

Article

Fundamental Limits of Coded Caching in Request-Robust D2D Communication Networks [†]

Wuqu Wang ¹, Zhe Tao ², Nan Liu ^{1,*} and Wei Kang ³

¹ National Mobile Communications Research Laboratory, Southeast University, Nanjing 211189, China; wuquwang@seu.edu.cn

² Huawei Technologies, Nanjing 210012, China

³ School of Information Science and Engineering, Southeast University, Nanjing 211189, China

* Correspondence: nanliu@seu.edu.cn

[†] This article was presented in part at the IEEE International Symposium on Networks, Computers and Communications (ISNCC 2023), Doha, Qatar, 23–26 October 2023.

Abstract: D2D coded caching, originally introduced by Ji, Caire, and Molisch, significantly improves communication efficiency by applying the multi-cast technology proposed by Maddah-Ali and Niesen to the D2D network. Most prior works on D2D coded caching are based on the assumption that all users will request content at the beginning of the delivery phase. However, in practice, this is often not the case. Motivated by this consideration, this paper formulates a new problem called *request-robust D2D coded caching*. The considered problem includes K users and a content server with access to N files. Only r users, known as requesters, request a file each at the beginning of the delivery phase. The objective is to minimize the average and worst-case delivery rate, i.e., the average and worst-case number of broadcast bits from all users among all possible demands. For this novel D2D coded caching problem, we propose a scheme based on uncoded cache placement and exploiting common demands and one-shot delivery. We also propose information-theoretic converse results under the assumption of uncoded cache placement. Furthermore, we adapt the scheme proposed by Yapar et al. for uncoded cache placement and one-shot delivery to the request-robust D2D coded caching problem and prove that the performance of the adapted scheme is order optimal within a factor of two under uncoded cache placement and within a factor of four in general. Finally, through numerical evaluations, we show that the proposed scheme outperforms known D2D coded caching schemes applied to the request-robust scenario for most cache size ranges.

Keywords: coded caching; device-to-device; request-robust; order-optimal scheme



Citation: Wang, W.; Tao, Z.; Liu, N.; Kang, W. Fundamental Limits of Coded Caching in Request-Robust D2D Communication Networks.

Entropy **2024**, *26*, 250. <https://doi.org/10.3390/e26030250>

Academic Editors: Shenghao Yang and Kenneth Shum

Received: 4 February 2024

Revised: 1 March 2024

Accepted: 7 March 2024

Published: 12 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, the demand for user throughput has greatly increased by applications based on fifth-generation (5G) mobile networks [1], such as short videos, self-driving vehicles, the Metaverse, etc. Fortunately, the data of such applications can be pre-stored in the user's storage during low-network consumption periods, preventing network congestion during peak hours. This approach is known as *caching* [2]. There are typically two phases in the caching process [3]. Before knowing any user requests, the server fills the users' caches in the *placement phase* during off-peak hours. The *delivery phase* follows in peak hours. The delivery signals will be designed and transmitted from terminals like the server or the users to satisfy all user demands when they are revealed. Technology for caching has advanced quickly in recent years, and it is currently regarded as one of the effective methods for relieving the congestion of wireless networks.

Traditional caching ignores the processing capability of the users, and therefore, the contents cached by the users and the signals transmitted by the server are both uncoded. In contrast to traditional caching, *coded caching* [3], proposed by Maddah-Ali and Niesen, uses a combination of coded multi-casting and device caching to simultaneously fulfill

multiple requests through coded transmissions. The coded caching strategy works for the prototypical network topology called the *single shared-link network*, e.g., vehicular networks [4]. Both the cache contents of the users and the delivery signal from the server are allowed to be coded in the coded caching problem. The aim is to design a caching and delivery scheme that minimizes the *average* and *worst-case* delivery rate, which is defined as the average and worst-case number of broadcast bits among all possible user requests. When the optimal caching and delivery scheme that achieves the lowest worst-case delivery rate can be found for any user cache size, the optimal tradeoff between rate and memory for the system is determined. If each user directly stores a subset of the files' bits in its cache without coding, it is referred to as an *uncoded* cache placement scheme; otherwise, it is referred to as a *coded* cache placement scheme. The original problem [3] studied in coded caching is *centralized*, assuming that all users present during the placement phase will each make a request for a file at the beginning of the delivery phase. *Decentralized coded caching* [5–7] also considers the possibility of users leaving or turning off during the delivery phase and explores less coordinated caching strategies.

Taking self-driving vehicles as an example: one promising approach to improve the communication efficiency is through the use of device-to-device (D2D) communication, which allows the users to directly exchange information with each other without the need for a server like a base station. This can be particularly useful in situations where the traditional infrastructure is limited or unavailable, such as in remote or rural areas. To solve the coded caching problem in these scenarios, a framework is proposed by Ji et al. in [8] for D2D coded caching. In the placement phase, similar to coded caching [3], the server fills the users' caches before the users make any requests. In the delivery phase, when the users reveal their demands, the server is disconnected from the users and it is up to the users to communicate with each other so that each user can decode the file it requested using the signals transmitted by the other users and the contents of its local cache. For the centralized D2D coded caching problem, the caching strategy of [3] (Algorithm 1), which is uncoded, is widely used in the placement phase, e.g., in [8,9] and so on. In [8], a novel delivery scheme was provided that is appropriate for the D2D scenario. Additionally, a well-known D2D coded caching converse was proposed in [8], and it has been demonstrated that, when the memory size is large, the proposed D2D caching and delivery scheme is order optimal within a constant factor. It is difficult to find the optimal caching and delivery scheme and the corresponding optimal rate–memory tradeoff for the centralized D2D coded caching problem. However, there are many researchers who try to find the fundamental limits of the centralized D2D coded caching problems under certain assumptions or additional constraints, e.g., [9,10].

With concern to the timeliness of the communication, *one-shot delivery*, which is defined to satisfy the condition that each user can decode any bit of its requested file from its own cache and the transmitted signal from at most one other user, is proposed in [9] for the centralized D2D coded caching problem. For example, one self-driving vehicle may quickly decode the requested map data after receiving the signals transmitted by another self-driving vehicle, without waiting for all the considered vehicles to complete the transmission of signals. The proposed caching and delivery scheme in [9] is optimal under the constraint of uncoded cache placement and one-shot delivery, and it is order optimal within a factor of two if the converse of the shared-link coded caching problem with uncoded cache placement [11] is used as the lower bound and order optimal within a factor of four compared to the general D2D coded caching converse results.

In addition to [9], many other researchers study variants of the D2D coded caching problems, such as allowing for coded placement with three users [10], private caching [12], private caching with a trusted server [13,14], distinct cache sizes [15], finite file packetizations [16], finite-length analysis [17], secure coded caching [18], secure delivery [19], wireless multi-hop D2D networks [20,21], partially cooperative D2D communication networks [22,23], constructions of placement delivery arrays (PDAs) [24], and so on. Among these papers, most of them assume that all users will request content at the begin-

ning of the delivery phase. However, in practice, this may not be true. For example, when assisted self-driving vehicles within a certain range carry out D2D communication, they may not request at the same time or some of them may be driven manually and do not need to access high-definition map data. In these situations, waiting for all users to request content will waste time, and setting the requests of the users who do not request some arbitrary file demand will waste communication resources. Note that in these scenarios, even though the users may not request data, they are still available to participate in the delivery phase by transmitting signals that are functions of their cached contents.

Hence, in this paper, we propose and study a new problem called *request-robust D2D coded caching*, where in the delivery phase, though all users in the placement phase are still present and may help with the transmission, some of them do not request any files. It is not known in the placement phase the number or identity of the users who do not request files. This problem is not the same as the decentralized D2D coded caching problem [8], where users who leave or turn off during the delivery phase do not make file requests, nor do they participate in the delivery. Note that this problem is similar to the user inactivity problem in the D2D caching setting [9,22], where each user may independently have a probability of being inactive, i.e., they do not make a file request at the beginning of the delivery phase. However, in the request-robust D2D coded caching problem, inactive users still help in the delivery phase by transmitting signals, whereas in the user inactivity problem, they do not.

1.1. Main Contributions

The main contributions of the paper can be summarized as follows:

- (1) For the request-robust D2D coded caching problem, we adapt the scheme from [9] for uncoded cache placement and one-shot delivery and call the adapted scheme the *adapted Yapar–Wan–Schaefer–Caire (YWSC) scheme*.
- (2) In order to find better performance, we present a new achievable scheme based on the uncoded cache placement and exploiting common demands [11] and one-shot delivery [9]. The caching strategy is the same as that proposed by Maddah-Ali and Niesen in [3] (Algorithm 1), while the delivery strategy divides the sub-files into three categories, and different delivery signals are designed for each category. We call the new scheme the *three-category-based scheme*. This scheme was presented in the conference version of this paper [25].
- (3) We propose an information-theoretic lower bound under uncoded cache placement based on seeking the converse of a problem called coded caching with inactive users. The problem of coded caching with inactive users was proposed in [26], where users are inactive with a certain probability in the traditional coded caching problem of [3]. Hence, the converse for the problem of coded caching with inactive users can serve as a converse for the request-robust D2D coded caching problem. Note that [26] only considers the optimization of the cache replication parameter and does not provide a converse for the caching and delivery scheme.
- (4) We prove that the performance of the adapted YWSC scheme is order optimal within a factor of two under the assumption of uncoded cache placement and within a factor of four in general.
- (5) Through numerical evaluation, we show that the three-category-based scheme outperforms the adapted YWSC scheme, as well as other known D2D coded caching schemes [3] applied to the request-robust scenario.

1.2. Notations

Throughout this paper, $H(\cdot)$ represents the entropy of random variables, $|\cdot|$ represents the cardinality of a set, \oplus denotes finite field addition, we let $\mathcal{X} \setminus \mathcal{Y} \triangleq \{x \in \mathcal{X} | x \notin \mathcal{Y}\}$, $[x : y : z] \triangleq \{x, x + y, x + 2y, \dots, z\}$, $[x : y] = [x : 1 : y]$ and $[n] = [1 : n]$. For two integers x , and y , if $x < y$ or $x \leq 0$, we let $\binom{y}{x} = 0$.

2. System Model and Related Background

2.1. System Model

We study the *request-robust D2D coded caching* problem, which is defined in the following. We consider a D2D coded caching system (see Figure 1) where a server is connected to a fixed content file database of N files, $\mathcal{W} \triangleq (W_1, \dots, W_N)$. Each file consists of F bits. There are K users in the system, each with a cache of size MF bits. We focus on the non-trivial scenario where $M \leq N$. Let \mathcal{K} be the set of user indices, i.e., $\mathcal{K} \triangleq [K]$.

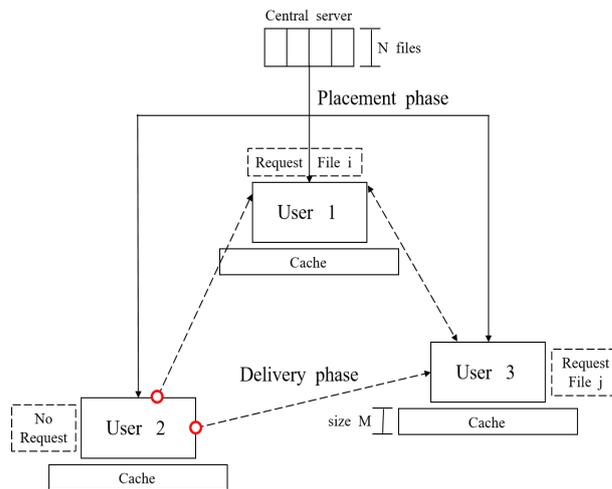


Figure 1. System model for request-robust D2D coded caching problem when there are 3 users. In this realization, User 2 does not request. Solid and dotted lines indicate placement and delivery phases, respectively.

The system operates in two phases. In the placement phase, each user’s cache is filled by the central server, which does not know the number of users or the identities of the users requesting files in the other phase. Denote the content in the cache of User k as Z_k , $k \in \mathcal{K}$. In the delivery phase, some of the K users will make file requests while others will not. We denote the set of users making file requests as \mathcal{R} , $\mathcal{R} \subseteq \mathcal{K}$. Each user in \mathcal{R} will request a single file. Let r denote the number of users requesting files, i.e., $r \triangleq |\mathcal{R}|$, and we assume that the file requests, i.e., which user requests which file and which users are not requesting any files, are known to all K users. Each of the K users will send a signal that will be received by the users in \mathcal{R} . It is required that each user in \mathcal{R} can decode its requested file by using the signals received and its own cache content. Note that Figure 1 is different from [9] (Figure 1), i.e., there exists a user who does not request any file in Figure 1, while in [9] (Figure 1), all users request files.

More specifically, a caching and delivery scheme for this system consists of

1. K caching functions

$$\varphi_k : [2^F]^N \rightarrow [2^{MF}], \quad k \in \mathcal{K},$$

which map the N files into cache contents of the users, denoted by $Z_k = \varphi_k(W_1, \dots, W_N)$, $k \in \mathcal{K}$. Thus, we have the following entropy constraint:

$$H(Z_k | W_1, W_2, \dots, W_N) = 0, \quad k \in \mathcal{K}.$$

2. $K \sum_{r=1}^K \binom{K}{r} N^r$ encoding functions

$$\phi_k^{D_{\mathcal{R}}} : [2^{MF}] \rightarrow [2^{RkF}], \quad k \in \mathcal{K}.$$

where $D_{\mathcal{R}}$ is the set of file requests made by the users in \mathcal{R} . For example, if there are $K = 4$ users, and Users 1 and 3 do not request files during the delivery phase,

the request vector is $D_{\{2,4\}} = (d_2, d_4)$. The encoding function $\phi_k^{D_{\mathcal{R}}}$ denotes the mapping of User k from its cached content to the signal it transmits, which is denoted as $X_k^{D_{\mathcal{R}}}$, i.e., $X_k^{D_{\mathcal{R}}} \triangleq \phi_k^{D_{\mathcal{R}}}(Z_k)$. Thus, we have

$$H(X_k^{D_{\mathcal{R}}}|Z_k) = 0, \quad k \in \mathcal{K}. \tag{1}$$

We assume the signal $X_k^{D_{\mathcal{R}}}$ consists of $R_k^{D_{\mathcal{R}}}F$ bits. The signals transmitted by all K users consist of $R^{D_{\mathcal{R}}}F$ bits, i.e., $R^{D_{\mathcal{R}}} = \sum_{k=1}^K R_k^{D_{\mathcal{R}}}$.

3. $\sum_{r=1}^K \binom{K}{r} r N^r$ decoding functions

$$\psi_k^{D_{\mathcal{R}}} : [2^{MF}] \times [2^{F \sum_{u \in \mathcal{K} \setminus \{k\}} R_u^{D_{\mathcal{R}}}}] \rightarrow [2^F], \quad k \in \mathcal{R},$$

which is the decoding function used by User k . For example, if there are $K = 4$ users, and Users 1 and 3 do not request any file during the delivery phase, the decoded files at Users 2 and 4 are $\hat{W}_{d_2} = \psi_2^{(d_2, d_4)}(Z_2, X_1^{(d_2, d_4)}, X_3^{(d_2, d_4)}, X_4^{(d_2, d_4)})$ and $\hat{W}_{d_4} = \psi_4^{(d_2, d_4)}(Z_4, X_1^{(d_2, d_4)}, X_2^{(d_2, d_4)}, X_3^{(d_2, d_4)})$, respectively.

Correct decoding by the users requesting files is given by $\hat{W}_{d_k} = W_{d_k}, k \in \mathcal{R}$, or in other words,

$$H(W_{d_k}|Z_k, X_{[K] \setminus \{k\}}^{D_{\mathcal{R}}}) = 0, \quad k \in \mathcal{R}, \tag{2}$$

which is called the decodability constraint. We find that by combining (1) and (2), one can decode any file by knowing the cache of all users, i.e.,

$$H(W_{[N]}|Z_{[K]}) = 0. \tag{3}$$

which implies that we are interested in the case where $KM \geq N$.

For any caching and delivery scheme that satisfies the decodability constraint, for a fixed \mathcal{R} with size r , we define $\mathcal{D}^{\mathcal{R}}$ as the set of all possible demands $\{1, \dots, N\}^r$. We are interested in two performance metrics: one is with respect to the average performance, and the other is with respect to the worst performance. More specifically, the average performance is defined as follows: we assume that the request vector $D_{\mathcal{R}}$ is uniformly distributed on $\mathcal{D}^{\mathcal{R}}$. Then, the average delivery rate with respect to the uniform demand $R_{\text{ave, req-rob}}^{\mathcal{R}}$ is defined as

$$R_{\text{ave, req-rob}}^{\mathcal{R}} = \mathbb{E}_{D_{\mathcal{R}}} [R^{D_{\mathcal{R}}}]$$

For a given r , we define the maximum average delivery rate with respect to the uniform demand $R_{\text{ave, req-rob}}^r$, where the maximum is over all request sets \mathcal{R} with size r , i.e.,

$$R_{\text{ave, req-rob}}^r = \max_{\mathcal{R}: |\mathcal{R}|=r} R_{\text{ave, req-rob}}^{\mathcal{R}}$$

The worst-case performance is defined as follows: first, the worst-case delivery rate $R_{\text{worst, req-rob}}^{\mathcal{R}}$ is defined as

$$R_{\text{worst, req-rob}}^{\mathcal{R}} = \max_{D_{\mathcal{R}}} R^{D_{\mathcal{R}}}.$$

For a given r , we define the maximum worst-case delivery rate $R_{\text{worst, req-rob}}^r$, where the maximum is over all request sets \mathcal{R} with size r , i.e.,

$$R_{\text{worst, req-rob}}^r = \max_{\mathcal{R}: |\mathcal{R}|=r} R_{\text{worst, req-rob}}^{\mathcal{R}}.$$

We would like to design caching and delivery schemes such that $R_{\text{ave,req-rob}}^r$ and $R_{\text{worst,req-rob}}^r$ are both the smallest for every $r = 1, 2, \dots, K$. As can be seen, this is a multi-objective optimization.

