On the Fundamental Limits of Device-to-Device Private Caching under Uncoded Cache Placement and User Collusion

Kai Wan, *Member, IEEE,* Hua Sun, *Member, IEEE,* Mingyue Ji, *Member, IEEE,* Daniela Tuninetti, *Fellow, IEEE,* and Giuseppe Caire, *Fellow, IEEE*

***Abstract*—In the coded caching problem, as originally formulated by Maddah-Ali and Niesen, a server communicates via a noiseless shared broadcast link to multiple users that have local storage capability. In order for a user to decode its demanded file from the coded multicast transmission, the demands of all the users must be globally known, which may violate the privacy of the users. To overcome this privacy problem, Wan and Caire recently proposed several schemes that attain coded multicasting gain while simultaneously guarantee information theoretic privacy of the users' demands. In Device-to-Device (D2D) networks, the demand privacy problem is further exacerbated by the fact that each user is also a transmitter, which appears to be needing the knowledge of the files demanded by the remaining users in order to form its coded multicast transmission. This paper shows how to solve this seemingly infeasible problem. The main contribution of this paper is the development of new achievable and converse bounds for D2D coded caching that are to within a constant factor of one another when privacy of the users' demands must be guaranteed even in the presence of colluding users (i.e., when some users share cached contents and demanded file indices). First, a D2D private caching scheme is proposed, whose key feature is the addition of virtual users in the system in order to "hide" the demands of the real users. By comparing the achievable D2D private load with an existing converse bound for the shared-link model without demand privacy constraint, the proposed scheme is shown to be order optimal, except for the very low memory size regime with more files than users. Second, in order to shed light into the open parameter regime, a new achievable scheme and a new converse bound under the constraint of uncoded cache placement (i.e., when each user stores directly a subset of the bits of the library) are developed for the case of two users, and shown to be to within a constant factor of one another for all system parameters. Finally, the two-user converse bound is extended to any number of users by a cut-set type argument. With this new converse bound, the virtual users scheme is shown to be order optimal in all parameter regimes under the constraint of uncoded cache placement and user collusion.**

The results of this paper were presented in parts at the 2020 IEEE International Conference on Communications, Dublin, Ireland, [1], and the 2020 IEEE International Symposium on Information Theory, Los Angeles, California, USA, [2].

K. Wan and G. Caire are with the Electrical Engineering and Computer Science Department, Technische Universität Berlin, 10587 Berlin, Germany (e-mail: kai.wan@tu-berlin.de; caire@tu-berlin.de). The work of K. Wan and G. Caire was partially funded by the European Research Council under the ERC Advanced Grant N. 789190, CARENET.

H. Sun is with the Department of Electrical Engineering, University of North Texas, Denton, TX 76203 (email: hua.sun@unt.edu). The work of H. Sun is supported in part by funding from NSF grants CCF-2007108 and CCF-2045656.

M. Ji is with the Electrical and Computer Engineering Department, University of Utah, Salt Lake City, UT 84112, USA (e-mail: mingyue.ji@utah.edu). The work of M. Ji was supported in part by NSF Awards 1817154 and 1824558.

D. Tuninetti is with the Electrical and Computer Engineering Department, University of Illinois Chicago, Chicago, IL 60607, USA (e-mail: danielat@uic.edu). The work of D. Tuninetti was supported in part by NSF Award 1910309.

I. Introduction

Internet data traffic has grown dramatically in the last decade because of on-demand video streaming. The users' demands concentrate on a relatively limited number of files (e.g., latest films and shows) and that the price of memory components in the devices is usually significantly less than the price of bandwidth. On the above observation, caching becomes an efficient and promising technique for future communication systems [3], which leverages the device memory to store data so that future demands can be served faster.

Coded caching was originally proposed by Maddah-Ali and Niesen (MAN) for shared-link networks [4]. In the MAN model, a server has access to a library of N equal-length files and is connected to K users through an error-free broadcast link. Each user can store up to M files in its cache. A caching scheme includes *placement* and *delivery* phases that are designed so as to minimize the *load* (i.e., the number of files sent on the shared link that suffices to satisfy every possible demand vector). In the original MAN model, no constraint is imposed in order to limit the amount of information that the delivery phase leaks to a user about the demands of the remaining users. Such a privacy constraint is critical in modern broadcast services, such as peer-to-peer networks, and is the focus of this paper.

In order to appreciate the main contributions of our work, in the next sub-section we briefly review the various models of coded caching studied in the literature, which will lead to the new problem formulation in this paper.

A. Brief review of coded caching models

Table I shows relevant known results and new results for various coded caching models. The complete memory-load tradeoff is obtained as the lower convex envelope of the listed points. These results are valid for any system parameters (N,K); other results that may lead to better tradeoffs but only apply to limited parameter regimes are not reported for sake of space.

TABLE I: Achievable loads for various coded caching models. Notation: $\mathbf{d}_{\setminus\{k\}}$ denotes the vector obtained from the demand vector $\mathbf{d}$ by removing the k-th element, and $N_e\left(\mathbf{d}_{\setminus\{k\}}\right)$ gives the number of distinct elements in $\mathbf{d}_{\setminus\{k\}}$.

(M,R)	No Privacy	With Privacy
Shared-link	$\left(t\frac{\mathsf{N}}{\mathsf{K}}, \frac{\binom{\mathsf{K}}{t+1}-\binom{\mathsf{K}-\min(\mathsf{N},\mathsf{K})}{t+1}}{\binom{\mathsf{K}}{t}}\right)$ $t\in[0:\mathsf{K}]$, from [5]	$\left(t\frac{1}{\mathsf{K}}, \frac{\binom{\mathsf{KN}}{t+1}-\binom{\mathsf{KN}-\mathsf{N}}{t+1}}{\binom{\mathsf{KN}}{t}}\right)$ $t\in[\mathsf{KN}]$, from [6], [7]
D2D	$\left(t\frac{\mathsf{N}}{\mathsf{K}}, \max\limits_{\mathbf{d}\in[\mathsf{N}]^{\mathsf{K}}} \frac{\binom{\mathsf{K}-1}{t}-\frac{1}{\mathsf{K}}\sum_{k\in[\mathsf{K}]}\binom{\mathsf{K}-1-N_e\left(\mathbf{d}_{\setminus\{k\}}\right)}{t}}{\binom{\mathsf{K}-1}{t-1}}\right)$ $t\in[\mathsf{K}]$, from [8]	$\left(\frac{\mathsf{N}+t-1}{\mathsf{K}}, \frac{\binom{\mathsf{N}(\mathsf{K}-1)}{t}-\binom{\mathsf{N}(\mathsf{K}-1)-\mathsf{N}}{t}}{\binom{\mathsf{N}(\mathsf{K}-1)}{t-1}}\right)$ $t\in[\mathsf{N}(\mathsf{K}-1)+1]$, Scheme A in this paper

1) Shared-link networks without privacy constraints: In the MAN placement phase, letting $t=\mathsf{KM}/\mathsf{N}\in[0:\mathsf{K}]$ represent the number of times a file can be copied in the network's aggregate memory (excluding the server), each file is partitioned into $\binom{\mathsf{K}}{t}$ equal-length subfiles, each of which is cached by a different t-subset of users. In the MAN delivery phase, each user demands one file. According to the users' demands, the server sends $\binom{\mathsf{K}}{t+1}$ *MAN multicast messages*, each of which has the size of a subfile and is useful to $t+1$ users simultaneously. The load of the MAN coded caching scheme is thus $\mathsf{R}=\frac{\binom{\mathsf{K}}{t+1}}{\binom{\mathsf{K}}{t}}=\frac{\mathsf{K}-t}{t+1}$.[1] The MAN scheme is said to achieve a *global coded caching gain*, also referred to as *multicasting gain*, equal to $t+1$ because the load with uncoded caching $\mathsf{R}_{\text{uncoded}}=\mathsf{K}-t=\mathsf{K}(1-\mathsf{M}/\mathsf{N})$ is reduced by a factor $t+1$. This gain scales linearly with network's aggregate memory size. Yu, Maddah-Ali, and Avestimehr (YMA) in [5] proved that $\binom{\mathsf{K}-N_e(\mathbf{d})}{t+1}$ of the MAN multicast messages are redundant when a file is requested simultaneously by multiple users, where $N_e(\mathbf{d})\in[\min(\mathsf{N},\mathsf{K})]$ is the number of distinct file requests in the demand vector $\mathbf{d}\in[\mathsf{N}]^{\mathsf{K}}$. The YMA scheme is known to be exactly optimal under the constraint of *uncoded cache placement* [5], and order optimal to within a factor of 2 otherwise [9], for both worst-case load and average load when files are requested independently and equally likely. The converse bound under the constraint of uncoded cache placement for the worst-case load was first derived by a subset of the authors in [10], [11] by exploiting the index coding acyclic converse bound in [12]. For the case $\mathsf{N}\geq\mathsf{K}=2$, the exact optimality without constraints on the type of placement was characterized in [13] by a non-trivial converse bound leveraging the symmetries in the coded caching problem.

2) Shared-link networks with privacy constraints: For the successful decoding of an MAN multicast message, the users need to know the composition of this message (i.e., which subfiles are coded together). As a consequence, users are aware of the demands of other users. In practice, schemes that leak information on the demand of a user to other users are highly undesirable. For example, this may reveal critical information on user behavior, and allow user profiling by discovering what types of content the users' request. Shared-link coded caching with private demands, which aims to preserve the privacy of the users' demands from other users, was originally discussed in [14] and formally analyzed in an information-theoretical framework by Wan and Caire in [6]. In the private coded caching model, the information about the cached content of each user is unknown to the other users and the composition of each coded multicast message sent by the server must be broadcasted along with the multicast message itself. Following the private coded caching model in [6], various private schemes were proposed in [6], [7], [15]–[18]. Relevant to this paper is the private coded caching scheme based on virtual user proposed in [6], which operates a MAN scheme as if there were KN users in total, i.e., $\mathsf{NK}-\mathsf{K}$ virtual users in addition to the K real users, and the demands of the virtual users as set such that each of the N files is demanded exactly K times. This choice of demands for the virtual users is such that any real user "appears" to have requested equally likely any of the files from the viewpoint of any other user, which guarantees the privacy of the demands. An improved private caching scheme based on virtual user strategy was proposed in [7], which used the YMA delivery instead of the MAN delivery. Compared to converse bounds for the shared-link model without privacy constraint, it can be shown that this scheme based on virtual users is order optimal in all regimes, except for $\mathsf{K}<\mathsf{N}$ and $\mathsf{M}<\frac{\mathsf{N}}{\mathsf{K}}$ [6].[2]

To the best of our knowledge, the only converse bound that truly accounts for privacy constraints in the system model of [6] was proposed in [19] for the case $\mathsf{K}=\mathsf{N}=2$. By combining the novel converse bound in [19] with existing bounds without privacy constraint, the exact optimality was fully characterized in [19] for $\mathsf{K}=\mathsf{N}=2$.

3) D2D networks without privacy constraints: In practice, the content of the library may have been already distributed across the users' local memories and thus can be delivered through peer-to-peer or Device-to-Device (D2D) communications. The shared-link coded caching model was extended to D2D networks in [20]. In the D2D delivery phase, each user broadcasts packets to all other users as functions of its cached content and the users' demands. The D2D load is the sum of the bits sent by all users normalized by the file length.

[1] In the MAN caching scheme, in order to allow each user to decode its demanded file, the composition of each coded multicast message sent by the server must be broadcasted along with the multicast message itself. This is akin to the "header" in linear network coding, that defines the structure of the linear combination of the files to enable decoding. Such composition requires to broadcast *metadata* along the coded multicast messages. Since the size of the metadata does not scale with the file size, the metadata overhead does not contribute to the load in the limit of large file size.

[2] The problem in this regime can be intuitively understood as follows: for $\mathsf{M}=0$ the achievable load in [6] is N while the converse bound is $\min(\mathsf{K},\mathsf{N})=\mathsf{K}$; the ratio of this two numbers can be unbounded.

