom{H-r}{b-\lambda}, \binom{H}{r+b-2\lambda}\binom{H-r-b+2\lambda}{\lambda}\right)$ *PDA* $\mathbf{P} = (p_{B,A})$, $B \in \mathcal{F}$, $A \in \mathcal{K}$, *where*

$$p_{B,A} = \begin{cases} ((A \cup B) - I, I) & \text{if } A \cap B = I \in \mathcal{I} \\ * & \text{otherwise} \end{cases} \quad (4)$$

*and*

- *an* $\left(\binom{H}{r}, \binom{H}{b}, \binom{H}{b} - \binom{r}{\lambda}\binom{H-r}{b-\lambda}, \binom{H}{r+b-2\lambda}\binom{r+b-2\lambda}{r-\lambda}\right)$ *PDA* $\mathbf{P}' = (p'_{B,A})$, $B \in \mathcal{F}$, $A \in \mathcal{K}$, *where*

$$p'_{B,A} = \begin{cases} ((A \cup B) - I, A - B) & \text{if } A \cap B = I \in \mathcal{I} \\ * & \text{otherwise} \end{cases} \quad (5)$$

*Example 2: When* $H = 4$, $r = 2$, $b = 2$, $\lambda = 1$, *we have*

$$\mathcal{F} = \mathcal{K} = \{\{0, 1\}, \{0, 2\}, \{0, 3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\},$$
$$\mathcal{I} = \{\{0\}, \{1\}, \{2\}, \{3\}\}.$$

*The following array* $\mathbf{P}$ *can be obtained by* (4).

| | 01 | 02 | 03 | 12 | 13 | 23 |
|---|---|---|---|---|---|---|
| 01 | * | (12, 0) | (13, 0) | (02, 1) | (03, 1) | * |
| 02 | (12, 0) | * | (23, 0) | (01, 2) | * | (03, 2) |
| 03 | (13, 0) | (23, 0) | * | * | (01, 3) | (02, 3) |
| 12 | (02, 1) | (01, 2) | * | * | (23, 1) | (13, 2) |
| 13 | (03, 1) | * | (01, 3) | (23, 1) | * | (12, 3) |
| 23 | * | (03, 2) | (02, 3) | (13, 2) | (12, 3) | * |

(6)

*Here we represent each subset as a string for short. Replacing the entries* (12, 0), (13, 0), (02, 1), (03, 1), (23, 0), (01, 2), (03, 2), (01, 3), (02, 3), (23, 1), (13, 2), (12, 3) *of* $\mathbf{P}$ *in* (6) *by* $0, 1, \ldots, 11$ *respectively, the array in* (1) *is obtained.*

*Lemma 1 ( [9], [19]): For any positive integers* $H$, $b$, $r$, $\lambda$ *satisfying* $0 < r, b < H$, $\lambda < \min\{r, b\}$ *and* $r + b \leq H+\lambda$, *there exists an* $\binom{H}{b}$*-division* $(\binom{H}{r}, M, N)$ *coded caching scheme with memory ratio and transmission rate as follows*:

$$\frac{M}{N} = 1 - \frac{\binom{r}{\lambda}\binom{H-r}{b-\lambda}}{\binom{H}{b}}$$

$$R = \frac{\binom{H}{r+b-2\lambda}}{\binom{H}{b}} \min\left\{\binom{H-(r+b-2\lambda)}{\lambda}, \binom{r+b-2\lambda}{r-\lambda}\right\}$$

## B. NEW SCHEMES FROM THE PDA IN [9]

In the following we will show that the PDA in [9] satisfies Theorem 2 with $Z' > 1$ for some parameters and obtain the following result.

*Theorem 3: For any positive integers* $H$, $b$, $r$, $\lambda$ *satisfying* $0 < r, b < H$, $\lambda < \min\{r, b\}$ *and* $r + b \leq H + \lambda$,, *there exists an* $(\binom{H}{b} - \sum_{i=0}^{\lambda-1}\binom{r}{i}\binom{H-r}{b-i})$*-division* $(\binom{H}{r}, M, N)$ *coded caching scheme with memory ratio and transmission rate as follows*:

$$\frac{M}{N} = 1 - \frac{\binom{r}{\lambda}\binom{H-r}{b-\lambda}}{\binom{H}{b} - \sum_{i=0}^{\lambda-1}\binom{r}{i}\binom{H-r}{b-i}}$$

$$R = \frac{\binom{H}{r+b-2\lambda}}{\binom{H}{b} - \sum_{i=0}^{\lambda-1}\binom{r}{i}\binom{H-r}{b-i}}\binom{H-(r+b-2\lambda)}{\lambda}$$