For easy presentability of the results, following the notation of [11], we denote $N_e(D_{\mathcal{R}})$ as the number of distinct files in a request vector $D_{\mathcal{R}}$. $D_{\mathcal{R} \setminus \{k\}}$ and $N_e(D_{\mathcal{R} \setminus \{k\}})$ are denoted as the request vector of users $\mathcal{R} \setminus \{k\}$ and the number of distinct files requested by all requesters but User k , respectively.

2.2. Preliminaries

In this subsection, we briefly summarize the related approaches, namely the uncoded symmetric placement scheme in [3] and the problem of coded caching with inactive users, which are critical for building our results for the request-robust D2D coded caching problem.

2.2.1. Uncoded Symmetric Placement Scheme

First, we introduce the uncoded symmetric placement scheme, which is useful for our scheme proposed in Section 4.

Definition 1. (Maddah-Ali Niesen [MAN] Uncoded Symmetric Placement Scheme): Define t as $t = KM/N$. When t is an integer, we have the MAN uncoded symmetric placement scheme as follows: Each file W_n is divided into $\binom{K}{t}$ disjoint sub-files denoted by $W_{n,\mathcal{T}}$, where $n \in [N]$, $\mathcal{T} \subseteq K$, $|\mathcal{T}| = t$, and $H(W_{n,\mathcal{T}}) = F/\binom{K}{t}$. Each user k caches all the bits of the sub-files $W_{n,\mathcal{T}}$, $n \in [N]$, for all $\mathcal{T} \ni k$. Since each file includes $\binom{K-1}{t-1}$ sub-files with $\mathcal{T} \ni k$, each user k satisfies the memory constraint $H(Z_k) = NF\binom{K-1}{t-1}H(W_{n,\mathcal{T}}) = NFt/K = MF$.

For the convenience of understanding and reference, we give the algorithm of this scheme in Algorithm 1.

Algorithm 1 MAN Uncoded Symmetric Placement Scheme ($N, K, M, W_{[N]}$)

- 1: $t \leftarrow KM/N$
 - 2: $\mathcal{T} \leftarrow \{\mathcal{T} \subseteq [K] : |\mathcal{T}| = t\}$
 - 3: **for** $n \in [N]$ **do**
 - 4: Divide file W_n into disjoint sub-files ($W_{n,\mathcal{T}} : \mathcal{T} \in \mathcal{T}$) with equal size
 - 5: **end for**
 - 6: **for** $k \in [K]$ **do**
 - 7: $Z_k \leftarrow (W_{n,\mathcal{T}} : n \in [N], \mathcal{T} \in \mathcal{T}, k \in \mathcal{T})$
 - 8: **end for**
-

The uncoded symmetric placement scheme is the optimal achievable placement scheme both for the shared-link model with uncoded cache placement [3] and D2D work with uncoded cache placement and one-shot delivery [9], which reveals that regardless of the number of users, using the uncoded symmetric placement scheme can satisfy the optimal rate of these models in all cases. Due to the superiority of the uncoded symmetric placement scheme, we use the scheme as the placement scheme in our scheme proposed in Section 4.

2.2.2. Problem of Coded Caching with Inactive Users

We denote our original D2D model with r users of K users requesting files independently in \mathcal{W} as System 1. In order to derive the converse of System 1, we consider another system model named *coded caching with inactive users*, denoted as System 2. This is the model where a central server responds to the users' requests, and some of the users do not request any files in the delivery phase. The central server connects to the whole file database.

The placement phase of System 2 is exactly the same as that of System 1. Thus, in System 2, Equation (3) is still satisfied. The delivery phase of System 2 is different from

that of System 1. Specifically, in System 1, the codewords are transmitted by the users, while in System 2, the codewords are transmitted by the central server. Since the central server has the whole database and connects to all K users, while in System 1, each user only caches a subset of the whole database and only connects to other $K - 1$ users, the optimal maximum average and worst-case delivery rate in System 2, denoted as $R_{ave,inactive}^{r*}$ and $R_{worst,inactive}^{r*}$, can not be larger than the delivery rate in System 1, respectively. In other words, we have the following inequality:

$$R_{ave,req-rob}^{r*} \geq R_{ave,inactive}^{r*} \tag{4}$$

$$R_{worst,req-rob}^{r*} \geq R_{worst,inactive}^{r*} \tag{5}$$

3. Main Result

In this section, we present the main results of this work. We propose two achievable schemes for the request-robust D2D coded caching problem in Theorems 1 and 2. We further propose a converse for the problem of coded caching with inactive users in Theorem 3, which also serves as a converse result to the request-robust D2D coded caching problem. Theorem 4 compares the performance gap between the achievability result in Theorem 1 and the converse result in Theorem 3 and shows that they are within a multiplicative gap.

The first achievable scheme is obtained by adapting the achievable scheme in [9] with uncoded cache placement and one-shot delivery to the request-robust D2D coded caching problem. More specifically, the adaptation is performed by assigning the users, who do not request, a demand that is most requested by the requesters. We call the adapted scheme the *adapted Yapar–Wan–Schaefer–Caire (YWSC) scheme*. We denote the adapted request vector as $D'_{\mathcal{K}}$ and the adapted request vector of users $\mathcal{K} \setminus \{k\}$ as $D'_{\mathcal{K} \setminus \{k\}}$. Hence, we have $N_e(D'_{\mathcal{K}}) = N_e(D_{\mathcal{R}})$ and obtain Theorem 1 as follows:

Theorem 1. *For the request-robust D2D coded caching problem, the optimal maximum average delivery rate with respect to the uniform demand is upper bounded by*

$$R_{ave,req-rob}^{r*} \leq \mathbb{E}_{D'_{\mathcal{K}}} \left\{ \frac{K \binom{K-1}{t} - \sum_{i=1}^K \binom{K-1-N_e(D'_{\mathcal{K} \setminus \{i\}})}{t} - f \left[\binom{K-r}{t} - \binom{K-r-1}{t} \right]}{t \binom{K}{t}} \right\}, \tag{6}$$

when $t = \frac{KM}{N}$ is an integer in $[K]$, where f is an integer equal to one if and only if each requester demands a distinct file, i.e., $N_e(D_{\mathcal{R}}) = r$; otherwise, $f = 0$. When $t \notin [K]$, $R_{ave,req-rob}^{r*}$ is upper bounded by the lower convex envelope of the values in (6) for integer values of $t \in [K]$.

For the maximum worst-case delivery rate, we have

$$R_{worst,req-rob}^{r*} \leq \max_{D'_{\mathcal{K}}} \left\{ \frac{K \binom{K-1}{t} - \sum_{i=1}^K \binom{K-1-N_e(D'_{\mathcal{K} \setminus \{i\}})}{t} - f \left[\binom{K-r}{t} - \binom{K-r-1}{t} \right]}{t \binom{K}{t}} \right\}, \tag{7}$$

$$= \begin{cases} \frac{K \binom{K-1}{t} - (K-r) \binom{K-r-1}{t} - r \binom{K-r}{t}}{t \binom{K}{t}}, & r \leq N, \\ \frac{K \binom{K-1}{t} - (2N-r) \binom{K-N}{t} - (K+r-2N) \binom{K-1-N}{t}}{t \binom{K}{t}}, & \text{otherwise,} \\ \frac{K \left[\binom{K-1}{t} - \binom{K-1-N}{t} \right]}{t \binom{K}{t}}, & r \geq 2N, \end{cases} \tag{8}$$

where $t = \frac{KM}{N}$ is an integer in $[K]$. When $t \notin [K]$, $R_{worst,req-rob}^{r*}$ is upper bounded by the lower convex envelope of the values in (7) for integer values of $t \in [K]$.

Proof. The proof of Theorem 1 is provided in Appendix A. \square

Theorem 1 is a simple adaptation of an existing scheme; in order to improve its performance, we propose a new scheme, called the *three-category-based scheme*, and obtain Theorem 2 as follows:

Theorem 2. For the request-robust D2D coded caching problem, the optimal maximum average delivery rate with respect to the uniform demand is upper bounded by

$$R_{ave,req-rob}^{r*} \leq \mathbb{E}_{D_{\mathcal{R}}} \left\{ \frac{\sum_{i=\max\{1,t+r-K\}}^{\min\{t-1,r-1\}} \binom{K-r}{t-i} \left[\binom{r}{i+1} - \binom{r-N_e(D_{\mathcal{R}})}{i+1} \right]}{\binom{K}{t}} + \frac{\binom{K-r}{t} N_e(D_{\mathcal{R}})}{\binom{K}{t}} + \frac{r \binom{r-1}{t} - \sum_{i \in \mathcal{R}} \binom{r-1-N_e(D_{\mathcal{R} \setminus \{i\}})}{t}}{t \binom{K}{t}} \right\}, \quad (9)$$

where $t = \frac{KM}{N}$ is an integer in $[K]$. When $t \notin [K]$, $R_{ave,req-rob}^{r*}$ is upper bounded by the lower convex envelope of the values in (9) for integer values of $t \in [K]$.

Then, for the maximum worst-case delivery rate, we have

$$R_{worst,req-rob}^{r*} \leq \max_{D_{\mathcal{R}}} \left\{ \frac{\sum_{i=\max\{1,t+r-K\}}^{\min\{t-1,r-1\}} \binom{K-r}{t-i} \left[\binom{r}{i+1} - \binom{r-N_e(D_{\mathcal{R}})}{i+1} \right]}{\binom{K}{t}} + \frac{\binom{K-r}{t} N_e(D_{\mathcal{R}})}{\binom{K}{t}} + \frac{r \binom{r-1}{t} - \sum_{i \in \mathcal{R}} \binom{r-1-N_e(D_{\mathcal{R} \setminus \{i\}})}{t}}{t \binom{K}{t}} \right\}, \quad (10)$$

$$= \begin{cases} \frac{\sum_{i=\max\{1,t+r-K\}}^{\min\{t-1,r-1\}} \binom{K-r}{t-i} \binom{r}{i+1}}{\binom{K}{t}} + \frac{r \binom{K-r}{t}}{\binom{K}{t}} + \frac{r \binom{r-1}{t}}{t \binom{K}{t}}, & r \leq N, \\ \frac{\sum_{i=\max\{1,t+r-K\}}^{\min\{t-1,r-1\}} \binom{K-r}{t-i} \left[\binom{r}{i+1} - \binom{r-N}{i+1} \right]}{\binom{K}{t}} + \frac{N \binom{K-r}{t}}{\binom{K}{t}} + \frac{r \binom{r-1}{t} - (2N-r) \binom{r-N}{t} - 2(r-N) \binom{r-1-N}{t}}{t \binom{K}{t}}, & \text{otherwise,} \\ \frac{\sum_{i=\max\{1,t+r-K\}}^{\min\{t-1,r-1\}} \binom{K-r}{t-i} \left[\binom{r}{i+1} - \binom{r-N}{i+1} \right]}{\binom{K}{t}} + \frac{N \binom{K-r}{t}}{\binom{K}{t}} + \frac{r \left[\binom{r-1}{t} - \binom{r-1-N}{t} \right]}{t \binom{K}{t}}, & r \geq 2N, \end{cases} \quad (11)$$

where $t = \frac{KM}{N}$ is an integer in $[K]$. When $t \notin [K]$, $R_{worst,req-rob}^{r*}$ is upper bounded by the lower convex envelope of the values in (10) for integer values of $t \in [K]$.

Proof. In proving Theorem 2, we propose a new scheme, i.e., the three-category-based scheme, where the sub-files are divided into three categories and different delivery signals are designed for each category. The detailed proof can be found in Section 4. \square

Remark 1. In Section 5, we numerically compare the delivery rates of the three-category-based scheme and the adapted YWSC scheme, and it can be seen that the proposed three-category-based scheme outperforms the adapted YWSC scheme in all cases cited (see Section 5 for the cited cases and comparison results).

In the following theorem, we characterize a converse for the request-robust D2D coded caching problem.

Theorem 3. For the request-robust D2D coded caching problem, the optimal maximum average delivery rate with respect to the uniform demand and under the constraint of uncoded placement is lower bounded by:

$$R_{ave,req-rob}^{r*} \geq \mathbb{E}_{D_{\mathcal{R}}} \left[\frac{\binom{K}{t+1} - \binom{K-N_e(D_{\mathcal{R}})}{t+1}}{\binom{K}{t}} \right], \quad (12)$$

where $t = \frac{KM}{N}$ is an integer in $[K]$, where $D_{\mathcal{R}}$ is uniformly distributed over $\mathcal{D}^{\mathcal{R}}$. When $t \notin [K]$, $R_{ave,req-rob}^{r*}$ is lower bounded by the lower convex envelope of the values in (12) for integer values of $t \in [K]$.

Then, for the maximum worst-case delivery rate, we have that the optimal maximum worst-case delivery rate under the constraint of uncoded placement is lower bounded by

$$R_{worst,req-rob}^{r*} \geq \frac{\binom{K}{t+1} - \binom{K-\min\{r,N\}}{t+1}}{\binom{K}{t}}, \tag{13}$$

where $t = \frac{KM}{N}$ is an integer in $[K]$. When $t \notin [K]$, $R_{worst,req-rob}^{r*}$ is lower bounded by the lower convex envelope of the values in (13) for integer values of $t \in [K]$.

Proof. Similar to [9], we use the converse of the central server version, i.e., System 2, as a converse for the request-robust D2D coded caching problem when the converse is under the constraint of uncoded placement. The detailed proof is provided in Appendix B. \square

We compare the rate achieved by the adapted YWSC scheme from Theorem 1 and the converse present in Theorem 3 and obtain a multiplicative gap result of Theorem 4 as follows:

Theorem 4. For the request-robust D2D coded caching problem, the upper bounds of the optimal maximum average and worst-case rates from Theorem 1 are order optimal within a factor of two under the constraint of uncoded cache placement and within a factor of four in general.

Proof. The proof of Theorem 4 is given in Appendix D. \square

Remark 2. It is hard to analytically prove that the rate achieved by the three-category-based scheme from Theorem 2 outperforms the rate achieved by the adapted YWSC scheme from Theorem 1 in all cases. However, since in the numerical comparisons in Section 5 the three-category-based scheme performs better than the adapted YWSC scheme in all cases cited, we conjecture that the rate achieved by the three-category-based scheme from Theorem 2 and the converse present in Theorem 3 also follow the multiplicative gap characterized by Theorem 4.

4. A Novel Achievable Scheme, i.e., Proof of Theorem 2

In this section, we present an achievable scheme for the request-robust D2D coded caching problem. The scheme achieves the rate stated in Theorem 2. We will first provide a general achievable scheme, which is based on uncoded cache placement and exploiting common demands [11] and one-shot delivery [9]. Then, we will characterize the performance of the proposed scheme and show that for any requester set \mathcal{R} and corresponding request vector $D_{\mathcal{R}}$, the proposed three-category-based scheme achieves the rate

$$R_{req-rob}^{D_{\mathcal{R}}} = \frac{\sum_{i=\max\{1,t+r-K\}}^{\min\{t-1,r-1\}} \binom{K-r}{t-i} \left[\binom{r}{i+1} - \binom{r-N_e(D_{\mathcal{R}})}{i+1} \right]}{\binom{K}{t}} + \frac{\binom{K-r}{t} N_e(D_{\mathcal{R}})}{\binom{K}{t}} + \frac{r \binom{r-1}{t} - \sum_{i \in \mathcal{R}} \binom{r-1-N_e(D_{\mathcal{R} \setminus \{i\}})}{t}}{t \binom{K}{t}}, \tag{14}$$

which, with the explicit characterization of the maximum worst-case delivery rate in Section 4.2, immediately proves Theorem 2. Finally, we will provide an example to aid in a better understanding of the proposed three-category-based scheme.

4.1. General Scheme

For the placement phase, because the central server does not know the number of requesters r , we use the highly adaptable MAN uncoded symmetric placement scheme

described in Definition 1, denoted as \mathcal{M}_{MAN} . In the following, we restrict to integer values of $t \in [1 : K]$. For cache size M where $t = KM/N$ is not an integer, memory-sharing will be performed [3,8].

For the delivery phase, let the set of requesters be \mathcal{R} with size r . The r requesters each demand a single file. The delivery strategy is divided into the following steps:

- (a) **Determining the leading requesters:** Each user $k \in \mathcal{K} \setminus \mathcal{R}$ who does not request arbitrarily selects a subset of $N_e(D_{\mathcal{R}})$ requesters, denoted by $\mathcal{U}^k = \{u_1^k, \dots, u_{N_e(D_{\mathcal{R}})}^k\} \subseteq \mathcal{R}$, that request $N_e(D_{\mathcal{R}})$ distinct files. Using the idea of leaders from [11], we name these requesters as the *leading requesters of User k* .
- (b) **Splitting the sub-files into three categories:** Recall that each sub-file is denoted as $W_{n,\mathcal{T}}$ and is cached by only users in \mathcal{T} . If $\mathcal{T} \subseteq \mathcal{K} \setminus \mathcal{R}$, then this sub-file belongs to the first category, which is the set of sub-files that are only cached by users who do not request any files. If \mathcal{T} contains some elements from \mathcal{R} and some elements from $\mathcal{K} \setminus \mathcal{R}$, then this sub-file belongs to the second category, which is the set of sub-files that are cached by both requesters and non-requesters. Finally, if $\mathcal{T} \subseteq \mathcal{R}$, then this sub-file belongs to the third category, which is the set of sub-files that are only cached by users who make file requests.

The three categories may not all exist or be required by requesters, and it depends on the value of r and t . When $t \in [K - r + 1, K]$ or $r = K$, the first category does not exist. When $t = 1$ or $r = K$, the second category does not exist, and when $r = 1$ or $t = K$, the second category is not required. When $t \in [r + 1, K]$, the third category does not exist, and when $t = r$ or $r = 1$, the third category is not required.