With the MAN cache placement where each file can be copied $t \in [0:\mathsf{K}]$ times in the aggregate network memory, the D2D coded caching scheme in [20] further partitions each MAN subfile into t equal-length sub-subfiles. Each user then acts as a shared-link server to convey its assigned sub-subfiles to the remaining users either with the MAN delivery [20] or the YMA delivery [8]. This scheme effectively splits the D2D network into K parallel shared-link models, each having N files and serving $\mathsf{K}-1$ users with memory parameter $t-1$. Yapar et *al.* in [8] proved that this scheme is order optimal to within a factor of 4, and exactly optimal under the constraint of uncoded cache placement and *one-shot delivery* (i.e., in a one-shot delivery, any user can recover any requested bit from the content of its own cache and the transmitted messages by at most one other user).

B. New D2D networks with privacy constraints

In D2D networks, the demand privacy problem is further exacerbated by the fact that each user is also a transmitter, which broadcasts coded multicast transmissions based on its cached content. Based on the intuition developed from the shared-link model, one is tempted to conclude that it is impossible to guarantee privacy in D2D networks as the demand vector knowledge appears to be necessary to design the coded multicast messages. Rather surprisingly, in this paper we show that it is possible to guarantee privacy of the users' demands against the other users also in a D2D setting. In our new D2D private caching model, the placement phase is similar to the shared-link private coded caching model. The delivery phase contains two steps. In the first step, each user broadcasts a query to the other users based on its local cached content and its demand; since the query size does not scale with the file size, this step does not contribute to the load in the limit for large file size. In the second step, after collecting all the queries from all the users, each user broadcasts coded multicast messages as a function of the queries and its cached content. In the large file size regime, the load of the system is defined as the load in the second step of D2D communication. The objective of this paper is to design a D2D private coded caching scheme for K users, N files and memory size $\mathsf{M} \geq \mathsf{N}/\mathsf{K}$ (so that the aggregate cache in the entire network suffices to store the entire library) with minimum transmitted load by all users in the delivery phase, while preserving the privacy of the users' demands against the other users.

In the Private Information Retrieval (PIR) problem [21] the privacy of the user's demand against the servers has been considered. In the PIR setting, a user wants to retrieve a desired file from some distributed non-colluding databases (servers), and the objective is to prevent any server from retrieving any information about the index of the user's demanded file. Recently, the authors in [22] characterized the information-theoretic capacity of the PIR problem by proposing a novel converse bound and a coded PIR scheme based on interference alignment. The T-privacy PIR problem with colluding servers were originally considered in [23], where it is imposed that any T-subset of queries sent from the user cannot reveal any information about the demand. The T-privacy PIR problem with at most T colluding servers where each server has a local coded storage was considered in [24], [25]. Since D2D communications have not been considered in the PIR literature, the D2D caching problem with private demands treated in this paper is not a special case of any existing PIR problem.

C. Contributions

We start by giving the first known information-theoretic formulation of the D2D coded caching problem with demand privacy. Then we organize the main contributions of this paper as follows.

a) **Results for general (N,K) from non-trivial extensions of past works**: we prove a constant gap result for all parameter regimes except for $\mathsf{N} > \mathsf{K}$ and $\mathsf{M} < 2\mathsf{N}/\mathsf{K}$ (i.e., the small memory regime with more files than users). More precisely, we propose:

(a.1) Coded Scheme A (Theorem 1): This scheme carefully combines the idea of introducing virtual users [7] with that of splitting the D2D network into multiple parallel shared links [8].

(a.2) Optimality (Theorem 2): By comparing Scheme A with a converse bound for the shared-link model without the privacy constraint in [9], we prove that Scheme A is order optimal to within a factor of 6 when $\mathsf{N} \geq \mathsf{K}$ and $\mathsf{MK}/\mathsf{N} \geq 2$, and of 12 when $\mathsf{N} < \mathsf{K}$ and $\mathsf{MK}/\mathsf{N} \geq 1$.

b) **Results specifically for the case $\mathsf{K}=2$ under uncoded cache placement**: at this point the regime $\mathsf{N} > \mathsf{K}$ and $\mathsf{MK}/\mathsf{N} \in [1,2)$ is open, which motivates the in-depth study of the simplest open case, namely the two-user case. We prove the first known general converse bound under uncoded cache placement that accounts for privacy constraints and leads to a constant gap result for any number of files and any memory regime. In particular, we propose:

(b.1) Coded Scheme B (Theorem 3): This scheme outperforms Scheme A for the two-user case.

(b.2) New Converse (Theorem 4): Inspired by the converse bounds for non-private shared-link caching models under uncoded cache placement from [11] and for PIR systems from [22], we propose a new converse bound under uncoded cache placement for the two-user case by fully considering the privacy constraint.[3]

(b.3) Optimality (Theorem 5): With the new converse bound, under the constraint of uncoded cache placement and $\mathsf{N} \geq \mathsf{K} = 2$, we show that Scheme B is exactly optimal when $\mathsf{M} \in [\mathsf{N}/2, (\mathsf{N}+1)/2]$ or $\mathsf{M} \in \left[\frac{\mathsf{N}(3\mathsf{N}-5)}{2(2\mathsf{N}-3)}, \mathsf{N}\right]$, and is order optimal to within a factor of 3 (numerical simulations suggest $4/3$) for the remaining memory size regime.

c) **Results for general (N,K) under uncoded cache placement and user collusion**: we leverage the new converse

[3]Our bound is not a generalization of the one for the shared-link private caching model with $\mathsf{N}=\mathsf{K}=2$ in [19], because the proposed converse bound heavily depends on the fact that the transmission of each user is a function of the queries and cached content of this user.

bound for the two-user case in a cut-set type bound and prove a constant gap result for all parameter regime, while at the same time considering a stronger notion of privacy that allows for colluding users. We propose:

(c.1) New Converse (Theorem 6): We extend the proposed two-user converse bound to the K-user system by dividing the K users into two groups, and derive a converse bound under uncoded cache placement and user collusion.

(c.2) Optimality (Theorem 7): Under the constraint of uncoded cache placement and user collusion, Scheme A is shown to be order optimal to within a factor of 18 (numerical simulations suggest $27/2$) when $\mathsf{N} > \mathsf{K}$ and $\mathsf{MK}/\mathsf{N} \in [1, 2)$. This proves that Scheme A is order optimal in all memory regimes (that is, also in the regime that was open under the converse bound for the non-private shared-link model) and it is robust to colluding users.

Remark 1 (The powerfulness of the two-user converse bound)**.** *It is rather surprising and quite remarkable to see that, in the considered D2D private coded caching problem, the converse for the case of $\mathsf{K} = 2$ users combined with a cut-set extension yields the order optimality for any system parameters under the constraint of uncoded cache placement and user collusion. This is in stark contrast to plenty of well-known multiuser information theory problems where the optimality results for the $\mathsf{K} = 2$ case do not generalize, and give in fact little or no hint to the $\mathsf{K} > 2$ case. Paramount examples include the general broadcast channel with degraded message sets [26], [27], the K-user Gaussian interference channel [28], [29], and the non-private shared-link coded caching [13].* □

Remark 2 (Cost of D2D)**.** *By using the result in [30], one can immediately infer that, under the constraint of uncoded cache placement and without privacy constraint, the gap between the achieved loads in the shared-link and D2D scenarios is at most 2. This is no longer the case when privacy is introduced, where the gap between the loads in private shared-link and private D2D scenarios can be arbitrarily large (i.e., the gap is larger than $\mathsf{N}/\min(\mathsf{N}, \mathsf{K})$ when $\mathsf{M} = \mathsf{N}/\mathsf{K}$, which can be unbounded). Similar observations were made in the context of secure shared-link pliable index coding [31], where the authors showed that problems that are feasible without security constraints became infeasible when security is considered (i.e., there is no constant gap factor independent of the system parameters).* □

D. Paper organization

The rest of this paper is organized as follows. Section II formulates the D2D private caching model. Section III provides an overview of all our technical results, and provides some numerical evaluations. Sections IV and V provide proofs of the proposed achievable schemes and converse bounds, respectively. Section VI concludes the paper. Some proofs (i.e., more technical lemmas and tedious gap derivations) may be found in the Appendices.

E. Notation convention

Calligraphic symbols denote sets, bold symbols denote vectors, and sans-serif symbols denote system parameters. We use $|\cdot|$ to represent the cardinality of a set or the length of a vector. Sets of consecutive integers are denoted as $[a:b] := \{a, a+1, \ldots, b\}$ and $[n] := [1:n]$. The symbol $\oplus$ represents bit-wise XOR. $a! = a \times (a-1) \times \cdots \times 1$ represents the factorial of a. We use the convention $\binom{x}{y} = 0$ if $x < 0$ or $y < 0$ or $x < y$.

II. System Model

A (K, N) D2D private caching system comprises the following elements.

- A library with N independently generated files, where each file is composed of B i.i.d. bits. The files are denoted by $(F_1, F_2, \ldots, F_{\mathsf{N}})$.
- K users, each equipped with a local cache.
- An error-free broadcast link from each user to all other users (e.g., a shared medium).[4]

The system operates in two phases.

- Placement Phase. Note that the placement phase is done without knowledge of later demand. Each user $k \in [\mathsf{K}]$ first generates some local randomness P_k, which is independent of the library $F_1, \ldots, F_{\mathsf{N}}$ and independent across users, and is only known at user $k \in [\mathsf{K}]$. Then user k stores Z_k in its cache, where

$$H\big(Z_k | P_k,\ F_1, \ldots, F_N\big) = 0 \quad \text{(placement constraint)}, \tag{1}$$

  The vector of all caches is $\mathbf{Z} := (Z_1, Z_2, \ldots, Z_{\mathsf{K}})$.
- Delivery Phase. User $k \in [\mathsf{K}]$ demands the file indexed by $d_k \in [\mathsf{N}]$. The demand vector is $\mathbf{d} := (d_1, d_2, \ldots, d_{\mathsf{K}})$. The delivery phase contains the following two steps.
  - Step 1: user $k \in [\mathsf{K}]$, given its randomness P_k, cached content Z_k and demand d_k, broadcasts the query ℓ_k to the other users.
  - Step 2: after having received all the queries, user $k \in [\mathsf{K}]$ broadcasts the signal X_k to the other users, where

$$H\big(X_k | Z_k, P_k, \ell_1, \ldots, \ell_{\mathsf{K}}\big) = 0, \ \text{(encoding constraint)}. \tag{2}$$

    Note that, the queries $\ell_1, \ldots, \ell_{\mathsf{K}}$ act as the metadata explained in Footnote 1, implying the composition of each coded multicast message.

  Successful decoding is guaranteed if

$$H\big(F_{d_k} | Z_k, P_k, d_k,\ \ell_1, \ldots, \ell_{\mathsf{K}},\ X_1, \ldots, X_{\mathsf{K}}\big) = 0,$$
$$\forall k \in [\mathsf{K}], \quad \text{(decoding constraint)}. \tag{3}$$

[4]D2D networks may be implemented at the physical/MAC layer, such that the nodes are physical devices sharing a common transmission medium, or at the logical or "application" layer, as for example in current peer-to-peer file sharing systems such as BitTorrent, Gnutella, Kazaa and several others. We do not make such distinction here and just compute the load as the sum of all nodes (or "peers") transmissions expressed in bits, necessary to satisfy the users demands. This load notion is compliant with the previously defined coded caching models for D2D and shared link systems.

Demand privacy[5] is guaranteed if

$$I\big(\mathbf{d}_{[\mathsf{K}]\setminus\{k\}}; Z_k, P_k, d_k,\ \ell_1,\ldots,\ell_{\mathsf{K}},\ X_1,\ldots,X_{\mathsf{K}}\big)=0, \quad \text{(privacy constraint)}, \tag{4}$$

where $\mathbf{d}_{\mathcal{S}}$ denotes the vector obtained from $\mathbf{d}$ by retaining only the elements indexed by $\mathcal{S}$.