*Proof:* Let $\mathbf{P}$ be the PDA generated by (4). From Theorem 2, we only need to count the number of useless stars in each column of $\mathbf{P}$. For any $A \in \mathcal{K} = \binom{[0,H)}{r}$ and $B \in \mathcal{F} = \binom{[0,H)}{b}$ satisfying $|A \cap B| < \lambda$, we have $p_{B,A} = *$ since $A \cap B \notin \mathcal{I} = \binom{[0,H)}{\lambda}$. Moreover, if $|A \cap B| < \lambda$, the star at $p_{B,A}$ is useless. Otherwise if $p_{B,A} = *$ is useful, which means that it occurs in a subarray $\mathbf{P}^{(C,I)}$ for some $C \in \binom{[0,H)}{r+b-2\lambda}$ and $I \in \binom{[0,H)}{\lambda}$. Then there must exist two subsets, say $A' \in \mathcal{K} \setminus \{A\}$ and $B' \in \mathcal{F} \setminus \{B\}$, such that

$$\begin{aligned} p_{B',A} &= ((A \cup B') - (A \cap B'), A \cap B') = p_{B,A'} \\ &= ((A' \cup B) - (A' \cap B), A' \cap B) = (C, I). \end{aligned}$$

Then we have $A \cap B' = A' \cap B = I$. So we have $I \subseteq A \cap B$, which contradicts with $|A \cap B| < \lambda$ since $|I| = \lambda$.

Since for each $A \in \mathcal{K}$ there are exactly $\sum_{i=0}^{\lambda-1}\binom{r}{i}\binom{H-r}{b-i}$ subsets $B \in \mathcal{F}$ satisfying $|A \cap B| < \lambda$. So each column has $Z' = \sum_{i=0}^{\lambda-1}\binom{r}{i}\binom{H-r}{b-i}$ useless stars. Based on the PDA $\mathbf{P}$ generated by (4) and using Theorem 2, our statement holds. □

*Theorem 4: For any positive integers* $H$, $r$, $b$, $\lambda$ *satisfying* $0 < r, b < H$, $\lambda < \min\{r, b\}$ *and* $r+b \leq H+\lambda$, *there exists an* $(\binom{H}{b} - \sum_{i=\lambda+1}^{\min\{r,b\}}\binom{r}{i}\binom{H-r}{b-i})$*-division* $(\binom{H}{r}, M, N)$ *coded caching scheme with memory ratio and transmission rate as follows*:

$$\frac{M}{N} = 1 - \frac{\binom{r}{\lambda}\binom{H-r}{b-\lambda}}{\binom{H}{b} - \sum_{i=\lambda+1}^{\min\{r,b\}}\binom{r}{i}\binom{H-r}{b-i}}$$

$$R = \frac{\binom{H}{r+b-2\lambda}}{\binom{H}{b} - \sum_{i=\lambda+1}^{\min\{r,b\}}\binom{r}{i}\binom{H-r}{b-i}}\binom{r+b-2\lambda}{r-\lambda}$$

*Proof:* Let $\mathbf{P}'$ be the PDA generated by (5). For any $A \in \mathcal{K} = \binom{[0,H)}{r}$ and $B \in \mathcal{F} = \binom{[0,H)}{b}$ with $|A \cap B| > \lambda$, we have $p'_{B,A} = *$ since $A \cap B \notin \mathcal{I} = \binom{[0,H)}{\lambda}$. We claim that such $p'_{B,A} = *$ is useless. Otherwise if the star at $p'_{B,A}$ is useful, then there must exist two subsets, say $A' \in \mathcal{K} \setminus \{A\}$ and $B' \in \mathcal{F} \setminus \{B\}$ satisfying $|A \cap B'| = \lambda$ and $|A' \cap B| = \lambda$, such that

$$\begin{aligned} p'_{B',A} &= ((A \cup B') - (A \cap B'), A - B') = p'_{B,A'} \\ &= ((A' \cup B) - (A' \cap B), A' - B). \end{aligned}$$

Then we have $A - B' = A' - B$. So we have $(A - B') \cap B = (A' - B) \cap B = \emptyset$. On the other hand, since $|A \cap B| > \lambda$ and $|A \cap B'| = \lambda$, there exists some $x \in [0, H)$ satisfying $x \in A \cap B$ and $x \notin A \cap B'$. Consequently we have $x \in A$, $x \in B$ and $x \notin B'$. So we have $x \in (A - B') \cap B$, which contradicts with $(A - B') \cap B = \emptyset$.