- (c) **Transmitting signals for the sub-files in the three categories:** we will discuss the delivery scheme for the sub-files in each of the three categories.
- (i) For the sub-files in the first category needed by Requester $k, k \in \mathcal{R}$, since these sub-files are not cached in any of the requesters, the users in $\mathcal{K} \setminus \mathcal{R}$ who cache these sub-files transmit them in an uncoded form. Suppose $W_{d_k,\mathcal{T}}$ is requested by User $k \in \mathcal{R}, \mathcal{T} \subseteq \mathcal{K} \setminus \mathcal{R}$, any of the t users in \mathcal{T} can transmit the sub-file in an uncoded form. However, we adopt the file-splitting strategy in [9] and allow each user in \mathcal{T} to transmit $1/t$ part of the sub-file, i.e., $W_{d_k,\mathcal{T}}$ is divided into t pieces, each consisting of $\frac{F}{t \binom{K}{t}}$ numbers of bits. The pieces are denoted as $W_{d_k,\mathcal{T},a}, a \in \mathcal{T}$, and User a transmits

$$X_a^{\text{1st},d_k,\mathcal{T}} = W_{d_k,\mathcal{T},a}. \tag{15}$$

We notice that $W_{d_k,\mathcal{T},a}$ may be needed by other requesters, i.e., there may be other requesters that request file d_k also. Hence, we let each user a transmit in sequence $X_a^{\text{1st},d_k,\mathcal{T}}$ for all $k \in \mathcal{U}^a$. Hence, the rate of transmitting total bits for the sub-files in the first category is

$$R^{\text{1st}} = \frac{t \binom{K-r}{t} N_e(D_{\mathcal{R}})}{t \binom{K}{t}} = \frac{\binom{K-r}{t} N_e(D_{\mathcal{R}})}{\binom{K}{t}}, \tag{16}$$

because for each \mathcal{T} , every user $a \in \mathcal{T}$ transmits $N_e(D_{\mathcal{R}}) \frac{F}{t \binom{K}{t}}$ bits, and there are t users in each \mathcal{T} , and a total of $\binom{K-r}{t}$ number of \mathcal{T} that are subsets of $\mathcal{K} \setminus \mathcal{R}$.

- (ii) Consider a sub-file in the second category needed by Requester $k, k \in \mathcal{R}$, denoted as $W_{d_k,\mathcal{T}}$, where $k \notin \mathcal{T}$. Denote the set of elements in \mathcal{T} that are in \mathcal{R} as \mathcal{B} , whose size is denoted as i , and we have $1 \leq i \leq r - 1$, because we know at least User k who is requesting a file is not in \mathcal{T} . Further denote the set of elements of \mathcal{T} that are in $\mathcal{K} \setminus \mathcal{R}$ as $\hat{\mathcal{B}}$, whose size is $t - i$, and we have $1 \leq t - i \leq K - r$. Hence, \mathcal{T} can be written as $\mathcal{T} = \mathcal{B} \cup \hat{\mathcal{B}}$. Furthermore, i must satisfy $i \in [\max\{1, t + r - K\}, \min\{t - 1, r - 1\}]$.

Consider the set $\hat{\mathcal{B}} \triangleq \{k\} \cup \mathcal{B}$, which is a set consisting of $i + 1$ requesters. The sub-file $W_{d_x, \hat{\mathcal{B}} \cup \hat{\mathcal{B}} \setminus \{x\}}$ is needed for $x \in \hat{\mathcal{B}}$. We ask that this sub-file be transmitted by the $t - i$ non-requesters, and $W_{d_x, \hat{\mathcal{B}} \cup \mathcal{B} \setminus \{x\}}$ for any $x \in \mathcal{B}$ be divided into $t - i$ equal-length disjoint sub-pieces of $\frac{F}{(t-i)\binom{K}{t}}$ bits, which are denoted by $W_{d_x, \hat{\mathcal{B}} \cup \mathcal{B} \setminus \{x\}, b'}$, where $b \in \hat{\mathcal{B}}$. Hence, if User $b \in \hat{\mathcal{B}}$ transmits

$$X_b^{2nd, d_x, \hat{\mathcal{B}} \cup \hat{\mathcal{B}} \setminus \{x\}} = \bigoplus_{x \in \hat{\mathcal{B}}} W_{d_x, \hat{\mathcal{B}} \cup \mathcal{B} \setminus \{x\}, b'} \tag{17}$$

the sub-piece retrieval can be accomplished for each requester in $\hat{\mathcal{B}}$ since User x has all the sub-pieces on the RHS of (17), except for $W_{d_x, \hat{\mathcal{B}} \cup \mathcal{B} \setminus \{x\}, b}$.

We ask each user $b \in \mathcal{K} \setminus \mathcal{R}$ to transmit $X_b^{2nd, d_x, \hat{\mathcal{B}} \cup \hat{\mathcal{B}} \setminus \{x\}}$ in sequence, only if $\hat{\mathcal{B}} \cap \mathcal{U}^{lb} \neq \emptyset$, i.e., user b will not transmit if the set $\hat{\mathcal{B}}$ consists of only non-leading requesters. We now count the amount of transmission for the second category. For a fixed i , the number of $\hat{\mathcal{B}}$, which is of size $t - i$, is $\binom{K-r}{t-i}$. For each $u \in \hat{\mathcal{B}}$, the number of transmitted bits is $\binom{r}{i+1} - \binom{r-N_e(D_{\mathcal{R}})}{i+1}$ times the size of a sub-piece, and there are a total of $t - i$ users in $\hat{\mathcal{B}}$. Hence, the rate of transmitting all the bits for the sub-files in the second category is

$$R^{2nd} = \frac{\sum_{i=\max\{1, t+r-K\}}^{\min\{t-1, r-1\}} \binom{K-r}{t-i} \left[\binom{r}{i+1} - \binom{r-N_e(D_{\mathcal{R}})}{i+1} \right]}{\binom{K}{t}}. \tag{18}$$

The next lemma shows that the scheme proposed satisfies the decodability constraint, even for the non-leading requesters.

Lemma 1. *The scheme proposed for the sub-files in the second category satisfies the decodability constraint, i.e., (2).*

The proof of Lemma 1 is based on showing the equivalence of the designed scheme and that in [11]. Hence, we state the following remark first:

Remark 3. *Notice that when the sub-files are in the second category and the parameter t is fixed, for each partition integer i , a user $b \in \hat{\mathcal{B}}$ generates its codewords exclusively from the sub-pieces $W_{d_k, \hat{\mathcal{B}} \cup \mathcal{B}, b}$, and there exist $\binom{r}{i}$ such sub-pieces in its cache. In addition, for any $c \in \hat{\mathcal{B}} \setminus \{b\}$, we have $W_{d_k, \hat{\mathcal{B}} \cup \mathcal{V}^b, b} \cap W_{d_l, \hat{\mathcal{B}} \cup \mathcal{V}^c, c} = \emptyset$ for any $\mathcal{V}^b, \mathcal{V}^c \subseteq \mathcal{R}, |\mathcal{V}^b| = |\mathcal{V}^c| = i, k \in \mathcal{V}^b, l \in \mathcal{V}^c$. That is to say, users in $\hat{\mathcal{B}}$ generate their codewords based on non-overlapping libraries of size $(t - i)N \binom{r}{i} \frac{F}{(t-i)\binom{K}{t}} = N \binom{r}{i} F / \binom{K}{t}$ bits. Also, observe that the cache of requester k contains $(t - i) \binom{r-1}{i-1}$ such $W_{d_k, \hat{\mathcal{B}} \cup \mathcal{V}^b, b}$ sub-pieces, which amount to $N(t - i) \binom{r-1}{i-1} \frac{F}{(t-i)\binom{K}{t}} = N \binom{r-1}{i-1} \frac{F}{\binom{K}{t}} = Ni \binom{r}{i} \frac{F}{r \binom{K}{t}}$ bits.*

Therefore, the proposed scheme is in fact composed of $(t - i)$ shared-link models [3] each with N files of size $F' = \binom{r}{i} F / \binom{K}{t}$ bits and $K' = r$ users with caches of size $M' = Ni/r$ units each. The corresponding parameter for each model is found to be $t' = \frac{K'M'}{N} = i$. To ensure the existence of sub-files in the second category, the partition integer must satisfy $i \in [\max\{1, t + r - K\}, \min\{t - 1, r - 1\}]$. Hence, for every $i \in [\max\{1, t + r - K\}, \min\{t - 1, r - 1\}]$, summing the achievable rates R_{sl} , which is defined as follows from [11]:

$$R_{sl} = \frac{\binom{K}{t+1} - \binom{K-N_e(\mathbf{d})}{t+1}}{\binom{K}{t}}, \tag{19}$$

of each $b \in \hat{\mathcal{B}}$ shared-link sub-system and replacing the shared-link system parameters F, K, M, t , and $N_e(\mathbf{d})$ with F', K', M', t' , and $N_e(D_{\mathcal{R}})$, respectively, we obtain (18).

We now prove that for each partition integer $i \in [\max\{1, t + r - K\}, \min\{t - 1, r - 1\}]$, each requester k is able to decode the needed sub-files in the second category with the partition integer i upon receiving the codewords $X_b^{2nd, d_k, \mathcal{B} \cup \mathcal{B} \setminus \{k\}}$ for all $b \in \hat{\mathcal{B}}$.

When k is a leading requester of the user b who does not request, i.e., $k \in \mathcal{U}^b$, it can decode any required sub-piece $W_{d_k, \mathcal{B} \cup \mathcal{P}^k, b}$, where $\mathcal{P}^k \subseteq \mathcal{R} \setminus \{k\}$, $|\mathcal{P}^k| = i$, from $X_b^{2nd, d_k, \mathcal{B} \cup \mathcal{P}^k}$, which is broadcast from user b by performing

$$W_{d_k, \mathcal{B} \cup \mathcal{P}^k, b} = \left(\bigoplus_{x \in \mathcal{P}^k} W_{d_x, \mathcal{B} \cup \mathcal{P}^k \cup \{k\} \setminus \{x\}, b} \right) \oplus X_b^{2nd, d_k, \mathcal{B} \cup \mathcal{P}^k},$$

as can be seen from (17).

When $k \notin \mathcal{U}^b$, it is less straightforward for the non-leading requester k to decode the needed sub-files, because not all of the corresponding codewords $X_b^{2nd, d_k, \mathcal{B} \cup \mathcal{P}^k}$ for its required sub-pieces $W_{d_k, \mathcal{B} \cup \mathcal{P}^k, b}$ are directly broadcast from user b . However, Requester k can generate these codewords simply based on the codewords received. To show this, we reformulate the following Lemma 2 from [11] (Lemma 1), which is applied to the codewords broadcast by the user b with the partition integer i .

Lemma 2. *Given an integer t , a partition integer i , a subset $\hat{\mathcal{B}} \subseteq \mathcal{K} \setminus \mathcal{R}$ of size $t - i$, a user $b \in \hat{\mathcal{B}}$, and a set of leading requesters \mathcal{U}^b , for any subset $\mathcal{C}^b \subseteq \mathcal{R}$ that includes \mathcal{U}^b , let \mathcal{V}_F^b be family of all subsets \mathcal{V}^b of \mathcal{C}^b such that each requested file in $D_{\mathcal{R}}$ is requested by exactly one user in \mathcal{V}^b . The following equation holds:*

$$\bigoplus_{\mathcal{V}^b \in \mathcal{V}_F^b} X_b^{2nd, d_k, \mathcal{B} \cup \{\mathcal{C}^b \setminus \mathcal{V}^b\} \setminus \{k\}} = 0, \tag{20}$$

if each $X_b^{2nd, d_k, \mathcal{B} \cup \{\mathcal{C}^b \setminus \mathcal{V}^b\} \setminus \{k\}}$ is defined in (17).

Proof. As we mentioned in Remark 3, for the sub-files needed in the second category, when the parameters t and i are fixed, the proposed scheme, in fact, corresponds to $(t - i)$ shared-link schemes. Thus, Ref. [11] (Lemma 1) can directly be applied to each b -th shared-link scheme. \square

Let us now consider any subset $\bar{\mathcal{B}}$ of $i + 1$ non-leading requesters of user b such that $\bar{\mathcal{B}} \cap \mathcal{U}^b = \emptyset$. Using (20), the following equation can be derived:

$$X_b^{2nd, d_k, \mathcal{B} \cup \bar{\mathcal{B}} \setminus \{k\}} = \bigoplus_{\mathcal{V}^b \in \mathcal{V}_F^b \setminus \{\mathcal{U}^b\}} X_b^{2nd, d_k, \mathcal{B} \cup \{\mathcal{C}^b \setminus \mathcal{V}^b\} \setminus \{k\}}, \tag{21}$$

where $\mathcal{C}^b = \mathcal{B}^i \cup \mathcal{U}^b$. Equation (21) shows that the codeword $X_b^{2nd, d_k, \mathcal{B} \cup \bar{\mathcal{B}} \setminus \{k\}}$ can be directly computed from the broadcast codewords transmitted to all the leading requesters of b , because all codewords on the RHS of (21) are directly broadcasted by user b . Hence, each requester k can obtain the value $X_b^{2nd, d_k, \mathcal{B} \cup \bar{\mathcal{B}} \setminus \{k\}}$ for any subset $\bar{\mathcal{B}}$ of $i + 1$ requesters and can decode its demanded sub-pieces as discussed before. Hence, Lemma 1 is proved.

(iii) Lastly, we consider the sub-files in the third category. Since all the sub-files needed for delivery are only cached in the requesters, the transmission will happen only among requesters. This is equivalent to the D2D coded caching model considered in [9], and we adopt its achievable scheme with uncoded cache placement and one-shot delivery, which we call the *Yapar–Wan–Schaefer–Caire (YWSC) scheme*.

More specifically, during the delivery phase, each sub-file is divided into t equal-length disjoint sub-pieces of $\frac{F}{t \binom{K}{i}}$ bits, which are denoted by $W_{n, \mathcal{T}, i}$, $i \in \mathcal{T}$. Further, each user i

in requester set \mathcal{R} , where $|\mathcal{R}| = r$, selects an arbitrary subset of $N_e(D_{\mathcal{R}\setminus\{i\}})$ users from $\mathcal{R} \setminus \{i\}$, denoted by $\mathcal{U}^i = \{u_1^i, \dots, u_{N_e(D_{\mathcal{R}\setminus\{i\}})}^i\}$, which request $N_e(D_{\mathcal{R}\setminus\{i\}})$ distinct files and are referred to as *leading demanders of user i*. Then, for all subsets $\mathcal{E}^i \subseteq \mathcal{R} \setminus \{i\}$ of t users, each user i transmits

$$X_i^{3rd} = \{Y_{\mathcal{E}^i}^i\}_{\mathcal{E}^i \cap \mathcal{U}^i \neq \emptyset}, \tag{22}$$

where

$$Y_{\mathcal{E}^i}^i = \bigoplus_{k \in \mathcal{E}^i} W_{d_k, \{\mathcal{E}^i \cup \{i\}\} \setminus \{k\}, i}. \tag{23}$$

In other words, since all users $k \in \mathcal{E}^i$ shall retrieve the needed sub-pieces $W_{d_k, \{\mathcal{E}^i \cup \{i\}\} \setminus \{k\}, i}$ from the transmissions of User i , using the idea of leaders from [11], each user i only needs to transmit in sequence the codewords $Y_{\mathcal{E}^i}^i$ for all subsets \mathcal{E}^i such that $\mathcal{E}^i \cap \mathcal{U}^i \neq \emptyset$, i.e., X_i^{3rd} .

As a result, when User k is a leading demander for User i , i.e., $k \in \mathcal{U}^i$, it can decode any needed sub-piece $W_{d_k, \mathcal{G}^k \cup \{i\}, i}$, where $\mathcal{G}^k \subseteq \mathcal{R} \setminus \{i, k\}$, $|\mathcal{G}^k| = t - 1$, from $Y_{\mathcal{G}^k \cup \{k\}}^i$, which is transmitted from User i by performing

$$W_{d_k, \mathcal{G}^k \cup \{i\}, i} = \left(\bigoplus_{x \in \mathcal{G}^k} W_{d_x, \{\mathcal{G}^k \cup \{i, k\}\} \setminus \{x\}, i} \right) \oplus Y_{\mathcal{G}^k \cup \{k\}}^i.$$

When User k is not a leading demander for User i , using the equation

$$\bigoplus_{\mathcal{V}^i \in \mathcal{V}_F^i} Y_{\mathcal{C}^i \setminus \mathcal{V}^i}^i = 0,$$

proved in [9] (Lemma 1), where subset $\mathcal{C}^i \subseteq \mathcal{R} \setminus \{i\}$ includes \mathcal{U}^i and \mathcal{V}_F^i denotes the family of all subsets \mathcal{V}^i of \mathcal{C}^i such that each requested file in $D_{\mathcal{R}\setminus\{i\}}$ is requested by exactly one user in \mathcal{V}^i , each user k can decode its requested sub-piece through obtaining the value $Y_{\mathcal{E}^i}^i$, for any subset \mathcal{E}^i of t users such that $\mathcal{E}^i \cap \mathcal{U}^i = \emptyset$, from the broadcast codewords by the following equation:

$$Y_{\mathcal{E}^i}^i = \bigoplus_{\mathcal{V}^i \in \mathcal{V}_F^i \setminus \{\mathcal{U}^i\}} Y_{\mathcal{C}^i \setminus \mathcal{V}^i}^i, \tag{24}$$

where $\mathcal{C}^i = \mathcal{E}^i \cup \mathcal{U}^i$. To sum up, for each $i \in \mathcal{R} \setminus \{k\}$, User k decodes its requested sub-pieces by following either one of the strategies above, depending on whether it is a leading demander of User i or not.

For each user $i \in \mathcal{R}$, the size of the transmitted signal amounts to $\binom{r-1}{t} - \binom{r-1-N_e(D_{\mathcal{R}\setminus\{i\}})}{t}$ times the size of a sub-piece. Hence, the rate of transmitting all the bits for the sub-files in the third category is

$$R^{3rd} = \frac{r \binom{r-1}{t} - \sum_{i \in \mathcal{R}} \binom{r-1-N_e(D_{\mathcal{R}\setminus\{i\}})}{t}}{t \binom{K}{t}}, \tag{25}$$

In other words, the equivalence of the above scheme and that of [9] can be seen in the following remark:

Remark 4. For the sub-files in the third category, the proposed scheme is equivalent to that of [9]. The central server has a library of N files, $\{W_{n, \mathcal{T}} | \mathcal{T} \subset \mathcal{R}, |\mathcal{T}| = t\}$, and each file has $\tilde{F} \triangleq \binom{F}{t} \binom{K}{t}$ bits. There are r D2D users in the system, each requesting a single file. Based on the MAN uncoded symmetric placement scheme adopted, the placement is the same as [9], where each sub-file has the

size of $\frac{F}{\binom{K}{t}}$, which is equal to $\frac{\tilde{F}}{\binom{K}{t}}$, and sub-file $W_{n,\mathcal{T}}$ is placed in users in \mathcal{T} . Thus, each user caches a total of $N\binom{r-1}{t-1}$ sub-files. Hence, each user has a memory of

$$\tilde{M} = N\binom{r-1}{t-1}\frac{F}{\binom{K}{t}}$$

bits, and we can check that

$$\frac{r\tilde{M}}{N\tilde{F}} = t,$$

just as in [9]. Hence, for the sub-files in the third category, the equivalence between the proposed cache placement scheme and that of [9] is established. In other words, corresponding to the parameters $(F, N, K, MF, t, \text{and } Ne(\mathbf{d}))$ in [9], which denote the file size, number of files, number of users, cache size of each user, the parameter defined as $t = \frac{KMF}{NF}$, and the number of users requesting different files, respectively, we have $(\tilde{F}, N, r, \tilde{M}, t, \text{and } Ne(D_{\mathcal{R}}))$ in our problem.