Assume that the length of (P_k,ℓ_k), $k\in[\mathsf{K}]$, does not scale with B. By the constraint of privacy, the number of transmissions in Step 2 of the delivery for different demand vectors should be the same. Thus a pair (M,R) is said to be achievable if all the above constraints are satisfied with

$$\limsup_{\mathsf{B}\to\infty}\frac{H(Z_k)}{\mathsf{B}}\le \mathsf{M},\ \forall k\in[\mathsf{K}],\quad \text{(cache size)}, \tag{5a}$$

$$\limsup_{\mathsf{B}\to\infty}\frac{\sum_{k\in[\mathsf{K}]}H(X_k)}{\mathsf{B}}\le \mathsf{R},\quad \text{(load)}. \tag{5b}$$

Our objective is to determine

$$\mathsf{R}^\star(\mathsf{M}):=\inf\{\mathsf{R}:(\mathsf{M},\mathsf{R}) \text{ is achievable as in (5)}\}. \tag{6}$$

We only consider the case $\min(\mathsf{K},\mathsf{N})\ge 2$, since the case $\mathsf{K}=1$, a single node network, does not make sense in a D2D network and when $\mathsf{N}=1$ each user knows the demand of the other users. In addition, we only need to consider $\mathsf{M}\in\left[\frac{\mathsf{N}}{\mathsf{K}},\mathsf{N}\right]$, since for $\mathsf{M}\ge\mathsf{N}$ each user can cache the whole library, thus no delivery is needed; and for $\mathsf{KM}<\mathsf{N}$ there is not enough space in the overall network memory to store the whole library, thus the problem is not feasible.

Uncoded Cache Placement. If each user $k\in[\mathsf{K}]$ directly copies some bits of the files into Z_k, the cache placement is said to be *uncoded*. The optimal load under the constraint of uncoded cache placement is denoted by $\mathsf{R}_{\mathrm{u}}^\star(\mathsf{M})$, which is defined as in (5b) but with the additional constraints that the cache placement phase is uncoded. Clearly, $\mathsf{R}^\star(\mathsf{M})\le\mathsf{R}_{\mathrm{u}}^\star(\mathsf{M})$.

Colluding Users. We say that the users in the system *collude* if they exchange the index of their demanded file and their cached content. Collusion is a natural consideration to increase the privacy level and is one of the most widely studied variants in the PIR problem [23], [32]–[34]. Privacy constraint against colluding users is a stronger notion than (4) and is defined as follows

$$I\big(\mathbf{d}_{[\mathsf{K}]\setminus\mathcal{S}};(Z_k,P_k:k\in\mathcal{S}),\mathbf{d}_{\mathcal{S}},\ell_1,\ldots,\ell_{\mathsf{K}},X_1,\ldots,X_{\mathsf{K}}\big)=0,\ \forall\mathcal{S}\subseteq[\mathsf{K}],\mathcal{S}\neq\emptyset. \tag{7}$$

The optimal load under uncoded cache placement and the privacy constraint in (7) is denoted by $\mathsf{R}_{\mathrm{u,c}}^\star(\mathsf{M})$. Clearly, $\mathsf{R}_{\mathrm{u,c}}^\star(\mathsf{M})\ge\mathsf{R}_{\mathrm{u}}^\star(\mathsf{M})\ge\mathsf{R}^\star(\mathsf{M})$.

Remark 3. *For $\mathsf{K}=2$, the privacy constraints in* (4) *and* (7) *are equivalent, and thus we have $\mathsf{R}_{\mathrm{u,c}}^\star(\mathsf{M})=\mathsf{R}_{\mathrm{u}}^\star(\mathsf{M})\ge\mathsf{R}^\star(\mathsf{M})$.* □

III. Main Results

In this section, we summarize all the new results in this paper and provide the main ingredients on how the bounds are derived.

[5]The privacy constraint in (4) corresponds to perfect secrecy in an information theoretic sense (see [27, Chapter 22]).

A. Results for general (N,K) by non-trivial extensions of known schemes

Inspired by the virtual-user strategy in [7], we propose a private coded caching scheme (referred to as Scheme A in the following) with a cache placement inspired by the D2D strategy [20]. More precisely, our scheme effectively divides the D2D network into K independent shared-link models, each of which serves $\mathsf{U}:=(\mathsf{K}-1)\mathsf{N}$ effective users, where $(\mathsf{K}-1)(\mathsf{N}-1)$ users are virtual. The achieved load is given in the following theorem; an example that highlights the main ingredients in Scheme A can be found in Section IV-A and the detailed general description on Scheme A can be found in Section IV-B.

Theorem 1 (Scheme A)**.** *For the (K,N) D2D private caching system, $\mathsf{R}_{\mathrm{u,c}}^\star$ is upper bounded by the lower convex envelope of the following points*

$$(\mathsf{M},\mathsf{R}_{\mathrm{A}})=\left(\frac{\mathsf{N}+t-1}{\mathsf{K}},\frac{\binom{\mathsf{U}}{t}-\binom{\mathsf{U}-\mathsf{N}}{t}}{\binom{\mathsf{U}}{t-1}}\right),\ \forall t\in[\mathsf{U}+1]. \tag{8}$$

□

Note that Scheme A satisfies the robust privacy constraint in (7) against colluding users. By comparing Scheme A in Theorem 1 and the converse bound for the shared-link caching problem without privacy constraint in [9], we have the following order optimality results, whose proof can be found in Appendix D.

Theorem 2 (Order optimality of Scheme A)**.** *For the (K,N) D2D private caching system, Scheme A in Theorem 1 is order optimal to within a factor of 6 if $\mathsf{N}\ge\mathsf{K}$ and $\mathsf{M}\ge 2\mathsf{N}/\mathsf{K}$, and 12 if $\mathsf{N}\le\mathsf{K}$.* □

Remark 4 (Reduction of Subpacketization for Scheme A)**.** *Scheme A in Theorem 1 divides each file into $\mathsf{K}\binom{\mathsf{U}}{t-1}$ equal-length subfiles, thus the subpacketization is $\mathsf{K}\binom{\mathsf{U}}{t-1}\approx\mathsf{K}2^{\mathsf{U}\mathcal{H}\left(\frac{t-1}{\mathsf{U}}\right)}$, where $\mathcal{H}(p)=-p\log_2(p)-(1-p)\log_2(1-p)$ is the binary entropy function. Hence, the maximal subpacketization of the virtual-user scheme (when $\frac{t-1}{\mathsf{U}}=\frac{1}{2}$) is exponential in U, which is much higher than the maximal subpacketization of the K-user MAN coded caching scheme (which is exponential in K). Very recently, after the original submission of this paper, the authors in [16] proposed a shared-link private coded caching scheme based on the cache-aided linear function retrieval [35], which can significantly reduce the subpacketization of the shared-link virtual-user private caching schemes in [6], [7]. In addition to the cached content by the MAN placement, the authors let each user privately cache some linear combinations of uncached subfiles in the MAN placement which are regarded as keys. In such way, the effective demand of each user in the delivery phase becomes the sum of these linear combinations and the subfiles of its desired file, such that the real remand is concealed. We can directly use the extension strategy in [20] to extend this shared-link private caching scheme to our D2D setting to*

obtain Scheme C, which achieves the lower convex envelope of $\left(\frac{\mathsf{K}}{\mathsf{N}}, \mathsf{N}\right)$ *and the following points*

$$(\mathsf{M}, \mathsf{R}_{\mathrm{C}}) = \left(\frac{t(\mathsf{N}-1)}{\mathsf{K}} + 1, \frac{\binom{\mathsf{K}-1}{t} - \binom{\mathsf{K}-1-\mathsf{N}}{t}}{\binom{\mathsf{K}-1}{t-1}} \right), \ \forall t \in [\mathsf{K}]. \tag{9}$$

The subpacketization of the scheme in (9) *is* $\mathsf{K}\binom{\mathsf{K}}{t} \approx \mathsf{K}2^{\mathsf{K}\mathcal{H}(t/\mathsf{K})}$*, which is the same as the* K*-user non-private D2D coded caching scheme in [20]. As the shared-link private caching scheme in [16], Scheme C also satisfies the robust privacy constraint in* (7) *against colluding users.* □

B. Results for $\mathsf{K} = 2$*: new converse bound to truly account for privacy constraints*

The order optimality results in Theorem 2 is derived from an existing converse bound without privacy constraint and does not cover the regime $\mathsf{N} > \mathsf{K}$ and $\mathsf{M} \in [\mathsf{N}/\mathsf{K}, 2\mathsf{N}/\mathsf{K})$. Hence, we are motivated to derive a new converse bound by fully incorporating the privacy constraint for the simplest open case, that is, for a two-user system.

When $\mathsf{K} = 2$, we observe that in Scheme A some cached content is redundant. By removing those redundancies we derive a new scheme (referred to as Scheme B in the following) whose achieved load is given in the following theorem; an example that highlights the main ingredients in Scheme B can be found in Section IV-C and the detailed general description on Scheme B can be found in Section IV-D.

Theorem 3 (Scheme B)**.** *For the* $(\mathsf{K}, \mathsf{N}) = (2, \mathsf{N})$ *D2D private caching system,* $\mathsf{R}^{\star}_{\mathrm{u}} = \mathsf{R}^{\star}_{\mathrm{u,c}}$ *is upper bounded by the lower convex envelope of* $(\mathsf{M}, \mathsf{R}_{\mathrm{B}}) = (\mathsf{N}, 0)$ *and the following points*

$$(\mathsf{M}, \mathsf{R}_{\mathrm{B}}) = \left(\frac{\mathsf{N}}{2} + \frac{\mathsf{N}t'}{2(\mathsf{N}+t'-1)}, \frac{\mathsf{N}(\mathsf{N}-1)}{(t'+1)(\mathsf{N}+t'-1)} \right), \quad \forall t' \in [0 : \mathsf{N}-1]. \tag{10}$$

□

In Appendix F we prove the following corollary.

Corollary 1. *By comparing Scheme A in Theorem 1 for* $\mathsf{K} = 2$ *and Scheme B in Theorem 3, we find* $\mathsf{R}_{\mathrm{B}} \leq \mathsf{R}_{\mathrm{A}}$*.* □

Next we turn our attention to converse bounds that truly incorporate the privacy constraint. The following converse bound is one of the key novelties of this paper. It truly accounts for the privacy constraint in the general setting $\mathsf{N} \geq 2$. The main idea is to derive several bounds that contain a "tricky" entropy term that needs to be bounded in a non-trivial way; in some bounds this entropy term appears with a positive sign and in others with a negative sign; by linearly combining the bounds, the "tricky" entropy term cancels out. Different from the converse bound in [19] for the shared-link caching with private demands for $\mathsf{N} = \mathsf{K} = 2$, our converse bound focuses on uncoded cache placement and works for any $\mathsf{N} \geq \mathsf{K} = 2$. Theorem 4 is proved in full generality in Section V-B. For the sake of clarity, an example of the key steps in the proof is provided Section V-A for the case of $\mathsf{N} = 2$ files.

Theorem 4 (New converse bound for the two-user system)**.** *For the* $(\mathsf{K}, \mathsf{N}) = (2, \mathsf{N})$ *D2D private caching system where* $\mathsf{N} \geq \mathsf{K} = 2$*, assuming* $\mathsf{M} = \frac{\mathsf{N}}{2} + y$ *where* $y \in \left[0, \frac{\mathsf{N}}{2}\right]$*, we have the following bounds*

$$\mathsf{R}^{\star}_{\mathrm{u}} \geq \mathsf{N} - 2y - \frac{4y + (\mathsf{N}-\mathsf{K}/2)h}{h+2} + \frac{h^2(\mathsf{N}-\mathsf{K}/2) - \mathsf{N}(2\mathsf{N}/\mathsf{K}-3) + h(\mathsf{N}+\mathsf{K}/2)}{(h+1)(h+2)} \frac{2y}{\mathsf{N}}, \quad \forall h \in [0 : \mathsf{N}-3], \ \text{only active for } \mathsf{N} \geq 3, \tag{11}$$

$$\mathsf{R}^{\star}_{\mathrm{u}} \geq \mathsf{K}\left(1 - \frac{3y}{\mathsf{N}}\right), \tag{12}$$

$$\mathsf{R}^{\star}_{\mathrm{u}} \geq \mathsf{K}\left(\frac{1}{2} - \frac{y}{\mathsf{N}}\right). \tag{13}$$

□

By comparing the new converse bound in Theorem 4 and Scheme B in Theorem 3, we have the following optimality result under the constraint of uncoded cache placement (the proof can be found in Appendix G).