Since for each $A \in \mathcal{K}$ there are exactly $\sum_{i=\lambda+1}^{\min\{r,b\}}\binom{r}{i}\binom{H-r}{b-i}$ subsets $B \in \mathcal{F}$ satisfying $|A \cap B| > \lambda$. So each column of $\mathbf{P}'$ has $Z' = \sum_{i=\lambda+1}^{\min\{r,b\}}\binom{r}{i}\binom{H-r}{b-i}$ useless stars. Based on the PDA $\mathbf{P}'$ generated by (5) and using Theorem 2, our statement holds. □

### C. PERFORMANCE ANALYSES

In this section, we assume the parameters $H$ and $r$ are fixed. The scheme from Lemma 1 is denoted by the original scheme, and the scheme with smaller transmission rate among the schemes from Theorem 3 and Theorem 4 is denoted by the new scheme. Since it is hard to propose a theoretic comparison between the original scheme and the new scheme, the trade off between the transmission rate and the memory ratio is shown in Figure 2 and Figure 3 for $H = 10$, $r = 5$ and $H = 15$, $r = 6$ respectively. From Figure 2 and Figure 3, it is easy to see that the transmission rate of the original scheme is much larger than that of the new scheme when the memory ratio is small. Moreover, the new scheme has more flexible memory ratio than the original scheme.

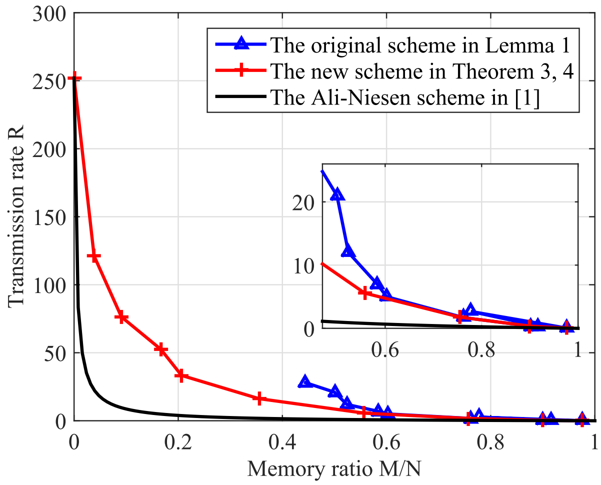


**FIGURE 2. The transmission rate versus the memory ratio for the original scheme from Lemma 1, the new scheme from Theorem 3 and Theorem 4, the Ali-Niesen scheme in [1] when $H = 10$, $r = 5$ and $K = \binom{H}{r} = 252$.**

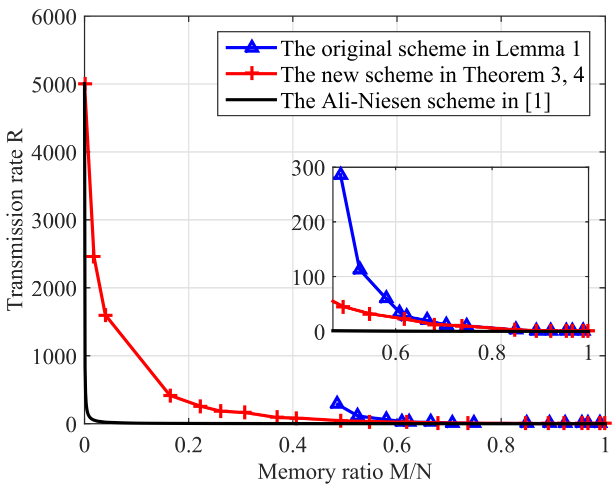


**FIGURE 3. The transmission rate versus the memory ratio for the original scheme from Lemma 1, the new scheme from Theorem 3 and Theorem 4, the Ali-Niesen scheme in [1] when $H = 15$, $r = 6$ and $K = \binom{H}{r} = 5005$.**

## V. CONCLUSION

In this article, we focused on coded caching schemes with low subpacketizations. After we observed that there are some cached packets not fully used in some well known coded caching schemes, i.e., do not generate multicasting opportunities in the delivery phase, we modified the uncoded placement of the scheme realized by an appropriate PDA to coded placement to overcome this drawback. Finally based on the scheme in [9], we obtain a new scheme which has smaller transmission rate and more flexible memory ratio than the original scheme in [9].