Summing up the three sub-schemes mentioned above, every requester can decode its requested file. Meanwhile, combining the rate from (16), (18), and (25), the total rate of this delivery scheme is $R_{\text{req-rob}}^{D_{\mathcal{R}}} = R^{1\text{st}} + R^{2\text{nd}} + R^{3\text{rd}}$, which results in the rate stated in (14). The description of this subsection with the explicit characterization of the maximum worst-case delivery rate in Section 4.2, directly proves Theorem 2. More specifically, since the proposed scheme is symmetric with respect to the users, any \mathcal{R} with $|\mathcal{R}| = r$ will offer the same average delivery rate $R_{\text{ave,req-rob}}^{\mathcal{R}}$ and the same worst-case delivery rate $R_{\text{worst,req-rob}}^{\mathcal{R}}$.

Remark 5. The difference in transmitting the sub-files in the three categories is that the transmissions of the sub-files in the first and third categories both adopt the one-shot delivery scheme in [9], while the transmission of the sub-files in the second category adopts the common demands scheme in [11]. Moreover, the transmission of the sub-files in the first category is in uncoded form, while the transmissions of the sub-files in the second and third categories are both with coded multi-casting opportunity.

We formally write the three-category-based scheme in Algorithm 2.

Algorithm 2 Three-category-based Scheme (N, K, M)

procedure PLACEMENT(W_1, \dots, W_N)

1: Apply Algorithm 1 MAN Uncoded Symmetric Placement Scheme $(N, K, M, W_{[N]})$

end procedure

procedure DELIVERY($\mathcal{R}, D_{\mathcal{R}}$)

2: $r \leftarrow |D_{\mathcal{R}}|$

3: $t \leftarrow KM/N$

4: $N_e(D_{\mathcal{R}}) \leftarrow$ the number of distinct elements in $D_{\mathcal{R}}$

5: **for** $k \in [K] \setminus \mathcal{R}$ **do**

6: $\mathcal{U}^{lk} \leftarrow \{u_1^{lk}, \dots, u_{N_e(D_{\mathcal{R}})}^{lk}\}$

7: **end for**

8: (i) For sub-files in the first category:

9: $\mathcal{T} \leftarrow \{\mathcal{T} \subseteq [K] \setminus \mathcal{R} : |\mathcal{T}| = t\}$

10: **for** $\mathcal{T} \in \mathcal{T}$ **do**

11: **for** $n \in [N]$ **do**

12: Divide sub-file $W_{n,\mathcal{T}}$ into t disjoint sub-pieces $(W_{n,\mathcal{T},a} : a \in \mathcal{T})$ with equal size

13: **end for**

14: **for** $a \in \mathcal{T}$ **do**

15: **for** $s \in \mathcal{U}^{la}$ **do**

16: User a transmit $X_a^{1\text{st},d_s,\mathcal{T}} = W_{d_s,\mathcal{T},a}$

```

17:   end for
18: end for
19: end for
20: (ii) For sub-files in the second category:
21: for  $\in [\max\{1, t+r-K\}, \min\{t-1, r-1\}]$  do
22:    $\mathcal{B} \leftarrow \{\hat{\mathcal{B}} \subseteq [K] \setminus \mathcal{R} : |\hat{\mathcal{B}}| = t-i\}$ 
23:   for  $\hat{\mathcal{B}} \in \mathcal{B}$  do
24:     for  $\mathcal{B} \subset \mathcal{R} : |\mathcal{B}| = i$  do
25:       for  $n \in [N]$  do
26:         Divide sub-file  $W_{n, \mathcal{B} \cup \hat{\mathcal{B}}}$  into  $t-i$  disjoint sub-pieces ( $W_{n, \mathcal{B} \cup \hat{\mathcal{B}}, b} : b \in \hat{\mathcal{B}}$ ) with
           equal size
27:       end for
28:     end for
29:     for  $b \in \hat{\mathcal{B}}$  do
30:       for  $\mathcal{B} \subseteq \mathcal{R} : |\mathcal{B}| = i+1$  do
31:         if  $\hat{\mathcal{B}} \cap \mathcal{U}^b == \emptyset$  then
32:           continue
33:         else
34:           User  $b$  transmits  $X_b^{2\text{nd}, d_x, \hat{\mathcal{B}} \cup \mathcal{B} \setminus \{x\}} = \bigoplus_{x \in \hat{\mathcal{B}}} W_{d_x, \hat{\mathcal{B}} \cup \mathcal{B} \setminus \{x\}, b}$ 
35:         end if
36:       end for
37:     end for
38:   end for
39: end for
40: (iii) For sub-files in the third category:
41:  $\mathcal{G} \leftarrow \{\mathcal{G} \subseteq [K] : |\mathcal{G}| = t\}$ 
42: for  $n \in [N]$  do
43:   for  $\mathcal{G} \in \mathcal{G}$  do
44:     Divide sub-file  $W_{n, \mathcal{G}}$  into  $t$  disjoint sub-pieces ( $W_{n, \mathcal{G}, i} : i \in \mathcal{G}$ ) with equal size
45:   end for
46: end for
47: for  $i \in \mathcal{R}$  do
48:    $N_e(D_{\mathcal{R} \setminus \{i\}}) \leftarrow$  the number of distinct elements in  $D_{\mathcal{R} \setminus \{i\}}$ 
49:    $\mathcal{U}^i \leftarrow \{u_1^i, \dots, u_{N_e(D_{\mathcal{R} \setminus \{i\}})}^i\}$ 
50:   for  $\mathcal{E}^i \subseteq \mathcal{R} \setminus \{i\} : |\mathcal{E}^i| = t$  users do
51:     if  $\mathcal{E}^i \cap \mathcal{U}^i == \emptyset$  then
52:       continue
53:     else
54:       User  $i$  transmits  $Y_{\mathcal{E}^i}^i = \bigoplus_{k \in \mathcal{E}^i} W_{d_k, \{\mathcal{E}^i \cup \{i\}\} \setminus \{k\}, i}$ 
55:     end if
56:   end for
57: end for
end procedure

```

4.2. The Maximum Worst-Case Delivery Rate

In this subsection, we characterize the performance of the proposed three-category-based scheme for the maximum worst-case delivery rate. The characterization is based on the observation that the binomial coefficient $\binom{n}{m}$ demonstrates a strictly ascending pattern with respect to n .

For the request-robust D2D coded caching problem, when $N, K, M, r,$ and \mathcal{R} do not change, since the upper bound rate $R_{\text{req-rob}}^{D_{\mathcal{R}}}$ from (14) decreases as $N_e(D_{\mathcal{R}})$ decreases,

the upper bound rate for the maximum worst-case delivery rate is the one at the maximum value of $N_e(D_{\mathcal{R}})$, i.e.,

$$N_e(D_{\mathcal{R}}) = \min\{r, N\}. \tag{26}$$

Then, for $r \geq 2N$, each file can be requested by at least two requesters, which leads to $N_e(D_{\mathcal{R} \setminus \{i\}}), \forall i \in \mathcal{R}$ having the maximum value of N . Hence, this case maximizes the upper bound rate $R_{\text{req-rob}}^{D_{\mathcal{R}}}$.

For $r < 2N$, a requester i may be the only user requesting a file or not, which leads to $N_e(D_{\mathcal{R} \setminus \{i\}}) = N_e(D_{\mathcal{R}}) - 1$ and $N_e(D_{\mathcal{R} \setminus \{i\}}) = N_e(D_{\mathcal{R}})$, respectively. Hence, due to (26), for $r \leq N$, we have $\binom{r-1-N_e(D_{\mathcal{R} \setminus \{i\}})}{t} = 0$, which proves the case where $r \leq N$.

For $N < r < 2N$, we have $N_e(D_{\mathcal{R}}) = N$ and then obtain that each file cannot be requested by more than two requesters. Thus, due to a total of r requesters, there are $2N - r$ requesters, each of which are the only users requesting a file while each of the remaining $2(r - N)$ requesters are not. Thus, we prove the case where $N < r < 2N$.

4.3. Example

To aid in better understanding, we provide an example to illustrate the proposed scheme in Section 4.1.

Let us consider a case where $N = 2, K = 6$, and $M = 2/3$. Hence, $t = KM/N = 2$. In the placement phase, each file is divided into $\binom{6}{2} = 15$ sub-files, and each sub-file's index is in \mathcal{T} , where \mathcal{T} is the family of all sets \mathcal{T} such that $\mathcal{T} \subset [6], |\mathcal{T}| = 2$. The user $k \in [6]$ caches the following sub-files for each $n \in \{1, 2\}$:

$$Z_k = \{W_{n,\mathcal{T}} | \mathcal{T} \in \mathcal{T}, k \in \mathcal{T}\}.$$

In the delivery phase, without a loss of generality, we consider only Users 1, 2, 3, and 4 as requesters, each requesting a single file, i.e., $\mathcal{R} = \{1, 2, 3, 4\}$, and the request vector is $D_{\{1,2,3,4\}} = (1, 2, 1, 1)$. Notice that $r = 4$ and $N_e(D_{\{1,2,3,4\}}) = 2$. Requesters 1, 2, 3, and 4 need the following missing sub-files:

$$\begin{aligned} W_1 \setminus Z_1 &= \{W_{1,\{2,3\}}, W_{1,\{2,4\}}, W_{1,\{2,5\}}, W_{1,\{2,6\}}, W_{1,\{3,4\}}, W_{1,\{3,5\}}, W_{1,\{3,6\}}, W_{1,\{4,5\}}, W_{1,\{4,6\}}, W_{1,\{5,6\}}\}, \\ W_2 \setminus Z_2 &= \{W_{2,\{1,3\}}, W_{2,\{1,4\}}, W_{2,\{1,5\}}, W_{2,\{1,6\}}, W_{2,\{3,4\}}, W_{2,\{3,5\}}, W_{2,\{3,6\}}, W_{2,\{4,5\}}, W_{2,\{4,6\}}, W_{2,\{5,6\}}\}, \\ W_1 \setminus Z_3 &= \{W_{1,\{1,2\}}, W_{1,\{1,4\}}, W_{1,\{1,5\}}, W_{1,\{1,6\}}, W_{1,\{2,4\}}, W_{1,\{2,5\}}, W_{1,\{2,6\}}, W_{1,\{4,5\}}, W_{1,\{4,6\}}, W_{1,\{5,6\}}\}, \\ W_1 \setminus Z_4 &= \{W_{1,\{1,2\}}, W_{1,\{1,3\}}, W_{1,\{1,5\}}, W_{1,\{1,6\}}, W_{1,\{2,3\}}, W_{1,\{2,5\}}, W_{1,\{2,6\}}, W_{1,\{3,5\}}, W_{1,\{3,6\}}, W_{1,\{5,6\}}\}. \end{aligned}$$

In Step (a), determining the leading requesters without a loss of generality, we assume that User 5 picks Users 1 and 2 and User 6 picks Users 2 and 3 as the leading requesters, i.e., $\mathcal{U}^5 = \{1, 2\}, \mathcal{U}^6 = \{2, 3\}$.

In Step (b), we split the sub-files into three categories. More specifically, $W_{n,\{5,6\}}, n \in [N]$ belong to the first category; $W_{n,\{1,5\}}, W_{n,\{1,6\}}, W_{n,\{2,5\}}, W_{n,\{2,6\}}, W_{n,\{3,5\}}, W_{n,\{3,6\}}, W_{n,\{4,5\}}, W_{n,\{4,6\}}, n \in [N]$ belong to the second category; and $W_{n,\{1,2\}}, W_{n,\{1,3\}}, W_{n,\{1,4\}}, W_{n,\{2,3\}}, W_{n,\{2,4\}}, W_{n,\{3,4\}}$ belong to the third category.

In Step (c) of delivering signals, for the sub-files belonging to the first category, i.e., sub-files only cached by Users 5 and 6, after splitting these sub-files into two equal-length sub-pieces, from (15), we know that User 5 transmits

$$W_{1,\{5,6\},5}, W_{2,\{5,6\},5},$$

and User 6 transmits

$$W_{1,\{5,6\},6}, W_{2,\{5,6\},6},$$

which are all directly needed by the requesters. The rate $R^{1st} = 1/30 \times 2 \times 2 = 2/15$, which coincides with (16).

For the sub-files in the second category, since i must satisfy $i \in [\max\{1, t + r - K\}, \min\{t - 1, r - 1\}]$, in this example, we only need to consider $i = 1$, which means that $t - i = 1$. The fact that $t - i = 1$ means that none of these sub-files need to be further split. Take, for example, User 1 in \mathcal{R} ; it requires the sub-file $W_{1,\{2,5\}}$. Thus, $\mathcal{T} = \{2, 5\} = \mathcal{B} \cup \hat{\mathcal{B}}$, where $\mathcal{B} = \{2\}$ and $\hat{\mathcal{B}} = \{5\}$. Consider the set $\bar{\mathcal{B}} = \{1\} \cup \mathcal{B} = \{1, 2\}$ of $i + 1 = 2$ users. Following from (17), if User $\{5\}$ transmits

$$W_{1,\{2,5\}} \oplus W_{2,\{1,5\}},$$

both Users 1 and 2 will be able to obtain the sub-file they want, i.e., $W_{1,\{2,5\}}$ and $W_{2,\{1,5\}}$, respectively. Similarly, we may consider all four users in \mathcal{R} and the sub-files that each of them need. We find that if User $\{5\}$ transmits

$$\begin{aligned} &W_{1,\{2,5\}} \oplus W_{2,\{1,5\}}, W_{1,\{3,5\}} \oplus W_{1,\{1,5\}}, \\ &W_{1,\{4,5\}} \oplus W_{1,\{1,5\}}, W_{2,\{3,5\}} \oplus W_{1,\{2,5\}}, \\ &W_{2,\{4,5\}} \oplus W_{1,\{2,5\}}, W_{1,\{4,5\}} \oplus W_{1,\{3,5\}}, \end{aligned}$$

and User $\{6\}$ transmits

$$\begin{aligned} &W_{1,\{2,6\}} \oplus W_{2,\{1,6\}}, W_{1,\{3,6\}} \oplus W_{1,\{1,6\}}, \\ &W_{1,\{4,6\}} \oplus W_{1,\{1,6\}}, W_{2,\{3,6\}} \oplus W_{1,\{2,6\}}, \\ &W_{2,\{4,6\}} \oplus W_{1,\{2,6\}}, W_{1,\{4,6\}} \oplus W_{1,\{3,6\}}, \end{aligned}$$

all the requesting users will be able to decode the necessary sub-files of the second category. However, recall that $\mathcal{U}^{i5} = \{1, 2\}$ and $\mathcal{U}^{i6} = \{2, 3\}$. Hence, the signal $W_{1,\{4,5\}} \oplus W_{1,\{3,5\}}$ corresponds to $\bar{\mathcal{B}} = \{3, 4, 5\}$, which has zero intersection with \mathcal{U}^{i5} . Hence, $W_{1,\{4,5\}} \oplus W_{1,\{3,5\}}$ need not be transmitted and can be calculated due to the fact that

$$(W_{1,\{3,5\}} \oplus W_{1,\{1,5\}}) \oplus (W_{1,\{4,5\}} \oplus W_{1,\{1,5\}}) \oplus (W_{1,\{4,5\}} \oplus W_{1,\{3,5\}}) = 0.$$

Similarly, $W_{1,\{4,6\}} \oplus W_{1,\{1,6\}}$ need not be transmitted due to the fact that

$$(W_{1,\{3,6\}} \oplus W_{1,\{1,6\}}) \oplus (W_{1,\{4,6\}} \oplus W_{1,\{3,6\}}) \oplus (W_{1,\{4,6\}} \oplus W_{1,\{1,6\}}) = 0.$$

Hence, the rate $R^{2nd} = 1/15 \times 5 \times 2 = 2/3$, which coincides with (18).

For the sub-files in the third category, User 1 has $\mathcal{R} \setminus \{1\} = \{2, 3, 4\}$, which means $N_e(D_{\mathcal{R} \setminus \{1\}}) = 2$. Suppose User 1 picks the leading demanders as $\mathcal{U}^1 = \{2, 4\}$. Similarly, User 2 has $\mathcal{R} \setminus \{2\} = \{1, 3, 4\}$, which means $N_e(D_{\mathcal{R} \setminus \{2\}}) = 1$. Suppose User 2 picks the leading demanders as $\mathcal{U}^2 = \{3\}$. User 3 has $\mathcal{R} \setminus \{3\} = \{1, 2, 4\}$, which means $N_e(D_{\mathcal{R} \setminus \{3\}}) = 2$. Suppose User 3 picks the leading demanders as $\mathcal{U}^3 = \{1, 2\}$, and User 4 has $\mathcal{R} \setminus \{4\} = \{1, 2, 3\}$, which means $N_e(D_{\mathcal{R} \setminus \{4\}}) = 2$. Suppose User 4 picks the leading demanders as $\mathcal{U}^4 = \{2, 3\}$.