Theorem 5 (Optimality for the two-user system)**.** *For the* $(\mathsf{K}, \mathsf{N}) = (2, \mathsf{N})$ *D2D private caching system where* $\mathsf{N} \geq \mathsf{K} = 2$*, Scheme B in Theorem 3 is exactly optimal under the constraint of uncoded cache placement when* $\frac{\mathsf{N}}{2} \leq \mathsf{M} \leq \frac{\mathsf{N}+1}{2}$ *or* $\frac{\mathsf{N}(3\mathsf{N}-5)}{2(2\mathsf{N}-3)} \leq \mathsf{M} \leq \mathsf{N}$*. Otherwise, Scheme B is order optimal to within a factor of* 3 *(numerical simulations suggest* $4/3$*).* □

From Theorem 5, we can directly derive the following corollary.

Corollary 2. *For the* $(\mathsf{K}, \mathsf{N}) = (2, \mathsf{N})$ *D2D private caching system Scheme B in Theorem 3 is exactly optimal under the constraint of uncoded cache placement in all memory regimes when* $\mathsf{N} \in \{2, 3\}$*.* □

C. Order optimality results for any system parameter when users may collude

In Section V-C we extend Theorem 4 to any $\mathsf{K} \geq 2$ with the consideration of the privacy constraint against colluding users in (7). The main idea is to divide the users into two groups in a cut-set-like fashion and generate a powerful aggregate user whose cache contains the caches of all users in each group (implying collusion). The derived converse bound is as follows.

Theorem 6 (New converse bound for the K-user system)**.** *For the* (K, N) *D2D private caching system where* $\mathsf{N} \geq \mathsf{K} \geq 3$*, assuming* $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{2y}{\mathsf{K}}$ *where* $y \in \left[0, \frac{\mathsf{N}}{2}\right]$*, we have*

$$\mathsf{R}^{\star}_{\mathrm{u,c}} \geq \frac{\lfloor \mathsf{K}/2 \rfloor}{\lceil \mathsf{K}/2 \rceil} \frac{\lfloor 2\mathsf{N}/\mathsf{K} \rfloor}{2\mathsf{N}/\mathsf{K}} \times \text{RHS eq(11)}, \quad \forall h \in [0 : \lfloor 2\mathsf{N}/\mathsf{K} - 3 \rfloor], \ \text{only active for } \mathsf{N}/\mathsf{K} \geq 3/2, \tag{14}$$

$$\mathsf{R}^{\star}_{\mathrm{u,c}} \geq \frac{\lfloor \mathsf{K}/2 \rfloor}{\lceil \mathsf{K}/2 \rceil} \times \text{RHS eq(12)}, \tag{15}$$

$$\mathsf{R}^{\star}_{\mathrm{u,c}} \geq \frac{\lfloor \mathsf{K}/2 \rfloor}{\lceil \mathsf{K}/2 \rceil} \times \text{RHS eq(13)}. \tag{16}$$

□

By comparing Scheme A in Theorem 1 and the combination of the new converse bound in Theorem 6 and the converse bound for shared-link caching without privacy in [11], we can characterize the order optimality of Scheme A under the constraint of uncoded cache placement and user collusion in all parameter regimes (the proof can be found in Appendix H).

Theorem 7 (Order optimality for the K-user system)**.** *For the* (K,N) *D2D private caching system where* $\mathsf{N}\geq\mathsf{K}$*, Scheme A in Theorem 1 is order optimal to within a factor of* 18 *(numerical simulations suggest* $27/2$*) under the constraint of uncoded cache placement and user collusion.*

□

Note that when $\mathsf{N}<\mathsf{K}$, Theorem 2 shows that Scheme A is generally order optimal to within a factor of 12. Hence, from Theorems 2 and 7, we can directly have the following conclusion.

Corollary 3. *For the* (K,N) *D2D private caching system, Scheme A in Theorem 1 is order optimal to within a factor of* 18 *under the constraint of uncoded cache placement and user collusion.* □

Remark 5 (Coded vs Uncoded Cache Placement)**.** *For the non-private shared-link coded caching problem in [4], by comparing the optimal coded caching scheme with uncoded cache placement in [5] and the general converse bound in [9], it was proved that the gain of coded cache placement is at most* 2*. Similarly, for the non-private D2D coded caching problem in [20], by comparing the coded caching scheme with uncoded cache placement in [8] and the general converse bound in [9], it was proved that the gain of coded cache placement is at most* 4*. However, for the considered D2D private coded caching problem, by comparing the proposed converse bounds under uncoded cache placement and Scheme C (which is with coded cache placement), it is interesting to find that the gain of coded cache placement is not always within a constant gap. More precisely, let us focus on the two-user system and consider* $\mathsf{M}=\frac{\mathsf{N}+1}{2}$*. By letting* $y=\frac{1}{2}$ *and* $h=0$ *in* (11)*, we have* $\mathsf{R}_{\mathrm{u}}^{\star}\geq\frac{\mathsf{N}-1}{2}$*. By letting* $t=2$ *in* (9)*, Scheme C achieves the memory-load pair* $(\mathsf{M},\mathsf{R}_{\mathrm{C}})=\left(\frac{\mathsf{N}+1}{2},1\right)$*. Hence, we have* $\frac{\mathsf{R}_{\mathrm{u}}^{\star}}{\mathsf{R}_{\mathrm{C}}}\geq\frac{\mathsf{N}-1}{2}$*, which can be unbounded (in the sense that it can be made larger than any constant by choosing a sufficiently large* N*).* □

D. Numerical evaluations

We conclude the overview of our main results with some numerical evaluations. For the achievable schemes, we plot Scheme A in Theorem 1, Scheme B in Theorem 3 (for the two-user system), and Scheme C in Remark 4 (with coded cache placement). We also plot the converse bound under uncoded cache placement in Theorem 4 for $\mathsf{K}=2$ and the converse bound under uncoded cache placement and user collusion in Theorem 6 for $\mathsf{K}\geq 3$. For sake of comparison, we also plot the converse bound in [9] and the converse bound under the constraint of uncoded cache placement in [11] for shared-link caching without privacy.

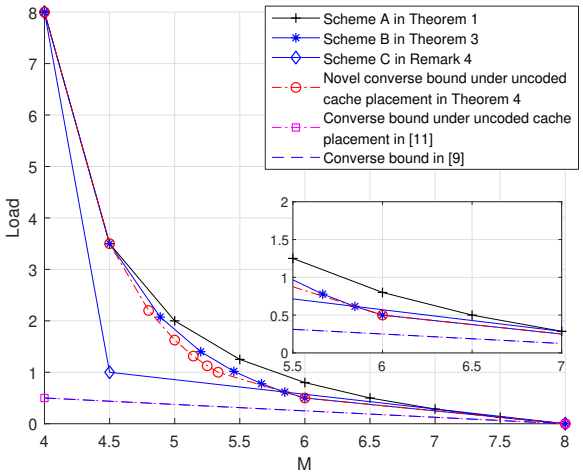

Fig. 1: The memory-load tradeoff for the D2D private caching system, where $\mathsf{K}=2$ and $\mathsf{N}=8$.

In Fig. 1, we consider the case where $\mathsf{K}=2$ and $\mathsf{N}=8$. Here the converse bounds in [11] and [9] are the same. It can be seen in Fig. 1 that, Scheme B and the proposed converse bound meet for all memories except $4.5\leq\mathsf{M}\leq 6$. When $4<\mathsf{M}\leq 5.8$, Scheme C, with coded cache placement, achieves a lower load than the converse bound under uncoded cache placement in Theorem 4.

In Fig. 2, we consider the case where $\mathsf{K}=10$ and $\mathsf{N}=40$. It can be seen in Fig. 2 that compared to the converse bound in [11], the proposed converse bound is tighter when M is small. This is mainly because in the proposed converse bound we treat $\mathsf{K}/2=5$ users as a powerful super-user, which loosens the converse bound when M grows. However, for the low memory size regime, this strategy performs well and gives the order optimality result of Scheme A, while the gap between the converse bound in [11] and Scheme A is not a constant. Hence, combining the proposed converse bound and the converse bound in [11], we can obtain the order optimality results of Scheme A for any memory size.

In Fig. 3, we consider the case where $\mathsf{K}=40$ and $\mathsf{N}=10$. It can be seen that the multiplicative gap between Scheme A and the converse bounds for non-private shared-link coded caching problem is to within a constant. In addition, Scheme A outperforms Scheme C for any $\mathsf{M}\in[0,\mathsf{N}]$.

IV. Achievable Schemes

In this section we provide the details of the achievable schemes together with illustrative examples.

A. Example of Scheme A

Before introducing Scheme A in full generality, we present an example to illustrate the main idea for the D2D private system with $\mathsf{K}=2$ users, $\mathsf{N}=3$ files, and $t=2$ (corresponding to cache size $\mathsf{M}=\frac{5}{2}$).

At a high level, we aim to create a "virtual users"-system with a total $\mathsf{KN}=6$ effective (i.e., real or virtual) users. We then effectively divide the "virtual users"-system into $\mathsf{K}=2$ independent shared-link models, in each of which a real user

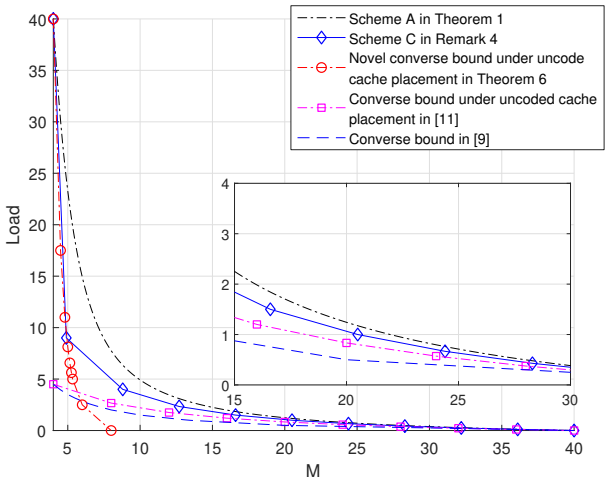

Fig. 2: The memory-load tradeoff for the D2D private caching system, where $\mathsf{K} = 10$ and $\mathsf{N} = 40$.

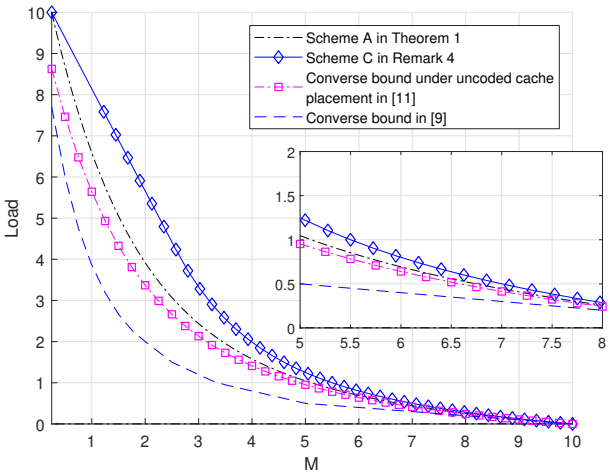

Fig. 3: The memory-load tradeoff for the D2D private caching system, where $\mathsf{K} = 40$ and $\mathsf{N} = 10$.

broadcasts coded multicast packets to $(\mathsf{K}-1)\mathsf{N} = 3$ effective users (including $\mathsf{K}-1 = 1$ real users and $(\mathsf{K}-1)(\mathsf{N}-1) = 2$ virtual users). The demand vector of the effective users served on each independent shared-link model is such that each file is requested exactly $\mathsf{K}-1 = 1$ times, thereby guaranteeing privacy.