## REFERENCES

[1] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.
[2] K. Wan, M. Jit, P. Piantanida, and D. Tuninetti, "On the benefits of asymmetric coded cache placement in combination networks with end-user caches," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 1550–1554.
[3] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless D2D networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 849–869, Feb. 2016.
[4] N. Karamchandani, U. Niesen, M. A. Maddah-Ali, and S. N. Diggavi, "Hierarchical coded caching," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3212–3229, Jun. 2016.
[5] A. Sengupta, R. Tandon, and T. C. Clancy, "Fundamental limits of caching with secure delivery," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 2, pp. 355–370, Feb. 2015.
[6] Q. Yan, M. Cheng, X. Tang, and Q. Chen, "On the placement delivery array design for centralized coded caching scheme," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5821–5833, Sep. 2017.
[7] K. Shanmugam, A. G. Dimakis, J. Llorca, and A. M. Tulino, "A unified Ruzsa-Szemerédi framework for finite-length coded caching," in *Proc. 51st ACSSC*, Pacific Grove, CA, USA, 2017, pp. 631–635.
[8] M. Cheng, J. Jiang, Q. Yan, and X. Tang, "Coded caching schemes for flexible memory sizes," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4166–4176, Jun. 2019.
[9] 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, Feb. 2018.
[10] S. Agrawal, K. V. S. Sree, and P. Krishnan, "Coded caching based on combinatorial designs," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 1227–1231.
[11] H. H. S. Chittoor, P. Krishnan, K. V. S. Sree, and M. V. N. Bhavana, "Subexponential and linear subpacketization coded caching via line graphs and projective geometry," 2020, *arXiv:2001.00399*. [Online]. Available: http://arxiv.org/abs/2001.00399
[12] C. Shangguan, Y. Zhang, and G. Ge, "Centralized coded caching schemes: A hypergraph theoretical approach," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5755–5766, Aug. 2018.
[13] K. Shanmugam, A. M. Tulino, and A. G. Dimakis, "Coded caching with linear subpacketization is possible using Ruzsa-Szeméredi graphs," in *Proc. IEEE ISIT*, Aachen, Germany, pp. 1237–1241, Jun. 2017.
[14] L. Tang and A. Ramamoorthy, "Coded caching schemes with reduced subpacketization from linear block codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3099–3120, Apr. 2018.
[15] K. Shanmugam, M. Ji, A. M. Tulino, J. Llorca, and A. G. Dimakis, "Finite-length analysis of caching-aided coded multicasting," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5524–5537, Oct. 2016.
[16] P. Krishnan, "Coded caching via line graphs of bipartite graphs," in *Proc. IEEE ITW*, Guangzhou, China, Nov. 2018, pp. 1–5.
[17] 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.
[18] J. H. van Lint, *Introduction to Coding Theory*. Berlin, Germany: Springer, 1999.
[19] M. Cheng, Y. Li, X. Zhong, and R. Wei, "Improved constructions of coded caching schemes for combination networks," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 5965–5975, Oct. 2020.
[20] J. J. Quinn and A. T. Benjamin, "Strong chromatic index of subset graphs," *J. Graph Theory*, vol. 24, no. 3, pp. 267–273, Mar. 1997.

**ZHANG MINGMING** received the M.S. degree in electrical engineering from Tianjin Polytechnic University, Tianjin, China, in 2020. She is currently a Research Assistant with Guangxi Normal University, Guilin, Guangxi, China. Her current research interests include combinatorics, coding theory, cryptography, and their interactions.

**CHENG MINQUAN** received the Ph.D. degree from the Department of Social Systems and Management, Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki, Japan, in 2012. In 2012, he joined Guangxi Normal University, Guilin, Guangxi, China, where he is currently a Full Professor with the School of Computer Science and Information Technology. His research interests include combinatorics, coding theory, and cryptography and their interactions.

**WANG JINYU** received the B.S. degree in information and computing science from Central South University, Changsha, China, in 2005, and the M.S. degree in computational mathematics from Zhejiang University, Hangzhou, China, in 2007. In 2007, she joined Guangxi Normal University, Guilin, Guangxi, China, where she is currently a Full Lecturer with the College of Mathematics and Statistics. Her research interests include the numerical solution of differential equations, combinatorics, and coding theory.

**ZHONG XI** received the B.S. degree in statistics from Guangxi Normal University, Guilin, Guangxi, China, in 2018, where she is currently pursuing the master's degree with the School of Computer Science and Technology. Her current research interests include coding theory, and caching networks and their interactions.

**YISHAN CHEN** received the B.S. degree in computer and application from the Central South Institute of Technology, Hengyang, Hunan, China, in 1998, and the master's degree from the School of Computer Science and Technology, Guangxi Normal University, Guilin, Guangxi, China, in 2011. His current research interests include caching networks and intelligent teaching systems.