Since $t = 2$, we split the sub-files in the third category into two equal-length sub-pieces. For User 1, \mathcal{E}^1 is of size $t = 2$ and is a subset of $\mathcal{R} \setminus \{1\} = \{2, 3, 4\}$. Hence, possible \mathcal{E}^1 s can be $\{2, 3\}$, $\{2, 4\}$, and $\{3, 4\}$, which all satisfy the condition of non-zero intersection with $\mathcal{U}^1 = \{2, 4\}$. Hence, from (22) and (23), User 1 transmits

$$Y_{\{2,3\}}^1 = W_{2,\{1,3\},1} \oplus W_{1,\{1,2\},1}, Y_{\{2,4\}}^1 = W_{2,\{1,4\},1} \oplus W_{1,\{1,2\},1}, Y_{\{3,4\}}^1 = W_{1,\{1,4\},1} \oplus W_{1,\{1,3\},1}.$$

Similarly, User 3 transmits

$$Y_{\{1,2\}}^3 = W_{1,\{2,3\},3} \oplus W_{2,\{1,3\},3}, Y_{\{1,4\}}^3 = W_{1,\{3,4\},3} \oplus W_{1,\{1,3\},3}, Y_{\{2,4\}}^3 = W_{2,\{3,4\},3} \oplus W_{1,\{2,3\},3},$$

and User 4 transmits

$$Y_{\{1,2\}}^4 = W_{1,\{2,4\},4} \oplus W_{2,\{1,4\},4}, Y_{\{1,3\}}^4 = W_{1,\{3,4\},4} \oplus W_{1,\{1,4\},4}, Y_{\{2,3\}}^4 = W_{2,\{3,4\},4} \oplus W_{1,\{2,4\},4}.$$

As for User 2, \mathcal{E}^2 is of size $t = 2$ and is a subset of $\mathcal{R} \setminus \{2\} = \{1, 3, 4\}$. Hence, possible \mathcal{E}^2 s can be $\{1, 3\}$, $\{1, 4\}$, and $\{3, 4\}$. Recall that $\mathcal{U}^2 = \{3\}$; hence, only $\{1, 3\}$ and $\{3, 4\}$ satisfy the condition of non-zero intersection with $\mathcal{U}^2 = \{3\}$. Hence, from (22) and (23), User 2 transmits

$$Y_{\{1,3\}}^2 = W_{1,\{2,3\},2} \oplus W_{1,\{1,2\},2}, \quad Y_{\{3,4\}}^2 = W_{1,\{2,4\},2} \oplus W_{1,\{2,3\},2}.$$

The signal $Y_{\{1,4\}}^2 = W_{1,\{2,4\},2} \oplus W_{1,\{1,2\},2}$ does not need to be transmitted and can be calculated since $Y_{\{1,3\}}^2 \oplus Y_{\{3,4\}}^2 \oplus Y_{\{1,4\}}^2 = 0$ (cf. (24)). Hence, the rate $R^{3rd} = 1/30 \times (3 \times 3 + 2) = 11/30$, which coincides with (25).

Thus, the total delivery rate achieved by the proposed scheme for the case $\mathcal{R} = \{1, 2, 3, 4\}$ and the request vector is $D_{\{1,2,3,4\}} = (1, 2, 1, 1)$ is $\frac{2}{15} + \frac{2}{3} + \frac{11}{30} = \frac{7}{6}$. In this case, if we directly use the YWSC scheme [9] by assigning Users 5 and 6 a demand that is the same as Users 1, 3, and 4, i.e., all six users make file requests, and the demand vector is $\mathbf{d} = \{1, 2, 1, 1, 1, 1\}$, then according to [9] (Equation (14)), the delivery rate is $\frac{6 \times 10 - (3 \times 5 + 6)}{30} = \frac{13}{10}$. If we assign Users 5 and 6 with some other demands, the delivery rate will be even higher. Hence, we see that for the request-robust D2D coded caching problem, the proposed scheme performs better than directly applying the YWSC scheme for this example.

The reason why the proposed three-category-based scheme has better performance is that directly applying the YWSC scheme may contain information that is useful for users who do not request files, which is useless for requesters and difficult to be excluded. We provide more details in the following remarks:

Remark 6. When each user requests a single file ($r = K$), our proposed scheme corresponds to the one originally presented in [9]. The improvement in our scheme is that when $r < K$, we take full advantage of the users who do not request files so that the broadcast codewords are only useful to the requesters. Moreover, through the numerical comparison in Section 5, we find that letting the users who do not request files broadcast pieces of sub-files in the first and second categories, i.e., sub-files in $\{W_{n,\mathcal{A}} | \mathcal{A} \ni \{k\}, k \in \mathcal{K} \setminus \mathcal{R}\}$, incurs a much smaller rate than letting all the users participate in broadcasting the required pieces of sub-files in all three categories.

Remark 7. The proposed scheme is symmetric in the placement phase. As mentioned in [9] (Remark 6), under the constraints of uncoded cache placement, the shared link models in [11,27] showed the optimality of symmetry in the placement phase [3]. This symmetry happens in the placement phase before the requesters are identified and reveal their demands, and any asymmetry in the placement will certainly not result in a better worst-case rate. However, due to the file-categorization step (i.e., Step (b)), the delivery phase of the proposed scheme is asymmetric, while if the value of $N_e(D_{\mathcal{R} \setminus \{i\}})$ is the same for every $i \in \mathcal{R}$, the delivery phase of directly applying the YWSC scheme (i.e., the adapted YWSC scheme (see Appendix A for the specific scheme)) is symmetric. Interestingly, the asymmetric delivery phase of the proposed scheme outperforms the possibly symmetric delivery phase of directly applying the YWSC scheme both for the maximum average and worst-case delivery rate in all cases cited, as shown in Section 5.

Remark 8. The delivery phase of the proposed scheme is actually one-shot, which is defined in [9] as meaning that each user k can recover the i -th needed bit denoted as $W_{d_k}^k(i)$ from its own cache and the transmission of a single other user whose index is $j_k(i)$, i.e., $H(W_{d_k}^k(i) | X_{j_k(i)}, Z_k) = 0$ holds. One-shot delivery allows all users to participate in the transmission without causing users to repeatedly broadcast the same codeword. However, it is difficult to confirm whether the proposed

scheme is optimal under the constraint of uncoded cache placement and one-shot delivery since the delivery scheme as mentioned in Remark 7 is asymmetric.

Remark 9. The rate achieved by the three-category-based scheme from (14) outperforms the rate achieved by the adapted YWSC scheme from (A1) in some specific cases. For example, for $r = 2$ and the maximum worst-case delivery rate, when $t \in [2, K - 1], r \leq N$, we have

$$\begin{aligned}
 R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}} |_{r=2, t \in [2, K-1], r \leq N} &= \frac{K \binom{K-1}{t} - (K-2) \binom{K-2-1}{t} - 2 \binom{K-2}{t}}{t \binom{K}{t}} \tag{27} \\
 &= \frac{(2K-t-1) \binom{K-2}{t-1}}{t \binom{K}{t}} \\
 &> \frac{(2K-t-2) \binom{K-2}{t-1}}{t \binom{K}{t}} \\
 &= \frac{\binom{K-2}{t-1} \binom{2}{1+1}}{\binom{K}{t}} + \frac{2 \binom{K-2}{t}}{\binom{K}{t}} \\
 &= R_{\text{req-rob}}^{D_{\mathcal{R}}} |_{r=2, t \in [2, K-1], r \leq N}, \tag{28}
 \end{aligned}$$

where (27) is from (8) and (28) is from (11). Meanwhile, when $t \in [2, K - 1], r > N$, i.e., $N = 1$, we have

$$\begin{aligned}
 R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}} |_{r=2, t \in [2, K-1], N=1} &= \frac{K \left[\binom{K-1}{t} - \binom{K-2}{t} \right]}{t \binom{K}{t}} \tag{29} \\
 &= \frac{1}{\binom{K}{t}} \cdot \frac{K(K-2)!}{t!(K-t-1)!} \\
 &> \frac{1}{\binom{K}{t}} \cdot \frac{(K-1)!}{t!(K-t-1)!} \\
 &= \frac{\binom{K-2}{t-1} \binom{2}{1+1}}{\binom{K}{t}} + \frac{\binom{K-2}{t}}{\binom{K}{t}} \\
 &= R_{\text{req-rob}}^{D_{\mathcal{R}}} |_{r=2, t \in [2, K-1], N=1}, \tag{30}
 \end{aligned}$$

where (29) is from (8) and (30) is from (11). Moreover, $t \geq K$, i.e., $M \geq N$, is trivial, and when $t = 1$, the three-category-based scheme and the adapted YWSC have the same performance. Hence, for $r = 2$ and the maximum worst-case delivery rate, the three-category-based scheme outperforms the adapted YWSC scheme for all values of K, N, t .

5. Numerical Evaluations

In this section, we compare the rate–memory tradeoff of the three-category-based scheme, the adapted YWSC scheme, and the achievable schemes in [8], adapted to the request-robust D2D coded caching scenario. The adaptation is performed by assigning the users, who do not request, a demand that is most requested by the requesters. We also plot the converse bounds on the optimal average and worst-case delivery rate of the request-robust D2D coded caching problem in Theorem 3.

We consider the cases where the value of K is from 1 to 60. For a fixed K , we consider that the value of N is from 1 to K . The performance metrics are the maximum average delivery rate with respect to the uniform demand and the maximum worst-case delivery rate, i.e., $R_{\text{ave, req-rob}}^r$ and $R_{\text{worst, req-rob}}^r$. We find that, in these cases, our proposed three-category-based scheme outperforms the adapted YWSC scheme and the adapted schemes of [8] for all possible r .

Take the example where $N = 10, K = 30$. As shown in Figure 2, the performance of the proposed scheme is given by the red solid line when $r = 20$ and the purple solid line with dots when $r = 5$. These lines are plotted according to the right-hand side (RHS) of (9) and (10). The performance of the adapted YWSC scheme is given by a blue dash-dot line when $r = 20$ and a cyan dash-dot line with dots when $r = 5$. These lines are plotted according to the RHS of (6) and (7). The proposed converse is given by the black dotted line when $r = 20$ and the orange dotted line with asterisks when $r = 5$. Since the achievable rate in [8] is independent of the demand, the performance of the adapted scheme from [8] does not change with the value of r and is given by the green dashed line. For the maximum worst-case rate, we also provide the lower bound in [8] adapted to the request-robust D2D coded caching scenario with the brown dashed line with dots. It can be seen that our proposed scheme outperforms the adapted YWSC scheme and adapted scheme of [8] in this case; meanwhile, the proposed converse is rather tight compared to the adapted lower bound in [8].

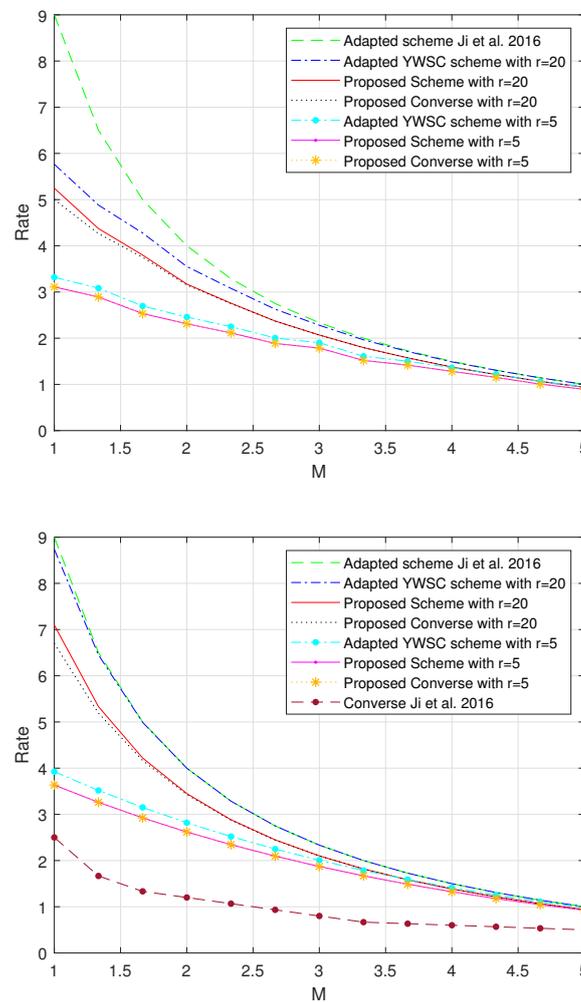


Figure 2. Consider the request-robust D2D coded caching problem from Section 2.1 where $N = 10$ and $K = 30$. The figure above is for the tradeoff between memory size and the maximum worst-case delivery rate for different requester numbers. The figure below shows the tradeoff between memory size and the maximum average delivery rate under uniform demand for different requester numbers. The scheme and converse proposed by Ji et al. [8] are both adapted to this request-robust D2D scenario.

6. Conclusions

In this paper, we propose a new problem called request-robust D2D coded caching, where in the delivery phase, though all users in the placement phase are still present, some of them do not request any files. We presented an achievable scheme for this problem based on uncoded cache placement and exploiting common demands and one-shot delivery. The caching strategy is the same as that proposed by Maddah-Ali and Niesen, while the delivery strategy divides the sub-files into three categories, and different delivery signals are designed for each category. We also characterized information-theoretic lower bounds for the request-robust D2D coded caching problem under the constraint of uncoded cache placement. The lower bounds are both for the maximum average delivery rate under uniform demand and the maximum worst-case delivery rate. We adapt the scheme proposed by Yapar et al. for uncoded cache placement and one-shot delivery to the request-robust D2D coded caching problem. The adaptation is performed by assigning the users, who do not request, a demand that is the most requested by the requesters. The performance of the adapted scheme is proved to be order optimal within a factor of two under uncoded cache placement and within a factor of four in general. Finally, by numerical evaluation, we show that the proposed scheme outperforms the known D2D coded caching schemes applied to the request-robust scenario.

Author Contributions: Conceptualization, W.W., Z.T., N.L. and W.K.; methodology, W.W., Z.T., N.L. and W.K.; formal analysis, W.W., Z.T., N.L. and W.K.; writing—original draft preparation, W.W.; writing—review and editing, W.W., N.L. and W.K.; supervision, N.L. and W.K.; funding acquisition, N.L. and W.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the National Natural Science Foundation of China under Grants 62361146853, 62071115, and 62371129 and the Research Fund of the National Mobile Communications Research Laboratory, Southeast University (no. 2024A03).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Author Zhe Tao was employed by the company Huawei Technologies. The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Adapted Yapar–Wan–Schaefer–Caire Scheme, i.e., Proof of Theorem 1

In this section, we present another achievable scheme that adapts the YWSC scheme from [9] for the request-robust D2D coded caching problem. The scheme achieves the rate stated in Theorem 1.

We will first provide the general achievable scheme, which is based on the uncoded cache placement and exploiting one-shot delivery [9]. Then, we will characterize the performance of the proposed scheme and show that for any requester set \mathcal{R} and corresponding request vector $D_{\mathcal{R}}$, the adapted YWSC scheme achieves the rate

$$R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}} = \frac{K \binom{K-1}{t} - \sum_{i=1}^K \binom{K-1-N_c(D'_{\mathcal{K} \setminus \{i\}})}{t} - f \left[\binom{K-r}{t} - \binom{K-r-1}{t} \right]}{t \binom{K}{t}}, \quad (\text{A1})$$

where $D'_{\mathcal{K}}$ is the request vector of all the users after adaption, and f is an integer equal to one if and only if each requester demands a distinct file; otherwise, $f = 0$. The rate $R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}$ with the explicit characterization of the maximum worst-case delivery rate in Appendix A.2 immediately proves Theorem 1. Finally, we will provide an example to aid in a better understanding of the adapted YWSC scheme.

Appendix A.1. General Scheme

Similar to Section 4.1, for the placement phase, we still use \mathcal{M}_{MAN} described in Definition 1. In the following, we restrict to integer values of $t \in [1 : K]$. For cache size M where $t = KM/N$ is not an integer, memory-sharing will be performed [3,8].

For the delivery phase, let the set of requesters be \mathcal{R} with size r . The r requesters each demand a single file. We find that the YWSC scheme from [9] is highly adaptable for the request-robust D2D coded caching scenario. The adaptation is performed by assigning the users, who do not request, a demand that is the most requested by the requesters. If there are multiple demands that are the most requested, the users, who do not request, would be assigned one of the demands. Recall that we denote the adapted request vector as $D'_{\mathcal{K}}$, and the adapted request vector of users $\mathcal{K} \setminus \{k\}$ as $D'_{\mathcal{K} \setminus \{k\}}$. For example, if $K = N = 4$, $\mathcal{R} = \{1, 2\}$, $D_{\{1,2\}} = \{1, 4\}$, we have $D'_{\{1,2,3,4\}} = \{1, 4, 1, 1\}$ or $D'_{\{1,2,3,4\}} = \{1, 4, 4, 4\}$. Obviously, $N_e(D'_{\mathcal{K}}) = N_e(D_{\mathcal{R}})$ and $N_e(D'_{\mathcal{K} \setminus \{k\}}) \geq N_e(D_{\mathcal{R} \setminus \{k\}})$ for each $k \in \mathcal{R}$.

Next, to adapt the YWSC scheme for the system model, the delivery strategy is divided into the following steps:

- (a) **Determining the leading demanders:** Recall that each sub-file is denoted as $W_{n,\mathcal{T}}$ and is cached by only users in \mathcal{T} . Each sub-file is divided into t equal-length disjoint sub-pieces of $\frac{F}{t \binom{K}{t}}$ bits, which are denoted by $W_{n,\mathcal{T},i}$, $i \in \mathcal{T}$. Further, each user i selects an arbitrary subset of $N_e(D'_{\mathcal{K} \setminus \{i\}})$ users from $\mathcal{K} \setminus \{i\}$, denoted by $\mathcal{U}^i = \{u_1^i, \dots, u_{N_e(D'_{\mathcal{K} \setminus \{i\}})}^i\}$, which request $N_e(D'_{\mathcal{K} \setminus \{i\}})$ distinct files and are referred to as *leading demanders of user i* .

Without a loss of generality, we assume that the leading demanders of each user i are determined from the requesters as much as possible and are denoted by $\bar{\mathcal{U}}^i$. For example, for $K = N = 3$, $\mathcal{R} = \{1, 2\}$, $D_{\{1,2\}} = \{1, 3\}$, $D'_{\{1,2,3\}} = \{1, 3, 1\}$, the leading demander of User 2 in the assumption is $\bar{\mathcal{U}}^2 = \{1\}$ and should not be $\{3\}$, while the leading demander of User 1 is $\bar{\mathcal{U}}^1 = \{2, 3\}$.