File Partitioning. Each file is partitioned into 6 equal-length subfiles as

$$F_i = \{F^1_{i,\{4,5\}}, F^1_{i,\{4,6\}}, F^1_{i,\{5,6\}}, \ F^2_{i,\{1,2\}}, F^2_{i,\{1,3\}}, F^2_{i,\{2,3\}}\}, \tag{17}$$

where $i \in [3]$. Each subfile contains $\mathsf{B}/6$ bits. The subfiles $(F^1_{i,\{4,5\}}, F^1_{i,\{4,6\}}, F^1_{i,\{5,6\}} : i \in [3])$ are to be delivered in the first independent shared-link model by real user 1 to the effective users indexed by $[\mathsf{N}+1 : 2\mathsf{N}] = [4:6]$. Similarly, the subfiles $(F^2_{i,\{1,2\}}, F^2_{i,\{1,3\}}, F^2_{i,\{2,3\}} : i \in [3])$ are to be delivered in the second independent shared-link model by real user 2 to the effective users indexed by $[\mathsf{N}] = [3]$.

Placement Phase. Real user 1 stores all the subfiles with superscript 1 (which it is charged to deliver in the delivery phase), and similarly, real user 2 must store all subfiles with superscript 2. In addition, each real user also stores other subfiles as follows. Real user $k \in [2]$ selects P_k uniformly i.i.d. over $[3]$. The realization of P_1 is unknown to real user 2, and similarly P_2 is unknown to real user 1. Real user $k \in [2]$ impersonates effective user $\theta_k = 3(k-1) + P_k$. Thus, the actual cache content of each real user $k \in [2]$ is

$$Z_k = \{F^k_{i,\mathcal{V}} : i \in [3], \forall \mathcal{V}\} \bigcup_{j \in [\mathsf{K}]\setminus\{k\}} \{F^j_{i,\mathcal{V}} : i \in [3], \theta_k \in \mathcal{V}\}. \tag{18}$$

For example, if we assume $P_1 = 1$ (real user 1 impersonates effective user 1) and $P_2 = 1$ (real user 2 impersonates effective user 4), then real users' cached contents are

$$Z_1 = (F^1_{i,\{4,5\}}, F^1_{i,\{4,6\}}, F^1_{i,\{5,6\}}, \ F^2_{i,\{1,2\}}, F^2_{i,\{1,3\}} : i \in [3]), \tag{19}$$

$$Z_2 = (F^1_{i,\{4,5\}}, F^1_{i,\{4,6\}}, \ F^2_{i,\{1,2\}}, F^2_{i,\{1,3\}}, F^2_{i,\{2,3\}} : i \in [3]), \tag{20}$$

each of $\mathsf{M} = 3\frac{5}{6}$ files.

Thus in the first shared-link model served by real user 1 with the library $(F^1_{i,\{4,5\}}, F^1_{i,\{4,6\}}, F^1_{i,\{5,6\}} : i \in [3])$, each effective user $k \in [4:6]$ caches $(F^1_{i,\mathcal{V}} : \mathcal{V} \in \{\{4,5\},\{4,6\},\{5,6\}\}, k \in \mathcal{V})$. In the second shared-link model served by real user 2 with the library $(F^2_{i,\{1,2\}}, F^2_{i,\{1,3\}}, F^2_{i,\{2,3\}} : i \in [3])$, each effective user $k \in [3]$ caches $(F^2_{i,\mathcal{V}} : \mathcal{V} \in \{\{1,2\},\{1,3\},\{2,3\}\}, k \in \mathcal{V})$.

Delivery Phase. In order to guarantee privacy, we want that each file is demanded the same number of times by the effective users served in each independent shared-link model. Therefore, we let real user $k \in [\mathsf{K}]$, who wants to retrieve the file indexed by d_k, choose uniformly i.i.d. at random one permutation among all permutations of $[\mathsf{N}]$ with P_k-th entry equal to d_k.

Assume that the demand vector is $(d_1, d_2) = (1, 1)$. Denote the demand of effective user k by q_k. Real user 1, who impersonates effective user 1 with demand $q_1 = 1$, randomly chooses (q_2, q_3) to be either $(2,3)$ or $(3,2)$, with equal probability. Real user 1 sends $\ell_1 = (q_1, q_2, q_3)$ as a query to real user 2. Similarly, real user 2, who impersonates effective user 4 with demand $q_4 = 1$, randomly chooses (q_5, q_6) to be either $(2,3)$ or $(3,2)$, with equal probability. Real user 2 sends $\ell_2 = (q_4, q_5, q_6)$ as a query to real user 1. It can be seen that in each independent shared-link model each file is demanded exactly once.

Real user 1 then sends

$$X_1 = F^1_{q_4,\{5,6\}} \oplus F^1_{q_5,\{4,6\}} \oplus F^1_{q_6,\{4,5\}}; \tag{21}$$

thus real user 2, who has cached $F^1_{q_5,\{4,6\}}, F^1_{q_6,\{4,5\}}$, can recover $F^1_{q_4,\{5,6\}}$. Similarity, real user 2 then sends

$$X_2 = F^2_{q_1,\{2,3\}} \oplus F^2_{q_2,\{1,3\}} \oplus F^2_{q_3,\{1,2\}}; \tag{22}$$

thus real user 1, who has cached $F^2_{q_2,\{1,3\}}$ and $F^2_{q_3,\{1,2\}}$, can recover $F^2_{q_1,\{2,3\}}$.

Performance. In the delivery phase, the load is $2\frac{1}{6}$, which coincides with (8). Privacy is guaranteed as, from the viewpoint of real user 1, who does not know the realization of

P_2, all the effective users in $[4:6]$ are equivalent; similarly for real user 2. The information theoretic proof on the privacy will be provided later for the general case. Note that $|P_k|, |\ell_k|$ where $k \in [2]$ do not scale with B, satisfying our assumption in Section II. In conclusion, the proposed scheme is decodable and secure.

B. Proof of Theorem 1: Description of Scheme A

We are now ready to generalize the example in Section IV-A.

Recall that $\mathsf{U} = (\mathsf{K}-1)\mathsf{N}$ denotes the number of virtual users. Let $t \in [\mathsf{U}+1]$. Similar to the "virtual users" scheme for the shared-link model in [7], we aim to contact a D2D system with $\mathsf{K}(\mathsf{N}-1)$ virtual users (in addition the K real users) and divide it into K independent shared-link models, each of which serves U effective users, where $(\mathsf{K}-1)(\mathsf{N}-1)$ are virtual users.

File Partitioning. Each file is partitioned into $\mathsf{K}\binom{\mathsf{U}}{t-1}$ equal-length subfiles as

$$\begin{aligned} F_i = &\{F^k_{i,\mathcal{V}} : k \in [\mathsf{K}], \mathcal{V} \subseteq [\mathsf{KN}] \setminus [(k-1)\mathsf{N}+1 : k\mathsf{N}], \\ &|\mathcal{V}| = t-1\}, \ \forall i \in [\mathsf{N}], \end{aligned} \tag{23}$$

where each subfile contains $\frac{\mathsf{B}}{\mathsf{K}\binom{\mathsf{U}}{t-1}}$ bits. Note that, for each $k \in [\mathsf{K}]$, in (23) we have eliminated the index interval $[(k-1)\mathsf{N}+1 : k\mathsf{N}]$, which is associated with real user k, from the set of all effective users $[\mathsf{KN}]$.

Placement Phase. Each real user $k \in [\mathsf{K}]$ selects $P_k \in [\mathsf{N}]$ uniformly at random and independently across users. We let real user $k \in [\mathsf{K}]$ impersonate effective user $\theta_k := (k-1)\mathsf{N} + P_k$ among the KN effective users. The realization of $P_k, k \in [\mathsf{K}]$, is unknown to all the other real users, that is, the other real users do not know the realization of $\theta_k \in [(k-1)\mathsf{N}+1 : k\mathsf{N}]$.

Each real user $k \in [\mathsf{K}]$ caches all sub-files $F^j_{i,\mathcal{V}}$ for which either $k = j$ or $\theta_k \in \mathcal{V}$, for all files $i \in [\mathsf{N}]$, requiring

$$\mathsf{M} = \mathsf{N}\frac{\binom{\mathsf{U}}{t-1} + (\mathsf{K}-1)\binom{\mathsf{U}-1}{t-2}}{\mathsf{K}\binom{\mathsf{U}}{t-1}} = \mathsf{N}\frac{1+(t-1)/\mathsf{N}}{\mathsf{K}}. \tag{24}$$

Delivery Phase. In the first step, each real user $k \in [\mathsf{K}]$ who demands $d_k \in [\mathsf{N}]$, uniformly and independently selects a vector $\ell_k = (q_{(k-1)\mathsf{N}+1}, \ldots, q_{k\mathsf{N}})$ among all permutations of $[\mathsf{N}]$ whose P_k-th element equals d_k. Then real user $k \in [\mathsf{K}]$ broadcasts ℓ_k to all the other real users. Thus, from the viewpoint of each of the other real users, the union of the demands of the effective users in $[(k-1)\mathsf{N}+1 : k\mathsf{N}]$ is always $[\mathsf{N}]$, which is key to guarantee privacy.

In the second step of the delivery phase, each real user $k \in [\mathsf{K}]$ performs a YMA delivery on the k-th shared-link model with sub-files

$$(F^k_{i,\mathcal{V}} : i \in [\mathsf{N}], \mathcal{V} \subseteq [\mathsf{KN}] \setminus [(k-1)\mathsf{N}+1 : k\mathsf{N}], |\mathcal{V}| = t-1), \tag{25}$$

for effective users $[\mathsf{KN}] \setminus [(k-1)\mathsf{N}+1 : k\mathsf{N}]$. More precisely, for each $i \in [N]$, the effective user with the smallest index in $[\mathsf{KN}] \setminus [(k-1)\mathsf{N}+1 : k\mathsf{N}]$ which requires F_i is chosen a leader for F_i. The leader set for the k-th shared-link model is denoted by $\mathcal{L}_k$. For each $\mathcal{S} \subseteq [\mathsf{KN}] \setminus [(k-1)\mathsf{N}+1 : k\mathsf{N}]$ where $|\mathcal{S}| = t$, we let

$$W^k_{\mathcal{S}} = \underset{j \in \mathcal{S}}{\oplus} F^k_{q_j, \mathcal{S}\setminus\{q_j\}}. \tag{26}$$

Then real user k broadcasts

$$\begin{aligned} X_k = &\big(W^k_{\mathcal{S}} : \mathcal{S} \subseteq [\mathsf{KN}] \setminus [(k-1)\mathsf{N}+1 : k\mathsf{N}], \\ &|\mathcal{S}| = t, \mathcal{S} \cap \mathcal{L}_k \neq \emptyset\big). \end{aligned} \tag{27}$$

Decodability. We focus on real user $k \in [\mathsf{K}]$. From X_j where $j \in [\mathsf{K}] \setminus \{k\}$, it was shown in [5, Lemma 1], real user k can reconstruct each multicast message $W^j_{\mathcal{S}}$ where $\mathcal{S} \subseteq ([\mathsf{KN}] \setminus [(j-1)\mathsf{N}+1 : j\mathsf{N}])$ and $|\mathcal{S}| = t$. Then real user k can recover each $F^j_{d_k,\mathcal{V}}$ where $\mathcal{V} \subseteq ([\mathsf{KN}] \setminus [(j-1)\mathsf{N}+1 : j\mathsf{N}])$, $|\mathcal{V}| = t-1$, and $\theta_k \notin \mathcal{V}$ from $W^j_{\mathcal{V}\cup\{\theta_k\}}$, since real user k caches all the subfiles in $W^j_{\mathcal{V}\cup\{\theta_k\}}$ except $F^j_{d_k,\mathcal{V}}$. In conclusion, real user k can recover all the uncached subfiles of F_{d_k} from $(X_j : j \in [\mathsf{K}] \setminus \{k\})$.

Privacy. We will prove that the privacy constraint in (4) holds.[6] By our construction, the cached content of each effective user is fixed. Hence, $(X_1, \ldots, X_{\mathsf{K}})$ only depends on the demands of the effective users. Since P_j, $j \in [\mathsf{K}]$, is chosen uniformly i.i.d over $[\mathsf{N}]$, θ_j is uniformly i.i.d. over $[(j-1)\mathsf{N}+1 : j\mathsf{N}]$. Hence, for any permutation of $[\mathsf{N}]$ denoted by $\mathbf{u}$, any $i \in [\mathsf{N}]$, and any $(j,k) \in [\mathsf{K}]^2$ where $j \neq k$, (assume that the p-th element of $\mathbf{u}$ is i)

$$\begin{aligned} &\Pr\{(q_{(j-1)\mathsf{N}+1}, \ldots, q_{j\mathsf{N}}) = \mathbf{u} | d_j = i, d_k, Z_k\} \\ &= \Pr\{(q_{(j-1)\mathsf{N}+1}, \ldots, q_{j\mathsf{N}}) = \mathbf{u} | d_j = i\} && \text{(28a)} \\ &= \Pr\{P_j = p | d_j = i\} \Pr\{(q_{(j-1)\mathsf{N}+1}, \ldots, q_{p-1}, \\ &q_{p+1}, \ldots, q_{j\mathsf{N}} | P_j = p, d_j = i\} && \text{(28b)} \\ &= \frac{1}{\mathsf{N}} \Pr\{(q_{(j-1)\mathsf{N}+1}, \ldots, q_{p-1}, q_{p+1}, \ldots, q_{j\mathsf{N}} | P_j = p, d_j = i\} && \text{(28c)} \\ &= \frac{1}{\mathsf{N}} \frac{1}{(\mathsf{N}-1)!}, && \text{(28d)} \end{aligned}$$

where (28a) follows since, given d_j, the demands of the effective users in $[\mathsf{KN}] \setminus [(j-1)\mathsf{N}+1 : j\mathsf{N}]$ are independent of the cached content, queries, and demands of other effective users; (28c) follows since P_j is chosen uniformly over $[\mathsf{N}]$ independent of d_j; and (28d) follows since, given P_j and d_j, the demand vector of the effective users in $[(j-1)\mathsf{N}+1 : j\mathsf{N}]$ is chosen uniformly among all permutations of $[\mathsf{N}]$ where the S_j-th element is d_j. From (28d), it can be seen that $\Pr\{(q_{(j-1)\mathsf{N}+1}, \ldots, q_{j\mathsf{N}}) | d_j, d_k, Z_k\}$ does not depend on (d_j, d_k, Z_k); thus

$$I(q_{(j-1)\mathsf{N}+1}, \ldots, q_{j\mathsf{N}}; d_j | d_k, Z_k) = 0. \tag{29}$$

Hence, from (29) and the fact that given d_j, the demands of the effective users in $[\mathsf{KN}] \setminus [(j-1)\mathsf{N}+1 : j\mathsf{N}]$ are independent of the cached content, queries, and demands of other effective users, we have

$$I(q_1, \ldots, q_{\mathsf{KN}}; \mathbf{d} | d_k, Z_k) = 0. \tag{30}$$

[6]Note that the privacy proof in [7] needs the constraint that the demand of each real user is uniformly i.i.d. over $[\mathsf{N}]$. In the following, we will show that this condition is not necessary.

Recall that $(X_1,\dots,X_{\mathsf{K}})$ only depends on the demands of the effective users; thus we can prove (4). Similarly, we can also prove the privacy constraint against colluding users in (7).

Performance. Each real user $k\in[\mathsf{K}]$ broadcasts $\binom{\mathsf{U}}{t}-\binom{\mathsf{U}-\mathsf{N}}{t}$ multicast messages, each of which contains $\frac{\mathsf{B}}{\mathsf{K}\binom{\mathsf{U}}{t-1}}$ bits. Hence, the achieved load is given by (8). Note that $|P_k|,|\ell_k|$ where $k\in[\mathsf{K}]$ do not scale with B, satisfying our assumption in Section II.

C. Example of Scheme B

We now focus on the case of $\mathsf{K}=2$ user and propose a scheme that does not introduce virtual users and removes the redundancy in the placement phase of Scheme A. Let us return to the example in Section IV-A but with $\mathsf{M}=\frac{9}{4}$ to illustrate the key insights.

Let us first go back to Scheme A. Recall that in Scheme A, each file is split as in (17), the cached contents of the real users are given by (19) and (20). and the transmitted signals are given by (21) and (22). Assume that the demand vector is $(d_1,d_2)=(1,1)$ and the queries are $\ell_1=\ell_2=(1,2,3)$. Thus the transmitted signals are

$$X_1=F^1_{1,\{5,6\}}\oplus F^1_{2,\{4,6\}}\oplus F^1_{3,\{4,5\}},\tag{31}$$
$$X_2=F^2_{1,\{2,3\}}\oplus F^2_{2,\{1,3\}}\oplus F^2_{3,\{1,2\}}.\tag{32}$$

Note that real user 2 caches $(F^1_{2,\{4,5\}},F^1_{2,\{4,6\}})$ but only uses $F^1_{2,\{4,6\}}$ in the decoding procedure. Similarly, real user 2 caches $(F^1_{3,\{4,5\}},F^1_{3,\{4,6\}})$ but only uses $F^1_{3,\{4,5\}}$ in the decoding procedure. In other words, the cached subfiles $F^1_{2,\{4,5\}}$ and $F^1_{3,\{4,6\}}$ are redundant for user 2. Similarly, the cached subfiles $F^2_{2,\{1,2\}}$ and $F^2_{3,\{1,3\}}$ are redundant for user 1. The same is true for any demand vector.

We propose Scheme B to remove this cache redundancy as follows.

File Partitioning. We partition each file into 4 subfiles as

$$F_i=\{F^1_{i,1},F^1_{i,2},F^2_{i,1},F^2_{i,2}\},\ i\in[3],\tag{33}$$

where each subfile contains $\mathsf{B}/4$ bits.

Placement Phase. User 1 selects $P_1=(p_{1,2},p_{2,2},p_{3,2})$ uniformly i.i.d. over $[2]^3$; user 2 selects $P_2=(p_{1,1},p_{2,1},p_{3,1})$ uniformly i.i.d. over $[2]^3$. Then user 1 caches $Z_1=(F^1_{i,1},F^1_{i,2},F^2_{i,p_{i,2}}:i\in[3])$, and user 2 caches $Z_2=(F^1_{i,p_{i,1}},F^2_{i,1},F^2_{i,2}:i\in[3])$. Hence, $\mathsf{M}=\frac{9}{4}$ files.

Delivery Phase. In the delivery phase, we assume that the demand vector is $(d_1,d_2)=(1,1)$. User 1 sends query $\ell_1=(q,p_{2,2},p_{3,2})$ to user 2, where $q\in[2]\setminus\{p_{1,2}\}$. After receiving ℓ_1, user 2 responds by transmitting

$$X_2=F^2_{1,q}\oplus F^2_{2,p_{2,2}}\oplus F^2_{3,p_{3,2}}.\tag{34}$$

User 2 sends query $\ell_2=(q',p_{2,1},p_{3,1})$ to user 1, where $q'\in[2]\setminus\{p_{1,1}\}$. After receiving ℓ_2, user 1 responds by transmitting

$$X_1=F^1_{1,q'}\oplus F^1_{2,p_{2,1}}\oplus F^1_{3,p_{3,1}}.\tag{35}$$

The same can be done for any demand vector.

Performance. Similar to the analysis of Scheme A, Scheme B is decodable and private. In this scheme, $|P_k|,|\ell_k|$ where $k\in[2]$ do not scale with B neither. In this example, Scheme B achieves the memory-load pair $\left(\frac{9}{4},\frac{1}{2}\right)$. Scheme A achieves the memory-load pairs $(2,1)$ for $t=1$, and $\left(\frac{5}{2},\frac{1}{3}\right)$ for $t=2$; hence, by memory-sharing Scheme A achieves the load $\frac{2}{3}$ when $\mathsf{M}=\frac{9}{4}$. Therefore, Scheme B outperforms Scheme A.

D. Proof of Theorem 3: Description of Scheme B

We now ready to provide the general description of Scheme B.

Placement Phase. Each file F_i, where $i\in[\mathsf{N}]$, is partitioned in two equal-length parts, denoted as $F_i=\{F^1_i,F^2_i\}$ where $|F^1_i|=|F^2_i|=\mathsf{B}/2$. For each $k\in[2]$, we further partition F^k_i into $\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}$ equal-length subfiles, denoted by $F^k_{i,1},\dots,F^k_{i,\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}}$, where each subfile has $\frac{\mathsf{B}}{2\left(\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}\right)}$ bits. We randomly generate a permutation of $\left[\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}\right]$, denoted by $\mathbf{p}_{i,k}=\left(p_{i,k}[1],\dots,p_{i,k}\left[\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}\right]\right)$, independently and uniformly over the set of all possible permutations.

We let $P_1=(\mathbf{p}_{i,2}:i\in[\mathsf{N}])$ and $P_2=(\mathbf{p}_{i,1}:i\in[\mathsf{N}])$. Then, we let user k cache all subfiles of F^k_i. In addition, we let the other user (i.e., the user in $[2]\setminus\{k\}$) cache $F^k_{i,p_{i,k}[1]},\dots,F^k_{i,p_{i,k}[\binom{\mathsf{N}-2}{t'-1}]}$.

Considering all the files, each user in total caches $\left(\binom{\mathsf{N}-1}{t'}+2\binom{\mathsf{N}-2}{t'-1}\right)\mathsf{N}$ subfiles, requiring memory

$$\mathsf{M}=\frac{\left(\binom{\mathsf{N}-1}{t'}+2\binom{\mathsf{N}-2}{t'-1}\right)\mathsf{N}}{2\left(\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}\right)}=\frac{\mathsf{N}}{2}+\frac{\mathsf{N}t'}{2(\mathsf{N}+t'-1)}.\tag{36}$$

Delivery Phase. We first focus on the transmission by user 1, in charge of delivery the subfiles with superscirpt 1. For each subset $\mathcal{S}\subseteq[\mathsf{N}]$ where $|\mathcal{S}|=t'+1$, we generate an XOR message containing exactly one subfile of each file in $\mathcal{S}$. More precisely, for each subset $\mathcal{S}\subseteq[\mathsf{N}]$ where $|\mathcal{S}|=t'+1$,

- If $d_2\in\mathcal{S}$, we pick a non-picked subfile among $F^1_{d_2,p_{d_2,1}[\binom{\mathsf{N}-2}{t'-1}+1]},\dots,F^1_{d_2,p_{d_2,1}[\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}]}$. In addition, for each $i\in\mathcal{S}\setminus\{d_2\}$, we pick a non-picked subfile among $F^1_{i,p_{i,1}[1]},\dots,F^1_{i,p_{i,1}[\binom{\mathsf{N}-2}{t'-1}]}$.
- If $d_2\notin\mathcal{S}$, for each $i\in\mathcal{S}$, we pick a non-picked subfile among $F^1_{i,p_{i,1}[\binom{\mathsf{N}-2}{t'-1}+1]},\dots,F^1_{i,p_{i,1}[\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}]}$.

We let $W^1_{\mathcal{S}}$ be the XOR of the picked $t'+1$ subfiles, where $|W^1_{\mathcal{S}}|=\frac{\mathsf{B}}{2\left(\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}\right)}$.

We proceed similarly for user 2. We let $W^2_{\mathcal{S}}$ be the binary sum of the picked $t'+1$ subfiles, where $|W^2_{\mathcal{S}}|=\frac{\mathsf{B}}{2\left(\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}\right)}$.

Finally, user 1 asks user 2 to transmit $X_1=(W^1_{\mathcal{S}}:\mathcal{S}\subseteq[\mathsf{N}],|\mathcal{S}|=t'+1)$, and user 2 asks user 1 to transmit $X_2=(W^2_{\mathcal{S}}:\mathcal{S}\subseteq[\mathsf{N}],|\mathcal{S}|=t'+1)$.[7]

Decodability. We focus on user 1. In each message $W^2_{\mathcal{S}}$ where $\mathcal{S}\subseteq[\mathsf{N}]$, $|\mathcal{S}|=t'+1$, and $d_1\in\mathcal{S}$, user 1 caches all subfiles except one subfile from F_{d_1}, so user 1 can recover this subfile. Hence, user 1 in total recovers $\binom{\mathsf{N}-1}{t'}$ uncached

[7] In other words, the query ℓ_k, $k\in[2]$, represents the indices of the subfiles in $W^k_{\mathcal{S}}$, where $\mathcal{S}\subseteq[\mathsf{N}]$ and $|\mathcal{S}|=t'+1$.

subfiles of F_{d_1}, and thus can recover F_{d_1}. Similarly, user 2 can also recover F_{d_2}.