- (b) **Pre-transmitting signals:** Then, for all subset $\mathcal{E}^i \subseteq \mathcal{K} \setminus \{i\}$ of t users, each user i prepares for transmitting:

$$X_i^{\text{YWSC}} = \{Y_{\mathcal{E}^i}^i\}_{\mathcal{E}^i \cap \bar{\mathcal{U}}^i \neq \emptyset}, \tag{A2}$$

where

$$Y_{\mathcal{E}^i}^i = \bigoplus_{k \in \mathcal{E}^i} W_{d_{k, \{\mathcal{E}^i \cup \{i\}\} \setminus \{k\}, i}} \tag{A3}$$

which is the same as (23). In other words, since all users $k \in \mathcal{E}^i$ shall retrieve the needed sub-pieces $W_{d_{k, \{\mathcal{E}^i \cup \{i\}\} \setminus \{k\}, i}}$ from the transmissions of User i , using the idea of leaders from [11], each user i only need to transmit in sequence the codewords $Y_{\mathcal{E}^i}^i$ for all subsets \mathcal{E}^i such that $\mathcal{E}^i \cap \bar{\mathcal{U}}^i \neq \emptyset$, i.e., X_i^{YWSC} .

- (c) **Removing unwanted codewords:** We notice that the codewords that are only useful for the users, who do not request, are unwanted codewords and do not need to be transmitted. Next, we will remove the unwanted codewords.

Recall that each user i only prepares for transmitting the codewords $Y_{\mathcal{E}^i}^i$ where $\mathcal{E}^i \cap \bar{\mathcal{U}}^i \neq \emptyset$. For user i , when $\bar{\mathcal{U}}^i \cap \mathcal{K} \setminus \mathcal{R} = \emptyset$, the signal X_i^{YWSC} does not have unwanted codewords. When $\bar{\mathcal{U}}^i \cap \mathcal{K} \setminus \mathcal{R} \neq \emptyset$, the signal X_i^{YWSC} may have the codewords $Y_{\mathcal{E}^i}^i$ where $\mathcal{E}^i \subseteq \mathcal{K} \setminus \mathcal{R}$, which are only useful for the users, who do not request, and should be removed. Obviously, if and only if each requester demands a distinct file, we have $\bar{\mathcal{U}}^i \cap \mathcal{K} \setminus \mathcal{R} = \emptyset$

for only one requester i in \mathcal{R} . Hence, for all subsets $\mathcal{E}^i \subseteq \mathcal{K} \setminus \{i\}$ of t users, each user i transmits

$$X_i^{\text{adapted-YWSC}} = \{Y_{\mathcal{E}^i}^i\}_{\mathcal{E}^i \cap \bar{\mathcal{U}}^i \neq \emptyset, \mathcal{E}^i \not\subseteq \mathcal{K} \setminus \mathcal{R}}$$

where $Y_{\mathcal{E}^i}^i$ is defined in (A3).

As a result, when Requester k is a leading demander for User i , i.e., $k \in \mathcal{U}^i$, it can decode any needed sub-piece $W_{d_k, \mathcal{G}^k \cup \{i\}, i}$, where $\mathcal{G}^k \subseteq \mathcal{K} \setminus \{i, k\}$, $|\mathcal{G}^k| = t - 1$, from $Y_{\mathcal{G}^k \cup \{k\}}^i$, which is transmitted from User i , by performing

$$W_{d_k, \mathcal{G}^k \cup \{i\}, i} = \left(\bigoplus_{x \in \mathcal{G}^k} W_{d_x, \{\mathcal{G}^k \cup \{i, k\}\} \setminus \{x\}, i} \right) \bigoplus Y_{\mathcal{G}^k \cup \{k\}}^i.$$

When Requester k is not a leading demander for User i , using the equation

$$\bigoplus_{\mathcal{V}^i \in \mathcal{V}_F^i} Y_{\mathcal{C}^i \setminus \mathcal{V}^i}^i = 0,$$

proved in [9] (Lemma 1), where subset $\mathcal{C}^i \subseteq \mathcal{K} \setminus \{i\}$ includes $\bar{\mathcal{U}}^i$ and \mathcal{V}_F^i denotes the family of all subsets \mathcal{V}^i of \mathcal{C}^i such that each requested file in $D'_{\mathcal{K} \setminus \{i\}}$ is requested by exactly one user in \mathcal{V}^i , each user k can decode its requested sub-piece through obtaining the value $Y_{\mathcal{E}^i}^i$ for any subset \mathcal{E}^i of t users such that $\mathcal{E}^i \cap \bar{\mathcal{U}}^i = \emptyset$ from the broadcast codewords by the following equation:

$$Y_{\mathcal{E}^i}^i = \bigoplus_{\mathcal{V}^i \in \mathcal{V}_F^i \setminus \{\bar{\mathcal{U}}^i\}} Y_{\mathcal{C}^i \setminus \mathcal{V}^i}^i$$

where $\mathcal{C}^i = \mathcal{E}^i \cup \bar{\mathcal{U}}^i$. To sum up, for each $i \in \mathcal{K} \setminus \{k\}$, Requester k decodes its requested sub-pieces by following either one of the strategies above depending on whether it is a leading demander of User i or not.

For each user $i \in \mathcal{K}$, the size of the transmitted signal amounts to $\binom{K-1}{t} - \binom{K-1-N_e(D'_{\mathcal{K} \setminus \{i\}})}{t}$ times the size of a sub-piece. Meanwhile, if we have $\bar{\mathcal{U}}^i \cap \mathcal{K} \setminus \mathcal{R} \neq \emptyset$ for only one requester i in \mathcal{R} , the size of the removed unwanted codewords amounts to $\binom{K-r}{t} - \binom{K-r-1}{t}$ times the size of a sub-piece. Hence, the rate of transmitting total bits coincides (A1).

For the convenience of understanding, we give the algorithm of the adapted YWSC scheme in Algorithm A1.

Algorithm A1 Adapted YWSC Scheme (N, K, M)

procedure PLACEMENT(W_1, \dots, W_N)

1: Apply Algorithm 1 MAN Uncoded Symmetric Placement Scheme($N, K, M, W_{[N]}$)

end procedure

procedure DELIVERY($\mathcal{R}, D_{\mathcal{R}}$)

2: $r \leftarrow |D_{\mathcal{R}}|$

3: $t \leftarrow KM/N$

4: Assign the users, who do not request, a demand that is the most requested by the requesters and has $D'_{\mathcal{K}} \leftarrow D_{\mathcal{R}}$

5: $\mathcal{G} \leftarrow \{\mathcal{G} \subseteq [K] : |\mathcal{G}| = t\}$

6: **for** $n \in [N]$ **do**

7: **for** $\mathcal{G} \in \mathcal{G}$ **do**

8: Divide sub-file $W_{n, \mathcal{G}}$ into t disjoint sub-pieces ($W_{n, \mathcal{G}, i} : i \in \mathcal{G}$) with equal size

9: **end for**

10: **end for**

11: **for** $i \in \mathcal{K}$ **do**

12: $N_e(D'_{\mathcal{K} \setminus \{i\}}) \leftarrow$ the number of distinct elements in $D'_{\mathcal{K} \setminus \{i\}}$

```

13:  $\bar{U}^i \leftarrow \{u_1^i, \dots, u_{N_e(D'_{\mathcal{K} \setminus \{i\}})}^i\}$  from  $\mathcal{R}$  as much as possible
14: for  $\mathcal{E}^i \subseteq \mathcal{K} \setminus \{i\} : |\mathcal{E}^i| = t$  users do
15:   if  $\mathcal{E}^i \cap \bar{U}^i == \emptyset$  then
16:     continue
17:   else if  $\mathcal{E}^i \subseteq \mathcal{K} \setminus \mathcal{R}$  then
18:     continue
19:   else
20:     User  $i$  transmits  $Y_{\mathcal{E}^i}^i = \bigoplus_{k \in \mathcal{E}^i} W_{d_k, \{\mathcal{E}^i \cup \{i\}\} \setminus \{k\}, i}$ 
21:   end if
22: end for
23: end for
end procedure

```

Appendix A.2. The Maximum Worst-Case Delivery Rate

In this section, we characterize the performance of the adapted YWSC scheme for the maximum worst-case delivery rate. The characterization is based on the observation that the binomial coefficient $\binom{n}{m}$ demonstrates a strictly ascending pattern with respect to n .

Again, the adaptation is performed by assigning the users, who do not request, a demand that is the most requested by the requesters, and we have $N_e(D'_{\mathcal{K}}) = N_e(D_{\mathcal{R}})$. Meanwhile, f from (A1) is an integer equal to one if and only if each requester demands a distinct file, i.e., $N_e(D_{\mathcal{R}}) = r$; otherwise, $f = 0$. Hence, for the request-robust D2D coded caching problem, when N, K, M, r , and \mathcal{R} do not change, since the upper bound rate $R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}$ from (A1) decreases as $N_e(D'_{\mathcal{K}})$ decreases, the upper bound rate for the maximum worst-case delivery rate is the one at the maximum value of $N_e(D'_{\mathcal{K}})$, i.e., $N_e(D'_{\mathcal{K}}) = \min\{r, N\}$.

Then, for $r \geq 2N$, each file can be requested by at least two users, which leads to $N_e(D'_{\mathcal{K} \setminus \{k\}}), \forall k \in \mathcal{K}$ having the maximum value of N . Meanwhile, in this case, $f = 0$. Hence, this case maximizes the upper bound rate $R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}$.

For $r < 2N$, User k may be the only user requesting a file or not, which leads to $N_e(D'_{\mathcal{K} \setminus \{k\}}) = N_e(D'_{\mathcal{K}}) - 1$ and $N_e(D'_{\mathcal{K} \setminus \{k\}}) = N_e(D'_{\mathcal{K}})$, respectively. Due to the adaption, for $r \leq N$, there are $r - 1$ users, each of which are the only users requesting a file while each of the remaining $K - r + 1$ users are not. Meanwhile, in the case $r \leq N$, $N_e(D'_{\mathcal{K}}) = r$ and $f = 1$. Thus, we prove the case where $r \leq N$.

For $N < r < 2N$, we have $N_e(D'_{\mathcal{K}}) = N$. Then, for $D_{\mathcal{R}}$, we obtain that each file cannot be requested by more than two requesters. Thus, due to a total of K users, there are $2N - r$ users, each of which are the only users requesting a file while each of the remaining $K + r - 2N$ users are not. Meanwhile, in this case, $f = 0$. Thus, we prove the case where $N < r < 2N$.

Appendix A.3. Example

To aid in better understanding, we provide an example to illustrate the adapted YWSC scheme in Appendix A.1.

Let us consider a case where $N = 2, K = 4$, and $M = 1$. Hence, $t = KM/N = 2$. In the placement phase, each file is divided into $\binom{4}{2} = 6$ sub-files, and each sub-file's index is in \mathcal{T} , where \mathcal{T} is the family of all sets \mathcal{T} such that $\mathcal{T} \subset [4], |\mathcal{T}| = 2$. The user $k \in [6]$ caches the following sub-files for each $n \in \{1, 2\}$:

$$Z_k = \{W_{n, \mathcal{T}} | \mathcal{T} \in \mathcal{T}, k \in \mathcal{T}\}.$$

In the delivery phase, without a loss of generality, we only consider Users 1 and 2, as requesters, each requesting a single file, i.e., $\mathcal{R} = \{1, 2\}$, and the request vector is $D_{\{1, 2\}} = (1, 2)$. Hence, we consider that the request vector after adaption is

$D'_{\{1,2,3,4\}} = (1, 2, 1, 1)$. Notice that $r = 4$ and $N_e(D'_{\{1,2,3,4\}}) = 2$. Requesters 1 and 2 need the following missing sub-files:

$$W_1 \setminus Z_1 = \{W_{1,\{2,3\}}, W_{1,\{2,4\}}, W_{1,\{3,4\}}\},$$

$$W_2 \setminus Z_2 = \{W_{2,\{1,3\}}, W_{2,\{1,4\}}, W_{2,\{3,4\}}\}.$$

In Step (a), since $t = 2$, we split the sub-files into two equal-length sub-pieces. Further, determining the leading demanders from the requesters as much as possible, without a loss of generality, we assume that User 1 picks Users 1 and 3, User 2 picks User 1, User 3 picks Users 1 and 2, and User 4 picks Users 1 and 2 as the leading demanders, i.e., $\bar{U}^1 = \{2, 3\}$, $\bar{U}^2 = \{1\}$, $\bar{U}^3 = \{1, 2\}$, and $\bar{U}^4 = \{1, 2\}$. Notice that $N_e(D'_{\mathcal{K} \setminus \{1\}}) = 2$, $N_e(D'_{\mathcal{K} \setminus \{2\}}) = 1$, $N_e(D'_{\mathcal{K} \setminus \{3\}}) = 2$, and $N_e(D'_{\mathcal{K} \setminus \{4\}}) = 2$.

In Step (b), for User 1, \mathcal{E}^1 is of size $t = 2$ and is a subset of $\mathcal{K} \setminus \{1\} = \{2, 3, 4\}$. Hence, possible \mathcal{E}^1 s can be $\{2, 3\}$, $\{2, 4\}$, and $\{3, 4\}$, which all satisfy the condition of non-zero intersection with $\bar{U}^1 = \{2, 3\}$. Hence, from (22) and (23), User 1 pre-transmits

$$Y_{\{2,3\}}^1 = W_{2,\{1,3\},1} \oplus W_{1,\{1,2\},1}, Y_{\{2,4\}}^1 = W_{2,\{1,4\},1} \oplus W_{1,\{1,2\},1}, Y_{\{3,4\}}^1 = W_{1,\{1,4\},1} \oplus W_{1,\{1,3\},1}.$$

Similarly, User 3 pre-transmits

$$Y_{\{1,2\}}^3 = W_{1,\{2,3\},3} \oplus W_{2,\{1,3\},3}, Y_{\{1,4\}}^3 = W_{1,\{3,4\},3} \oplus W_{1,\{1,3\},3}, Y_{\{2,4\}}^3 = W_{2,\{3,4\},3} \oplus W_{1,\{2,3\},3},$$

and User 4 pre-transmits

$$Y_{\{1,2\}}^4 = W_{1,\{2,4\},4} \oplus W_{2,\{1,4\},4}, Y_{\{1,3\}}^4 = W_{1,\{3,4\},4} \oplus W_{1,\{1,4\},4}, Y_{\{2,3\}}^4 = W_{2,\{3,4\},4} \oplus W_{1,\{2,4\},4}.$$

As for User 2, \mathcal{E}^2 is of size $t = 2$ and is a subset of $\mathcal{K} \setminus \{2\} = \{1, 3, 4\}$. Hence, possible \mathcal{E}^2 s can be $\{1, 3\}$, $\{1, 4\}$, and $\{3, 4\}$. Recall that $\bar{U}^2 = \{1\}$; hence, only $\{1, 3\}$ and $\{1, 4\}$ satisfy the condition of non-zero intersection with $\bar{U}^2 = \{1\}$. Hence, from (A2) and (A3), User 2 pre-transmits

$$Y_{\{1,3\}}^2 = W_{1,\{2,3\},2} \oplus W_{1,\{1,2\},2}, \quad Y_{\{1,4\}}^2 = W_{1,\{2,4\},2} \oplus W_{1,\{1,2\},2}.$$

The signal $Y_{\{3,4\}}^2 = W_{1,\{2,3\},2} \oplus W_{1,\{2,4\},2}$ does not need to be pre-transmitted and can be calculated since $Y_{\{1,3\}}^2 \oplus Y_{\{1,4\}}^2 \oplus Y_{\{3,4\}}^2 = 0$ (cf. (24)).

In Step (c), removing unwanted codewords, we notice that the codeword $Y_{\{3,4\}}^1$ is only useful for the users, who do not request, and hence are removed. The other codewords are transmitted in the way of Step (b).

Thus, the total delivery rate achieved by the adapted YWSC scheme for the case $\mathcal{R} = \{1, 2\}$ and $D'_{\{1,2,3,4\}} = (1, 2, 1, 1)$ is $R_{\text{adapted-YWSC}}^{\{1,2\}} = \frac{2+3+3+2}{12} = \frac{5}{6}$, which coincides with (A1). In this case, if we directly use the YWSC scheme [9] by assigning Users 3 and 4 a demand that is the same as User 1, then according to [9] (Equation (14)), the delivery rate is $\frac{4 \times 3 - 1}{12} = \frac{11}{12}$. If we assign Users 3 and 4 with some other demands, the delivery rate will be the same. However, if we use the proposed three-category-based scheme from Section 4.1, then according to (A1), the delivery rate is $R_{\text{req-rob}}^{\{1,2\}} = \frac{2}{3}$. Hence, we see that for the request-robust D2D coded caching problem, the proposed three-category-based scheme performs better than the adapted YWSC scheme for this example.

The reason why the proposed three-category-based scheme and the adapted YWSC scheme have better performance is in the following remark:

Remark A1. The achievable rate $R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}$ of our model is similar to the achievable rate $R^*(\mathbf{d}, \mathcal{M}_{\text{MAN}})$ of Yapar's model [9]. The main difference is that the value of $N_e(\mathbf{d}_{\setminus \{i\}})$ in $R^*(\mathbf{d}, \mathcal{M}_{\text{MAN}})$ depends on all the users' request vector \mathbf{d} , while the value of $N_e(D'_{\mathcal{K} \setminus \{i\}})$ in

$R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}$ depends on the r requesters' request vector $D_{\mathcal{R}}$. When $N, K,$ and M is the same, for User i , there is $N_e(D'_{K \setminus \{i\}}) \leq N_e(\mathbf{d}_{\setminus \{i\}})$, and hence $R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}} \leq R^*(\mathbf{d}, \mathcal{M}_{\text{MAN}})$ for uniform demand distribution and the worst case. However, using the adapted YWSC scheme, the signal $X_i^{\text{adapted-YWSC}}$ may contain useless information. For example, in Appendix A.3, the signal $Y_{\{2,3\}}^1$ contains the useless sub-piece $W_{1,\{1,2\},1}$, which is transmitted for User 3, who does not request. We find that removing useless sub-pieces does not affect the size of the rate but makes signals contain too small uncoded sub-pieces. The proposed three-category-based scheme in Section 4 does not contain useless information and maximizes the global gain brought from the code by utilizing the users, who do not request, to transmit signals as much as possible.

Appendix B. Proof of Theorem 3

Appendix B.1. The Maximum Average Delivery Rate

In this subsection, we propose the converse bound given in Theorem 3 for the maximum average delivery rate of the request-robust D2D coded caching problem. Let us consider the problem of coded caching with inactive users defined in Section 2.2.2 first. For a problem of coded caching with inactive users with the rate R_{inact} , given the same placement Z and requester demand vector $D_{\mathcal{R}}$ as in Section 2.1, the encoding function on the central server is

$$\phi^{D_{\mathcal{R}}} : [2^{NF}] \rightarrow [2^{R_{\text{inact}}F}],$$

and the delivery information X' is denoted as $X' = \phi^{D_{\mathcal{R}}}(W_1, \dots, W_N)$. The r requesters decode the request message according to caching content and delivery information, and the decoding function is defined as

$$\psi_k^{D_{\mathcal{R}}} : [2^{M_kF}] \times [2^{R_{\text{inact}}F}] \rightarrow [2^F], \quad k \in \mathcal{R}.$$

The file decoded by User k is denoted as $\hat{W}_{d_k} = \psi_k^{D_{\mathcal{R}}}(X', Z_k)$. If the error probability of this system satisfies

$$\mathbb{P}(\hat{W}_{d_k} \neq W_{d_k}) \leq \epsilon,$$

we call the system ϵ -achievable.

Given $D_{\mathcal{R}}$ and Z , we define the minimum achievable rate of the ϵ -achievable system as $R_{\epsilon, \text{inact}}^{D_{\mathcal{R}}^*}(Z)$. For a fixed \mathcal{R} with size r , $\mathcal{D}^{\mathcal{R}}$ is defined as the set of all possible demands $\{1, \dots, N\}^r$. When $D_{\mathcal{R}}$ is uniformly distributed on $\mathcal{D}^{\mathcal{R}}$, given placement Z , the average delivery rate with respect to the uniform demand $R_{\epsilon, \text{ave, inact}}^{\mathcal{R}^*}(Z)$ is defined as

$$R_{\epsilon, \text{ave, inact}}^{\mathcal{R}^*}(Z) = \mathbb{E}_{D_{\mathcal{R}}} [R_{\epsilon, \text{inact}}^{D_{\mathcal{R}}^*}(Z)].$$

For a given r , we define the maximum average delivery rate with respect to the uniform demand, denoted as $R_{\text{ave, inact}}^r$, where the maximum is over all the request sets \mathcal{R} with size r . Then, the optimal maximum average delivery rate of the memory–rate tradeoff is essentially the maximum average delivery rate for the minimum value given an arbitrary cache size M , and at this rate, the user can decode the requested file with a sufficiently large size without error; that is,

$$R_{\text{ave, inact}}^{r*} = \sup_{\epsilon > 0} \lim_{F \rightarrow +\infty} \sup_Z \min_{\mathcal{R}: |\mathcal{R}|=r} \max R_{\epsilon, \text{ave, inact}}^{\mathcal{R}^*}(Z).$$

Similarly, for a fixed \mathcal{R} with size r , given placement Z , the worst-case delivery rate is

$$R_{\epsilon, \text{worst, inact}}^{\mathcal{R}^*}(Z) = \max_{D_{\mathcal{R}}} R_{\epsilon, \text{inact}}^{D_{\mathcal{R}}^*}(Z),$$

and for a given r , the optimal maximum worst-case delivery rate we want to find is

$$R_{\text{worst,inact}}^{r*} = \sup_{\epsilon > 0} \limsup_{F \rightarrow +\infty} \min_Z \max_{\mathcal{R}:|\mathcal{R}|=r} R_{\epsilon,\text{worst,inact}}^{\mathcal{R}*}(Z).$$

For the problem of coded caching with inactive users, we have the following results:

Theorem A1. For the problem of coded caching with inactive users with K users, a database of N files, uncoded cache sizes of M files at each user, only r users as requesters demanding files during the delivery phase, and parameter $t = \frac{KM}{N}$, we have

$$R_{\text{ave,inact}}^{r*} = \mathbb{E}_{D_{\mathcal{R}}} \left[\frac{\binom{K}{t+1} - \binom{K-N_e(D_{\mathcal{R}})}{t+1}}{\binom{K}{t}} \right], \tag{A4}$$

for $t \in \mathcal{K}$, where $D_{\mathcal{R}}$ is uniformly random on $\mathcal{D}^{\mathcal{R}} = \{1, \dots, N\}^r$ and $N_e(D_{\mathcal{R}})$ denotes the number of distinct requests in $D_{\mathcal{R}}$. When $t \notin \mathcal{K}$, $R_{\text{ave,inact}}^{r*}$ equals the lower convex envelope of the values in (A4) for integer values of $t \in \mathcal{K}$.

Moreover, for the worst-case rate, we have

$$R_{\text{worst,inact}}^{r*} = \frac{\binom{K}{t+1} - \binom{K-\min\{r,N\}}{t+1}}{\binom{K}{t}}, \tag{A5}$$

for $t \in \mathcal{K}$. When $t \notin \mathcal{K}$, $R_{\text{worst,inact}}^{r*}$ equals the lower convex envelope of the values in (A5) for integer values of $t \in \mathcal{K}$.

Proof. The tight lower bounds of the average delivery rate and worst-case delivery rate are derived in the rest of this section. The caching and delivery scheme that achieves the optimal maximum average and worst-case rates is described in Appendix C. \square

Due to inequality (4), to prove Theorem 3, we just need to prove Theorem A1. Motivated by [11], to lower bound the achievable average rate, we first introduce the concept of demand-type division. For the problem of coded caching with inactive users with the number of initial users K , the number of requesters r , and the fixed request set \mathcal{R} , given the request vector $D_{\mathcal{R}}$, we define $s_{\mathcal{R}}(D_{\mathcal{R}})$ as its statistics such that the i^{th} element of $s_{\mathcal{R}}(D_{\mathcal{R}})$ is equal to the number of the i -th most requested file. For example, when $K = 6$ and $N = 4$, and during the delivery phase $r = 4$, $\mathcal{R} = \{1, 2, 3, 4\}$, and $D_{\mathcal{R}} = \{1, 1, 3, 4\}$, the statistic for $D_{\mathcal{R}}$ is $s_{\{1,2,3,4\}}(D_{\mathcal{R}}) = (2, 1, 1, 0)$. For convenience, we simply the statistics $s_{\mathcal{R}}(D_{\mathcal{R}})$ as $s_{\mathcal{R}}$. Then, we denote the set of all possible statistics by $\mathcal{S}_{\mathcal{R}}$. Through this statistical method, the set of request vectors $\mathcal{D}^{\mathcal{R}}$ can be divided into some subsets as type, and type $\mathcal{D}^{s_{\mathcal{R}}}$ is defined as the set of queries with statistics $s_{\mathcal{R}}$.

Note that for each request vector $D_{\mathcal{R}}$, the value of $N_e(D_{\mathcal{R}})$ only depends on its statistic $s_{\mathcal{R}}(D_{\mathcal{R}})$, so for the same type of request vector, the rate is the same. For convenience, the $N_e(D_{\mathcal{R}})$ of the type $\mathcal{D}^{s_{\mathcal{R}}}$ is called $N_e(s_{\mathcal{R}})$.

Given the number of requesters r , the request set \mathcal{R} , and the placement Z , for each type $\mathcal{D}^{s_{\mathcal{R}}}$ in the problem of coded caching with inactive users, the average delivery rate $R_{\epsilon,\text{ave,inact}}^{s_{\mathcal{R}}*}(Z)$ is defined as

$$R_{\epsilon,\text{ave,inact}}^{s_{\mathcal{R}}*}(Z) = \frac{1}{|\mathcal{D}^{s_{\mathcal{R}}}|} \sum_{D_{\mathcal{R}} \in \mathcal{D}^{s_{\mathcal{R}}}} R_{\epsilon,\text{inact}}^{D_{\mathcal{R}}*}(Z),$$

For all types of request vectors, we have

$$R_{\text{ave,inact}}^{r*} = \sup_{\epsilon > 0} \limsup_{F \rightarrow +\infty} \min_Z \max_{\mathcal{R}:|\mathcal{R}|=r} \mathbb{E}_{s_{\mathcal{R}}} \left[R_{\epsilon,\text{ave,inact}}^{s_{\mathcal{R}}*}(Z) \right]$$

$$\geq \sup_{\epsilon > 0} \limsup_{F \rightarrow +\infty} \max_{\mathcal{R}: |\mathcal{R}|=r} \mathbb{E}_{s_{\mathcal{R}}} \left[\min_Z R_{\epsilon, \text{ave}, \text{inact}}^{s_{\mathcal{R}}^*}(Z) \right]. \tag{A6}$$

Hence, the lower bound of $R_{\text{ave}, \text{inact}}^{r^*}$ can be derived through bounding the minimum value of $R_{\epsilon, \text{ave}, \text{inact}}^{s_{\mathcal{R}}^*}(Z)$ for each type $\mathcal{D}^{s_{\mathcal{R}}}$ individually. We use the following lemma to lower bound the average rates within each type:

Lemma A1. Consider the problem of coded caching with inactive users with K initial users, N files, MF cache sizes, r requesters, and request set \mathcal{R} during the delivery phase; for each type $\mathcal{D}^{s_{\mathcal{R}}}$, the minimum value of $R_{\epsilon, \text{ave}, \text{inact}}^{s_{\mathcal{R}}^*}(Z)$ is lower bounded by

$$\min_Z R_{\epsilon, \text{ave}, \text{inact}}^{s_{\mathcal{R}}^*}(Z) \geq \text{Conv} \left(\frac{\binom{K}{t+1} - \binom{K-N_{\epsilon}(s_{\mathcal{R}})}{t+1}}{\binom{K}{t}} \right) - \left(\frac{1}{F} + N_{\epsilon}^2(s_{\mathcal{R}})\epsilon \right), \tag{A7}$$

where $\text{Conv}(f(t))$ denotes the lower convex envelope of the following points: $\{(t, f(t)) | t \in \{0, 1, \dots, K\}\}$.

Proof. We notice that solving the fundamental limits of the problem of coded caching with inactive users is essentially the caching problem defined in [11]. The only difference between the problem of coded caching with inactive users and the caching problem defined in [11] is the size of the request vector. Therefore, applying [11] (Lemma 2) and replacing the caching problem parameters demand d ; statistics of demand s ; set of all possible statistics \mathcal{S} ; set of all possible demands \mathcal{D} ; and type of demand \mathcal{D}_s with $D_{\mathcal{R}}, s_{\mathcal{R}}(D_{\mathcal{R}}), \mathcal{S}_{\mathcal{R}}, \mathcal{D}^{\mathcal{R}}$, and $\mathcal{D}^{s_{\mathcal{R}}}$, we obtain (A7). \square

From (A6) and Lemma A1, the lower bound of $R_{\text{ave}, \text{inact}}^{r^*}$ can be further derived as

$$R_{\text{ave}, \text{inact}}^{r^*} \geq \mathbb{E}_{s_{\mathcal{R}}} \left[\text{Conv} \left(\frac{\binom{K}{t+1} - \binom{K-N_{\epsilon}(s_{\mathcal{R}})}{t+1}}{\binom{K}{t}} \right) \right]. \tag{A8}$$

Because the sequence

$$c_n = \frac{\binom{K}{n+1} - \binom{K-N_{\epsilon}(s_{\mathcal{R}})}{n+1}}{\binom{K}{n}},$$

is convex, the order of the expectation and the Conv in (A8) can be switched. Therefore, $R_{\text{ave}, \text{inact}}^{r^*}$ is lower bounded by the rate defined in Theorem A1, which immediately proves Theorem 3 for the maximum average delivery rate.

Appendix B.2. The Maximum Worst-Case Delivery Rate

Due to (5), to prove Theorem 3 for the maximum worst-case delivery rate, we just need to prove Theorem A1 for the maximum worst-case delivery rate. Given the number of requesters r , the request set \mathcal{R} , and placement Z , for each type $\mathcal{D}^{s_{\mathcal{R}}}$ in the problem of coded caching with inactive users, we define the worst-case delivery rate $R_{\epsilon, \text{worst}, \text{inact}}^{s_{\mathcal{R}}^*}(Z)$ as

$$R_{\epsilon, \text{worst}, \text{inact}}^{s_{\mathcal{R}}^*}(Z) = \max_{D_{\mathcal{R}} \in \mathcal{D}^{s_{\mathcal{R}}}} R_{\epsilon, \text{inact}}^{D_{\mathcal{R}}^*}(Z).$$

Note that

$$\begin{aligned} R_{\text{worst}, \text{inact}}^{r^*} &= \sup_{\epsilon > 0} \limsup_{F \rightarrow +\infty} \min_Z \max_{\mathcal{R}: |\mathcal{R}|=r} \max_{s_{\mathcal{R}}} R_{\epsilon, \text{worst}, \text{inact}}^{s_{\mathcal{R}}^*}(Z) \\ &\geq \sup_{\epsilon > 0} \limsup_{F \rightarrow +\infty} \max_{\mathcal{R}: |\mathcal{R}|=r} \max_{s_{\mathcal{R}}} \min_Z R_{\epsilon, \text{worst}, \text{inact}}^{s_{\mathcal{R}}^*}(Z). \end{aligned}$$

For each $s_{\mathcal{R}} \in \mathcal{S}_{\mathcal{R}}$, using Lemma A1, we can know that

$$\begin{aligned} \min_Z R_{\epsilon, \text{worst, inact}}^{s_{\mathcal{R}}^*}(Z) &\geq \min_Z R_{\epsilon, \text{ave, inact}}^{s_{\mathcal{R}}^*}(Z) \\ &\geq \text{Conv} \left(\frac{\binom{K}{t+1} - \binom{K-N_e(s_{\mathcal{R}})}{t+1}}{\binom{K}{t}} \right) - \left(\frac{1}{F} + N_e^2(s_{\mathcal{R}})\epsilon \right). \end{aligned}$$

When N, K, M, r , and \mathcal{R} do not change, since the worst-case delivery rate $R_{\epsilon, \text{worst, inact}}^{s_{\mathcal{R}}^*}(Z)$ decreases as $N_e(s_{\mathcal{R}})$ decreases, the optimal worst-case delivery rate is the one at the maximum value of $N_e(s_{\mathcal{R}})$. Meanwhile, there is

$$N_e(s_{\mathcal{R}}) \leq \min\{r, N\}. \tag{A9}$$

Consequently,

$$\begin{aligned} R_{\text{worst, inact}}^{r^*} &\geq \sup_{\epsilon > 0} \limsup_{F \rightarrow +\infty} \max_{\mathcal{R}: |\mathcal{R}|=r} \max_{s_{\mathcal{R}}} \text{Conv} \left(\frac{\binom{K}{t+1} - \binom{K-N_e(s_{\mathcal{R}})}{t+1}}{\binom{K}{t}} \right) - \left(\frac{1}{F} + N_e^2(s_{\mathcal{R}})\epsilon \right) \\ &= \text{Conv} \left(\frac{\binom{K}{t+1} - \binom{K-\min\{r, N\}}{t+1}}{\binom{K}{t}} \right). \end{aligned}$$

Therefore, $R_{\text{worst, inact}}^{r^*}$ is lower bounded by the rate defined in Theorem A1 for the maximum worst-case delivery rate, which immediately proves Theorem 3 for the maximum worst-case delivery rate.

Appendix C. Optimal Scheme for Problem of Coded Caching with Inactive Users Achieved in Theorem A1

In this section, we present the optimal scheme for the problem of coded caching with inactive users. The scheme achieves the rate stated in Theorem A1. We will characterize the performance of the proposed scheme and show that for any requester set \mathcal{R} and corresponding request vector $D_{\mathcal{R}}$, the optimal scheme achieves the rate

$$R_{\text{inact}} = \frac{\binom{K}{t+1} - \binom{K-N_e(D_{\mathcal{R}})}{t+1}}{\binom{K}{t}}. \tag{A10}$$

The rate R_{inact} with the explicit characterization of the maximum worst-case delivery rate in (A9) immediately proves Theorem A1.

Similar to Section 4.1, we restrict to integer values of $t \in \mathcal{K}$ and use the MAN uncoded symmetric placement scheme given in Definition 1 in the placement phase. For non-integer values of t , the pair (M, R_{inact}) achieves the lower convex envelope of the achievable points for integer values of $t \in \mathcal{K}$.

For the delivery phase, the central server arbitrarily selects a subset of $N_e(D_{\mathcal{R}})$ requesters, denoted by $\hat{\mathcal{U}} = \{\hat{u}_1, \dots, \hat{u}_{N_e(D_{\mathcal{R}})}\}$, that requests $N_e(D_{\mathcal{R}})$ distinct files, $\hat{\mathcal{U}} \subseteq \mathcal{R}$. Following the idea of leaders from [11], we name these requesters as *leaders*.

Given an arbitrary subset \mathcal{H} of $t + 1$ users, each requester $k \in \mathcal{H} \cap \mathcal{R}$ needs the sub-file $W_{d_k, \mathcal{H} \setminus \{k\}}$, which is known by all other users in \mathcal{H} . Precisely, all the requesters $k \in \mathcal{H} \cap \mathcal{R}$ shall retrieve the needed sub-files $W_{d_k, \mathcal{H} \setminus \{k\}}$ from the transmissions of the central server. By letting the central server broadcast the codeword

$$Y_{\mathcal{H}} = \bigoplus_{x \in \mathcal{H} \cap \mathcal{R}} W_{d_x, \mathcal{H} \setminus \{x\}}, \quad \text{where } \mathcal{H} \cap \mathcal{R} \neq \emptyset, \tag{A11}$$

this sub-file retrieval can be accomplished since each requester $k \in \mathcal{H} \cap \mathcal{R}$ has all the sub-files on the RHS of (A11) except for $W_{d_k, \mathcal{H} \setminus \{k\}}$.