Privacy. Let us focus on user 1. Since user 1 does not know the random permutations generated in the placement phase, from its viewpoint, all subfiles in F_i^1 where $i \in [\mathsf{N}]$ are equivalent.[8] X_1 contains $\binom{\mathsf{N}}{t'}$ messages, each of which corresponds to a different $(t'+1)$-subset of $[\mathsf{N}]$ and contains exactly one subfile of each file in the subset. Hence, the compositions of X_1 for different demands of user 2 are equivalent from the viewpoint of user 1. In addition, X_2 is generated independent of d_2, and thus X_2 cannot reveal any information of d_2. As a result, the demand of user 2 is private against user 1. Similarly, the demand of user 1 is private against user 2.

Performance. Each user broadcasts $\binom{\mathsf{N}}{t'+1}$ messages, each of which contains $\frac{\mathsf{B}}{2\left(\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}\right)}$ bits. Hence, the achieved load is

$$\mathsf{R} = \frac{2\binom{\mathsf{N}}{t'+1}}{2\left(\binom{\mathsf{N}-1}{t'}+\binom{\mathsf{N}-2}{t'-1}\right)} = \frac{\mathsf{N}(\mathsf{N}-1)}{(t'+1)(\mathsf{N}+t'-1)}. \tag{37}$$

Note that $|P_k|, |\ell_k|$ where $k \in [2]$ do not scale with B, satisfying our assumption in Section II.

V. New Converse Bounds Under the Constraint of Uncoded Cache Placement and User Collusion

In this section, we provide the proofs of our new converse bounds in Theorems 4 and 6. We first introduce the proposed converse bound for the two-user system and then extend it to the K-user system. We start by introducing an example to illustrate in the simplest possible case the new ideas needed to derive our new converse bound.

A. Example of converse

We consider the D2D private system with $(\mathsf{K}, \mathsf{N}) = (2, 2)$ and $\mathsf{M} = 6/5$, for which the achieved load by both Scheme A and Scheme B is $\mathsf{R} = 7/5$. The converse bound under the constraint of uncoded cache placement and one-shot delivery for D2D caching without privacy in [8] gives $\mathsf{R}_{\mathrm{u}}^{\star}(6/5) \geq 4/5$.[9] In the following, we prove that $\mathsf{R}_{\mathrm{u}}^{\star}(6/5) = 7/5$.

Assume we have a working system, that is, a system where all encoding, decoding and privacy constraints listed in Section II are met. In the following, in order not to clutter the derivation with unnecessary "epsilons", we shall neglect the terms P_k, ℓ_k where $k \in [\mathsf{K}]$ that contribute $\epsilon_{\mathsf{B}} = o(\mathsf{B})$ when $\mathsf{B} \to \infty$ to bounds like the one in (40). Finally, without loss of generality (see Remark 7), each user caches a fraction $\mathsf{M}/\mathsf{N} = 3/5$ of each file and each bit in the library is cached by at least one user.

Assume that the cache configurations of the two users are Z_1^1 and Z_2^1, where $Z_1^1 \cup Z_2^1 = \{F_1, F_2\}$. For the demand vector $(d_1, d_2) = (1, 1)$, any working scheme must produce transmitted signals (X_1, X_2) such that the demand vector $(d_1, d_2) = (1, 1)$ can be satisfied. The following observation is critical: because of the privacy constraint, from the viewpoint of user 1, there must exist a cache configuration of user 2, denoted by Z_2^2, such that $Z_1^1 \cup Z_2^2 = \{F_1, F_2\}$, $H(X_2|Z_2^2) = 0$, and F_2 can be decoded from (X_1, Z_2^2). If such a cache configuration Z_2^2 did not exist, then user 1 would know that the demand of user 2 is F_1 from (Z_1^1, X_1, X_2, d_1), which is impossible in a working private system. Similarly, from the viewpoint of user 2, there must exist a cache configuration of user 1, denoted by Z_1^2, such that $Z_1^2 \cup Z_2^1 = \{F_1, F_2\}$, $H(X_1|Z_1^2) = 0$, and F_2 can be decoded from (X_2, Z_1^2).

From (Z_1^1, Z_2^1), because of Remark 7, for each file $F_i,\ i \in [2]$, we have[10]

$$|F_i \cap Z_1^1| = \frac{\mathsf{BM}}{\mathsf{N}} = \frac{3\mathsf{B}}{5}, \tag{38a}$$

$$|F_i \setminus Z_1^1| = |F_i \setminus Z_2^1| = \mathsf{B} - \frac{3\mathsf{B}}{5} = \frac{2\mathsf{B}}{5}, \tag{38b}$$

$$|F_i \cap Z_1^1 \cap Z_2^1| = \frac{\mathsf{B}}{5}. \tag{38c}$$

Similarly, since $Z_1^1 \cup Z_2^1 = Z_1^1 \cup Z_2^2 = \{F_1, F_2\}$, we also must have

$$|F_i \cap Z_1^1 \cap Z_2^2| = \frac{\mathsf{B}}{5}, \tag{38d}$$

$$F_i \setminus Z_1^1 \subseteq F_i \cap Z_2^1 \cap Z_2^2. \tag{38e}$$

Inspired by the genie-aided converse bound for shared-link caching networks without privacy in [5], [11], we construct a genie-aided super-user with cached content

$$Z' = \big(Z_2^1,\ Z_2^2 \setminus (F_1 \cup Z_2^1)\big), \tag{39}$$

who is able to recover the whole library from (X_1, Z'). Indeed, after file F_1 is reconstructed from (X_1, Z_2^1), the combination of $(F_1 \cup Z_2^1)$ and $Z_2^2 \setminus (F_1 \cup Z_2^1)$ gives Z_2^2; now, file F_2 can be reconstructed from (X_1, Z_2^2). Therefore, we have

$$\begin{aligned}
2\mathsf{B} &= H(F_1, F_2) \leq H\big(X_1, Z'\big) && \text{(40a)}\\
&= H\big(X_1, Z_2^1, Z_2^2 \setminus (F_1 \cup Z_2^1)\big) && \text{(40b)}\\
&= H\big(X_1, Z_2^1\big) + H\big(Z_2^2 \setminus (F_1 \cup Z_2^1)|X_1, Z_2^1, F_1\big) && \text{(40c)}\\
&\leq H(X_1) + H(Z_2^1) + H\big(Z_2^2|Z_2^1, F_1\big) && \text{(40d)}\\
&= H(X_1) + H(Z_2^1) + H(F_2 \cap Z_2^2 \cap Z_1^1|Z_2^1) && \text{(40e)}\\
&= H(X_1) + \underbrace{H(Z_2^1)}_{\leq \mathsf{MB}} + \underbrace{H(F_2 \cap Z_2^2 \cap Z_1^1)}_{\leq \mathsf{B}/5}\\
&\quad - H(\underbrace{F_2 \cap Z_2^2 \cap Z_1^1 \cap Z_2^1}_{:=\mathcal{Q}}), && \text{(40f)}
\end{aligned}$$

where (40e) follows since, from (40d), only the bits in F_2 are left, and $Z_2^2 \setminus Z_2^1 = (Z_2^2 \cap Z_1^1) \setminus Z_2^1$ following the reasoning leading to (38e); the last step in (40f) follows since the bits in a file are independent.

[8] In our paper, the statement that from the viewpoint of a user A and B are equivalent, means that given the known information of this user, A and B are identically distributed.

[9] For $\mathsf{K} = 2$, any D2D caching scheme is one-shot.

[10] Intuitively, with uncoded cache placement, each file is split into disjoint pieces as $F_i = (F_{i,\{1\}}, F_{i,\{2\}}, F_{i,\{1,2\}}), i \in [2]$, and the users cache $Z_1 = \cup_{i=1}^2 (F_{i,\{1\}}, F_{i,\{1,2\}}), Z_2 = \cup_{i=1}^2 (F_{i,\{2\}}, F_{i,\{1,2\}})$; by symmetry, let $x \in [0, 1]$ with $|F_{i,\{1\}}| = |F_{i,\{2\}}| = \mathsf{B}x/2$ and $|F_{i,\{1,2\}}| = \mathsf{B}(1-x)$ such that $x/2 + 1 - x = \mathsf{M}/\mathsf{N} = 3/5 \to x = 2(1 - \mathsf{M}/\mathsf{N}) = 4/5$. In the proof, one can think of different cache configurations as different ways to split the files.

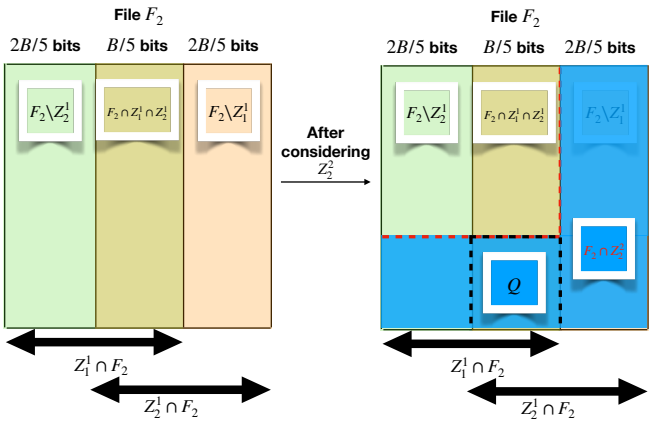

Fig. 4: Illustration of the composition of $\mathcal{Q} := F_2 \cap Z_1^1 \cap Z_2^1 \cap Z_2^2$.

At this point, we need a bound that can be combined with the one in (40) such that it contains on the right hand side the term $H(X_2)$, so that $H(X_1) + H(X_2)$ can be bounded by BR_u, and a term that allows one to get rid of the negative entropy of the random variable

$$\mathcal{Q} := F_2 \cap Z_1^1 \cap Z_2^1 \cap Z_2^2, \tag{41}$$

which is illustrated in Fig. 4.

In the next step, we will introduce another approach to construct a genie-aided super-user, in order to derive an inequality eliminating $\mathcal{Q}$ in (40f). We then focus on the cache configurations Z_1^1 and Z_1^2, and the transmitted packets X_2. Recall that F_1 can be reconstructed from (Z_1^1, X_2), and F_2 can be reconstructed from (Z_1^2, X_2). Furthermore, by recalling the definition of $\mathcal{Q}$ in (41), it can be seen that the bits in $(F_2 \cap Z_1^1) \setminus \mathcal{Q}$ are independent of X_2. Thus F_1 can be reconstructed from $(Z_1^1 \cap F_1, \mathcal{Q}, X_2)$. Hence, we can construct a super-user with cached content

$$Z'' = (Z_1^1 \cap F_1, Z_1^2 \cap F_2, \mathcal{Q}), \tag{42}$$

who can decode both files. Thus

$$\begin{aligned} 2\mathsf{B} = H(F_1, F_2) &\le H(X_2, Z'') && \text{(43a)}\\ &\le H(X_2) + \underbrace{H(Z_1^1 \cap F_1)}_{\le 3\mathsf{B}/5} + \underbrace{H(Z_1^2 \cap F_2)}_{\le 3\mathsf{B}/5} + H(\mathcal{Q}). && \text{(43b)} \end{aligned}$$

Finally, by summing (40f) and (43b), we have that any achievable rate under uncoded cache placement must satisfy

$$\mathsf{R}_\mathrm{u} \ge \frac{H(X_1) + H(X_2)}{\mathsf{B}} \ge \frac{7}{5}. \tag{44}$$

The bound in (44) shows that Scheme A and Scheme B are indeed optimal for the considered memory point.