In order to achieve the rate in (A10), the central server only needs to transmit the codeword $Y_{\mathcal{H}}$, which is useful for at least one leader, i.e., $X^{\text{inact}} = \{Y_{\mathcal{H}}\}_{\mathcal{H} \cap \hat{\mathcal{U}} \neq \emptyset}$. The above delivery scheme totally transmits $\binom{K}{t+1} - \binom{K-N_e(D_{\mathcal{R}})}{t+1}$ codewords each with a size of $F/\binom{K}{t}$ bits, which achieves the rate in (A10).

Remark A2. We notice that the proposed scheme is in fact the shared-link model [3] with the same file number N , file size F , user number K , cache size M , and corresponding parameter t . The only difference is the size of the request vector. Replacing the shared-link system parameter $N_e(\mathbf{d})$ in (19) with $N_e(D_{\mathcal{R}})$, we obtain (A10).

We now prove that each requester can decode the file requested by the above delivery scheme. When k is a leader, i.e., $k \in \hat{\mathcal{U}}$, it can decode any required sub-file $W_{d_k, \mathcal{H}'}$, where $\mathcal{H}' \not\ni k, |\mathcal{H}'| = t$, from $Y_{\mathcal{H}' \cup \{k\}}$ by performing

$$W_{d_k, \mathcal{H}'} = \left(\bigoplus_{x \in \mathcal{H}'} W_{d_x, \mathcal{H}' \cup \{k\} \setminus \{x\}} \right) \oplus Y_{\mathcal{H}' \cup \{k\}}, \tag{A12}$$

where $Y_{\mathcal{H}' \cup \{k\}}$ is defined in (A11).

When $k \in \mathcal{R} \setminus \hat{\mathcal{U}}$, not all codewords $Y_{\mathcal{H}' \cup \{k\}}$ are transmitted, and the requester k needs to decode the required codewords not transmitted directly. We use Lemma A2 to explain that the non-leader requester k can also decode all the required sub-files, even if the central server does not transmit $Y_{\mathcal{H}' \cup \{k\}}$, where $\mathcal{H}' \cap \hat{\mathcal{U}} = \emptyset$.

Lemma A2. Given the request vector $D_{\mathcal{R}}$ and picking a set of leaders $\hat{\mathcal{U}}$, for any set $\mathcal{I} \subseteq \mathcal{K}$, let \mathcal{V}_F be the family of all subsets \mathcal{V} of \mathcal{I} such that each requested file in $D_{\mathcal{R}}$ is requested by exactly one user in \mathcal{V} , and we have

$$\bigoplus_{\mathcal{V} \in \mathcal{V}_F} Y_{\mathcal{I} \setminus \mathcal{V}} = 0, \tag{A13}$$

where $Y_{\mathcal{I} \setminus \mathcal{V}}$ is defined in (A11).

Proof. As mentioned in Remark A2, the proposed scheme for the problem of coded caching with inactive users is actually the shared-link scheme. Thus, Lemma 1 in [11] can be directly applied to the proposed scheme. \square

Consider any subset \mathcal{H} of $t + 1$ non-leader users but containing requesters. From Lemma A2, the codeword $Y_{\mathcal{H}}$ can be directly computed from the transmitted codewords by using the following equation:

$$Y_{\mathcal{H}} = \bigoplus_{\mathcal{V} \in \mathcal{V}_F \setminus \{\hat{\mathcal{U}}\}} Y_{\mathcal{I} \setminus \mathcal{V}}, \tag{A14}$$

where $\mathcal{I} = \mathcal{H} \cup \hat{\mathcal{U}}$, given the fact that all codewords on the RHS of (A14) are broadcast, because each $\mathcal{I} \setminus \mathcal{V}$ has a size of $t + 1$ and contains at least one leader. Hence, each requester k can obtain the value $Y_{\mathcal{H}}$ for any subset \mathcal{H} , where $\mathcal{H} \cap \mathcal{R} \neq \emptyset$ of $t + 1$ users, and can decode its demanded files as discussed in (A11). The proposed scheme for the problem of coded caching with inactive users achieves the rate in (A10), which proves the achievability of Theorem A1.

For the convenience of understanding, we give the algorithm of the caching and delivery scheme with inactive users in Algorithm A2.

Algorithm A2 Caching and Delivery Scheme with Inactive Users(N, K, M)

procedure PLACEMENT(W_1, \dots, W_N)

1: Apply Algorithm 1 MAN Uncoded Symmetric Placement Scheme($N, K, M, W_{[N]}$)

end procedure

```

procedure DELIVERY( $\mathcal{R}, D_{\mathcal{R}}$ )
2:  $t \leftarrow KM/N$ 
3:  $N_e(D_{\mathcal{R}}) \leftarrow$  the number of distinct elements in  $D_{\mathcal{R}}$ 
4:  $\hat{\mathcal{U}} \leftarrow \{\hat{u}_1, \dots, \hat{u}_{N_e(D_{\mathcal{R}})}\}$ 
5: for  $\mathcal{H} \subseteq [K] : |\mathcal{H}| = t + 1$  do
6:   if  $\mathcal{H} \cap \hat{\mathcal{U}} == \emptyset$  then
7:     continue
8:   else
9:     Central server transmits  $Y_{\mathcal{H}} = \bigoplus_{x \in \mathcal{H}: d_x \in D_{\mathcal{R}}} W_{d_x, \mathcal{H} \setminus \{x\}}$ 
10:  end if
11: end for
end procedure

```

We provide an example to explain how the proposed scheme works.

Let us consider a case that $N = 2, K = 3, M = 2/3$, and $t = KM/N = 1$. In the placement phase, each file is divided into $\binom{3}{1} = 3$ sub-files. The users $k \in [3]$ cache the following sub-files:

$$Z_1 = \{W_{1,\{1\}}, W_{2,\{1\}}\}, \quad Z_2 = \{W_{1,\{2\}}, W_{2,\{2\}}\}, \quad Z_3 = \{W_{1,\{3\}}, W_{2,\{3\}}\}.$$

In the delivery phase, only User 1 and 2 as requesters request a single file, i.e., $\mathcal{R} = \{1, 2\}$, and the request vector is $D_{\{1,2\}} = (1, 1)$. Notice that $r = 2$ and $N_e(D_{\{1,2\}}) = 1$.

Without a loss of generality, assume that the central server picks User 1 as the leader, i.e., $\hat{\mathcal{U}} = \{1\}$. Then, the central server transmits the following codewords:

$$Y_{\{1,2\}} = W_{1,\{2\}} \oplus W_{1,\{1\}}, \quad Y_{\{1,3\}} = W_{1,\{3\}},$$

without transmitting $Y_{\{2,3\}} = W_{1,\{3\}}$, this will cause $Y_{\{2,3\}} \oplus Y_{\{1,3\}} = 0$ from (A13). No matter what the leader central server picks, it can transmit one less codeword, denoted as $Y_{\mathcal{H}}$, where $\mathcal{H} \cap \hat{\mathcal{U}} = \emptyset$, through this method.

From the transmitted codewords just mentioned, all the requesters can decode all their needed files by performing (A12). The rate $R_{\text{inact}} = 1/3 \times 2 = 2/3$ and could be directly calculated by (A10).

Remark A3. The proposed scheme for the problem of coded caching with inactive users is with secure delivery [28] for $t \geq 1$ and $r \geq 2$, where the users who do not request and the external wiretapper cannot decode any files since the transmitted codewords only consist of the sub-files needed by requesters, and the sub-files whose index only consists of requesters would not be directly transmitted and cannot be decoded.

Appendix D. Order Optimality of the Adapted YWSC Scheme in Appendix A.1 i.e., Proof of Theorem 4

As discussed in Remark 1, in this section, under the constraint of the uncoded cache placement, we only compare the achievable rate $R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}$ from (A1) with the lower bound rate for $R_{\text{ave,req-rob}}^{r*}$ and $R_{\text{worst,req-rob}}^{r*}$ from (12) and (13), respectively.

For the maximum average delivery rate, from (12), (A4), and (A10), we have that $\mathbb{E}_{D_{\mathcal{R}}}[R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}] \geq R_{\text{ave,req-rob}}^{r*} \geq R_{\text{ave,inact}}^{r*} = \mathbb{E}_{D_{\mathcal{R}}}[R_{\text{inact}}]$. Furthermore, we observe that $R_{\text{inact}} \geq \frac{t}{t+1} R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}$ by the following:

$$\frac{t}{t+1} R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}} \leq \frac{\frac{1}{t+1} K \binom{K-1}{t} - \frac{1}{t+1} \sum_{i=1}^K \binom{K-1-N_e(D'_{\mathcal{K} \setminus \{i\}})}{t}}{\binom{K}{t}}$$

$$\begin{aligned}
&= \frac{\binom{K}{t+1} - \sum_{i=1}^K \frac{1}{K - N_e(D'_{\mathcal{K}\setminus\{i\}})} \binom{K - N_e(D'_{\mathcal{K}\setminus\{i\}})}{t+1}}{\binom{K}{t}} \\
&\leq \frac{\binom{K}{t+1} - \min_i \frac{K}{K - N_e(D'_{\mathcal{K}\setminus\{i\}})} \binom{K - N_e(D'_{\mathcal{K}\setminus\{i\}})}{t+1}}{\binom{K}{t}} \\
&\leq \frac{\binom{K}{t+1} - \binom{K - N_e(D_{\mathcal{R}})}{t+1}}{\binom{K}{t}} \\
&= R_{\text{inact}},
\end{aligned} \tag{A15}$$

where (A15) is because $1 \leq N_e(D_{\mathcal{R}\setminus\{i\}}) \leq N_e(D'_{\mathcal{K}\setminus\{i\}}) \leq N_e(D'_{\mathcal{K}}) = N_e(D_{\mathcal{R}})$ for all $i \in [K]$.

Therefore, we have that

$$\begin{aligned}
\mathbb{E}_{D_{\mathcal{R}}}[R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}] &\geq R_{\text{ave,req-rob}}^{r*} \geq R_{\text{ave,inact}}^{r*} = \mathbb{E}_{D_{\mathcal{R}}}[R_{\text{inact}}] \\
&\geq \frac{t}{t+1} \mathbb{E}_{D_{\mathcal{R}}}[R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}] \geq \frac{1}{2} \mathbb{E}_{D_{\mathcal{R}}}[R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}]
\end{aligned}$$

gives the same result, which is

$$\begin{aligned}
\max_{D_{\mathcal{R}}} R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}} &\geq R_{\text{worst,req-rob}}^{r*} \geq R_{\text{worst,inact}}^{r*} = \max_{D_{\mathcal{R}}} R_{\text{inact}} \\
&\geq \frac{t}{t+1} \max_{D_{\mathcal{R}}} R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}} \geq \frac{1}{2} \max_{D_{\mathcal{R}}} R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}
\end{aligned}$$

which can be achieved for the maximum worst-case delivery rate by using similar steps. These results prove that the achievable rate $R_{\text{adapted-YWSC}}^{D_{\mathcal{R}}}$ is order optimal within a factor of two under the constraint of uncoded cache placement. In addition, by the proved order optimality of the shared-link scheme within a factor of two [29] for allowing coded placement, and as discussed in Remark A2 that the only difference between the delivery scheme with inactive users and a shared-link scheme is the size of the request vector, it immediately proves that the achieved rate is within a factor of four in general and completes the proof of Theorem 4.

References

1. Akpakwu, G.A.; Silva, B.J.; Hancke, G.P.; Abu-Mahfouz, A.M. A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges. *IEEE Access* **2018**, *6*, 3619–3647. [\[CrossRef\]](#)
2. Golrezaei, N.; Molisch, A.F.; Dimakis, A.G.; Caire, G. Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution. *IEEE Commun. Mag.* **2013**, *51*, 142–149. [\[CrossRef\]](#)
3. Maddah-Ali, M.A.; Niesen, U. Fundamental Limits of Caching. *IEEE Trans. Inf. Theory* **2014**, *60*, 2856–2867. [\[CrossRef\]](#)
4. Wei, Q.; Wang, L.; Xu, L.; Tian, Z.; Han, Z. Hierarchical Coded Caching for Multiscale Content Sharing in Heterogeneous Vehicular Networks. *IEEE Trans. Veh. Technol.* **2022**, *71*, 5770–5786. [\[CrossRef\]](#)
5. Maddah-Ali, M.A.; Niesen, U. Decentralized Coded Caching Attains Order-Optimal Memory-Rate Tradeoff. *IEEE/ACM Trans. Netw.* **2015**, *23*, 1029–1040. [\[CrossRef\]](#)
6. Cheng, H.; Li, C.; Xiong, H.; Frossard, P. Optimal decentralized coded caching for heterogeneous files. In Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO), Kos, Greece, 28 August–2 September 2017; pp. 2531–2535. [\[CrossRef\]](#)
7. Zheng, L.; Chen, Q.; Yan, Q.; Tang, X. Decentralized Coded Caching Scheme With Heterogeneous File Sizes. *IEEE Trans. Veh. Technol.* **2020**, *69*, 818–827. [\[CrossRef\]](#)
8. Ji, M.; Caire, G.; Molisch, A.F. Fundamental Limits of Caching in Wireless D2D Networks. *IEEE Trans. Inf. Theory* **2016**, *62*, 849–869. [\[CrossRef\]](#)
9. Yapar, C.; Wan, K.; Schaefer, R.F.; Caire, G. On the Optimality of D2D Coded Caching With Uncoded Cache Placement and One-Shot Delivery. *IEEE Trans. Commun.* **2019**, *67*, 8179–8192. [\[CrossRef\]](#)
10. Wang, W.; Liu, N.; Kang, W. Three-User D2D Coded Caching With Two Random Requesters and One Sender. *IEEE Trans. Commun.* **2023**, *71*, 6967–6978. [\[CrossRef\]](#)
11. Yu, Q.; Maddah-Ali, M.A.; Avestimehr, A.S. The Exact Rate-Memory Tradeoff for Caching With Uncoded Prefetching. *IEEE Trans. Inf. Theory* **2018**, *64*, 1281–1296. [\[CrossRef\]](#)

12. Wan, K.; Sun, H.; Ji, M.; Tuninetti, D.; Caire, G. On the Fundamental Limits of Device-to-Device Private Caching under Uncoded Cache Placement and User Collusion. *IEEE Trans. Inf. Theory* **2022**, *68*, 5701–5729. [[CrossRef](#)]
13. Wan, K.; Sun, H.; Ji, M.; Tuninetti, D.; Caire, G. Device-to-Device Private Caching with Trusted Server. In Proceedings of the ICC 2020–2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6. [[CrossRef](#)]
14. Wan, K.; Sun, H.; Ji, M.; Tuninetti, D.; Caire, G. Novel Converse for Device-to-Device Demand-Private Caching with a Trusted Server. In Proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020; pp. 1705–1710. [[CrossRef](#)]
15. Ibrahim, A.M.; Zewail, A.A.; Yener, A. Device-to-Device Coded-Caching With Distinct Cache Sizes. *IEEE Trans. Commun.* **2020**, *68*, 2748–2762. [[CrossRef](#)]
16. Woolsey, N.; Chen, R.R.; Ji, M. Towards Finite File Packetizations in Wireless Device-to-Device Caching Networks. *IEEE Trans. Commun.* **2020**, *68*, 5283–5298. [[CrossRef](#)]
17. Zhang, X.; Ji, M. Finite-length Analysis of D2D Coded Caching via Exploiting Asymmetry in Delivery. In Proceedings of the 2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC), Oulu, Finland, 4–6 July 2022; pp. 1–5. [[CrossRef](#)]
18. Zewail, A.A.; Yener, A. Device-to-Device Secure Coded Caching. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 1513–1524. [[CrossRef](#)]
19. Awan, Z.H.; Sezgin, A. Fundamental limits of caching in D2D networks with secure delivery. In Proceedings of the 2015 IEEE International Conference on Communication Workshop (ICCW), London, UK, 8–12 June 2015; pp. 464–469. [[CrossRef](#)]
20. Ji, M.; Chen, R.R.; Caire, G.; Molisch, A.F. Fundamental limits of distributed caching in multihop D2D wireless networks. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 2950–2954. [[CrossRef](#)]
21. Lee, M.C.; Ji, M.; Molisch, A.F. Optimal Throughput-Outage Analysis of Cache-Aided Wireless Multi-Hop D2D Networks. *IEEE Trans. Commun.* **2021**, *69*, 2489–2504. [[CrossRef](#)]
22. Tebbi, A.; Sung, C.W. Coded caching in partially cooperative D2D communication networks. In Proceedings of the 2017 9th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Munich, Germany, 6–8 November 2017; pp. 148–153. [[CrossRef](#)]
23. Phatak, A.; Varanasi, M.K. An improved coded caching scheme for partially cooperative D2D networks. In Proceedings of the 2022 IEEE International Symposium on Information Theory (ISIT), Espoo, Finland, 26 June–1 July 2022; pp. 1121–1126. [[CrossRef](#)]
24. Li, J.; Chang, Y. New Constructions of D2D Placement Delivery Arrays. *IEEE Commun. Lett.* **2023**, *27*, 85–89. [[CrossRef](#)]
25. Wang, W.; Liu, N.; Kang, W. Coded Caching in Request-robust D2D Communication Networks. In Proceedings of the 2023 International Symposium on Networks, Computers and Communications (ISNCC), Doha, Qatar, 23–26 October 2023; pp. 1–6. [[CrossRef](#)]
26. Liao, J.; Tirkkonen, O. Coded Caching in Presence of User Inactivity. In Proceedings of the 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 10–13 April 2022; pp. 1437–1442. [[CrossRef](#)]
27. Wan, K.; Tuninetti, D.; Piantanida, P. On the optimality of uncoded cache placement. In Proceedings of the 2016 IEEE Information Theory Workshop (ITW), Cambridge, UK, 11–14 September 2016; pp. 161–165. [[CrossRef](#)]
28. Sengupta, A.; Tandon, R.; Clancy, T.C. Fundamental Limits of Caching With Secure Delivery. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 355–370. [[CrossRef](#)]
29. Yu, Q.; Maddah-Ali, M.A.; Avestimehr, A.S. Characterizing the Rate-Memory Tradeoff in Cache Networks Within a Factor of 2. *IEEE Trans. Inf. Theory* **2019**, *65*, 647–663. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.