Remark 6 (A high-level explanation of the converse technique)**.** *The key take-away points in the example in Section V-A are as follows.*

- *By exploiting the privacy constraints, we note that from the viewpoint of each user k (i.e., given cache Z_k and transmitted packets (X_1, X_2)), any demand of the other user is equally possible. Hence, there must exist a cache configuration of the other user that allow for the decoding of any file using the same (X_1, X_2).*
- *We introduce an auxiliary random variable $\mathcal{Q}$ to represents the set of bits $F_2 \cap Z_1^1 \cap Z_2^1 \cap Z_2^2$. We then use two different approaches to construct genie-aided super-users to decode the whole library, in such a way that we can get rid of "tricky" entropy term when the various bounds are summed together:*

  1) In the first approach, we focus on (X_1, Z_2^1, Z_2^2) and construct a genie-aided super-user who can reconstruct the whole library by receiving X_1. The bits in $\mathcal{Q}$ belong to the overlap of Z_2^1 and Z_2^2. Hence, the size of the genie-aided super-user's cache decreases when $|\mathcal{Q}|$ increases. In other words, the needed transmitted load increases when $|\mathcal{Q}|$ increases (see (40f)*).*

  2) In the second approach, we focus on (X_2, Z_1^1, Z_1^2) and construct a genie-aided super-user who can reconstruct the whole library by receiving X_2. Now the bits in $\mathcal{Q}$ are in the cache of the super-user. Hence, the size of the genie-aided super-user's cache increases when $|\mathcal{Q}|$ increases. In other words, the needed transmitted load decreases when $|\mathcal{Q}|$ increases (see (43b)*).*

  Finally, by summing (40f) *and* (43b)*, the effect of $\mathcal{Q}$ is fully cancelled, such that we derive* (44)*.*

□

Remark 7 (On Optimality of Symmetric Placement)**.** *To derive the converse bound under the constraint of uncoded cache placement in the above example, we assumed that every user caches a fraction M/N of each file. This assumption is without loss of generality. Assume that there exists a caching scheme where users cache different fraction of the files. By taking a permutation of $[\mathsf{N}]$ and by using the same strategy to fill the users' caches, we can get another caching scheme. By symmetry, these two caching schemes have the same load. Hence, by considering all possible permutations and taking memory-sharing among all such cache schemes, we have constructed a scheme where every user caches the same fraction of each file, with the same achieved load as the original caching scheme.*

In addition, in the example, we also assumed the total number of cached bits by each user is exactly MB, i.e., the cache of each user is full. Assume that the total number of cached bits by user k is $\mathsf{M}_k\mathsf{B}$. By reasoning as above, we can prove that for any caching scheme, there must exist a caching scheme where $\mathsf{M}_1 = \cdots = \mathsf{M}_\mathsf{K}$ and with the same load as the above scheme. Furthermore, the converse bounds in Theorem 4 and Theorem 6 derived under the assumption that $\mathsf{M}_1 = \cdots = \mathsf{M}_\mathsf{K} = \mathsf{M}$, are non-increasing with the increase of M. Hence, the assumption that the total number of cached bits by each user is exactly MB bits, is also without loss of generality.

Hence, in the proof of our new converse bounds, without loss of generality, we can assume each uses caches a fraction $\frac{\mathsf{M}}{\mathsf{N}}$ of each file. □

B. Proof of Theorem 4: Two-user system

We focus on uncoded cache placement. Without loss of generality, each uses caches a fraction $\frac{\mathsf{M}}{\mathsf{N}}$ of each file (as explained in Remark 7). Let

$$\mathsf{M} = \frac{\mathsf{N}}{2} + y, \tag{45}$$

where $y \in \left[0, \frac{\mathsf{N}}{2}\right]$.

Assume the cache configurations of the two users are (Z_1^1, Z_2^1), where $Z_1^1 \cup Z_2^1 = \{F_1, \ldots, F_{\mathsf{N}}\}$. For the demand vector $(d_1, d_2) = (1, 1)$, any achievable scheme must produce transmitted packets (X_1, X_2), such that the demand vector $(d_1, d_2) = (1, 1)$ can be satisfied. By the privacy constraint in (4), from the viewpoint of user 1 with cache configuration Z_1^1, there must exist some cache configuration Z_2^j such that $Z_1^1 \cup Z_2^j = \{F_1, \ldots, F_{\mathsf{N}}\}$, $H(X_2|Z_2^j) = 0$, and $H(F_j|X_1, Z_2^j) = 0$, for any $j \in [\mathsf{N}]$; otherwise, user 1 will know that the demand of user 2 is not F_j. Similarly, we have the following lemmas.

Lemma 1. *For any $i \in [\mathsf{N}]$ and $j \in [\mathsf{N}]$, there must exist some cache configurations Z_1^i and Z_2^j, such that*

$$Z_1^i \cup Z_2^1 = Z_1^1 \cup Z_2^j = \{F_1, \ldots, F_{\mathsf{N}}\}; \tag{46a}$$

$$H(X_1|Z_1^i) = H(X_2|Z_2^j) = 0; \tag{46b}$$

$$H(F_i|X_2, Z_1^i) = H(F_j|X_1, Z_2^j) = 0. \tag{46c}$$

Lemma 2. *From Z_1^i and Z_2^j where $i, j \in [\mathsf{N}]$ as in Lemma 1, it must hold*

- *consider Z_1^i where $i \in [\mathsf{N}]$. For any $j' \in [\mathsf{N}]$, there must exist a cache configuration denoted by $Z_2^{(i,j')}$ such that $Z_1^i \cup Z_2^{(i,j')} = \{F_1, \ldots, F_{\mathsf{N}}\}$, $H(X_2|Z_2^{(i,j')}) = 0$, and $H(F_{j'}|X_1, Z_2^{(i,j')}) = 0$; and*
- *consider Z_2^j where $j \in [\mathsf{N}]$. For any $i' \in [\mathsf{N}]$, there must exist a cache configuration denoted by $Z_1^{(i',j)}$ such that $Z_1^{(i',j)} \cup Z_2^j = \{F_1, \ldots, F_{\mathsf{N}}\}$, $H(X_1|Z_1^{(i',j)}) = 0$, and $H(F_{i'}|X_2, Z_1^{(i',j)}) = 0$.*

In addition, by definition of Lemma 1,

- *when $i = 1$, we have $Z_2^{(1,j')} = Z_2^{j'}$ for each $j' \in [\mathsf{N}]$; when $j = 1$, we have $Z_1^{(i',1)} = Z_1^{i'}$ for each $i' \in [\mathsf{N}]$; and*
- *when $j' = 1$, we have $Z_2^{(i,1)} = Z_2^1$ for each $i \in [\mathsf{N}]$; when $i' = 1$, we have $Z_1^{(1,j)} = Z_1^1$ for each $j \in [\mathsf{N}]$.*

We can represent the construction of the cache configurations in Lemmas 1 and 2 by an N-ary tree, as illustrated in Fig. 5.

- Two vertices (assumed to be represented by cache configurations Z_1' and Z_2') are connected by an edge with superscript (i, j), if $Z_1' \cup Z_2' = \{F_1, \ldots, F_{\mathsf{N}}\}$, $H(X_1|Z_1') = H(X_2|Z_2') = 0$, and $H(F_i|X_2, Z_1') = H(F_j|X_1, Z_2') = 0$.
- For each $i \in [\mathsf{N}]$, Z_1^i is connected to exactly N vertices, which are $Z_2^{(i,j')}$ where $j' \in [\mathsf{N}]$.
- For each $j \in [\mathsf{N}]$, Z_2^j is connected to exactly N vertices, which are $Z_1^{(i',j)}$ where $i' \in [\mathsf{N}]$.

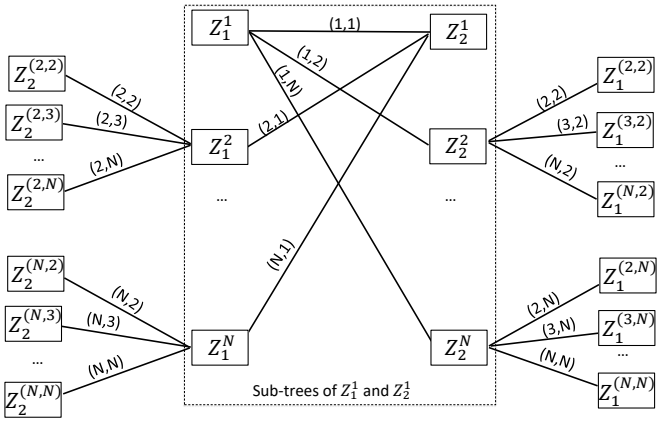

Fig. 5: Construction of cache configurations in Lemmas 1 and 2.

Consider Z_1^i where $i \in [\mathsf{N}]$. Recall that $\mathsf{M} = \mathsf{N}/2 + y$, and that for each $j' \in [\mathsf{N}]$, we have $Z_1^i \cup Z_2^{(i,j')} = \{F_1, \ldots, F_{\mathsf{N}}\}$. For each file F_p where $p \in [\mathsf{N}]$, by defining

$$Z_{1,p}^i := Z_1^i \cap F_p,\ Z_{2,p}^{(i,j')} := Z_2^{(i,j')} \cap F_p,\ \forall j' \in [\mathsf{N}], \tag{47a}$$

we have

$$|F_p \setminus Z_{1,p}^i| = |F_p \setminus Z_{2,p}^{(i,1)}| = \cdots = |F_p \setminus Z_{2,p}^{(i,\mathsf{N})}| = \frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}}; \tag{47b}$$

$$|Z_{1,p}^i \cap Z_{2,p}^{(i,j')}| = \frac{2y\mathsf{B}}{\mathsf{N}},\ \forall j' \in [\mathsf{N}]; \tag{47c}$$

$$(F_p \setminus Z_{1,p}^i) \subseteq Z_{2,p}^{(i,j')},\ \forall j' \in [\mathsf{N}]. \tag{47d}$$

For each file $p \in [\mathsf{N}]$, we define that

$$\mathcal{Q}_{1,p}^i = Z_{1,p}^i \cap Z_{2,p}^{(i,1)} \cap \cdots \cap Z_{2,p}^{(i,\mathsf{N})}, \tag{48}$$

and that $q_{1,p}^i = |\mathcal{Q}_{1,p}^i|$.

Similarly, focus on Z_2^j where $j \in [\mathsf{N}]$, and we have

$$Z_{2,p}^j := Z_2^j \cap F_p,\ Z_{1,p}^{(i',j)} := Z_1^{(i',j)} \cap F_p,\ \forall i' \in [\mathsf{N}]; \tag{49a}$$

$$|F_p \setminus Z_{2,p}^j| = |F_p \setminus Z_{1,p}^{(1,j)}| = \cdots = |F_p \setminus Z_{1,p}^{(\mathsf{N},j)}| = \frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}}; \tag{49b}$$

$$|Z_{2,p}^j \cap Z_{1,p}^{(i',j)}| = \frac{2y\mathsf{B}}{\mathsf{N}},\ \forall i' \in [\mathsf{N}]; \tag{49c}$$

$$(F_p \setminus Z_{2,p}^j) \subseteq Z_{1,p}^{(i',j)},\ \forall i' \in [\mathsf{N}]. \tag{49d}$$

For each file $p \in [\mathsf{N}]$, we define that

$$\mathcal{Q}_{2,p}^j = Z_{2,p}^j \cap Z_{2,p}^{(1,j)} \cap \cdots \cap Z_{2,p}^{(\mathsf{N},j)}, \tag{50}$$

and that $q_{2,p}^j = |\mathcal{Q}_{2,p}^j|$.

After the above definitions, we are ready to prove Theorem 4. As illustrated in the example in Section V-A, we will use two different approaches to construct powerful super-users.

First approach: Consider Z_1^i where $i \in [\mathsf{N}]$. We then focus the connected vertices of Z_1^i in Fig. 5, i.e., $Z_2^{(i,j')}$ where $j' \in [\mathsf{N}]$. By the construction, from $(X_1, Z_2^{(i,j')})$, we can reconstruct $F_{j'}$. The first approach is inspired by the

acyclic index coding converse bound in [5], [11] for shared-link caching without privacy. We pick a permutation of $[\mathsf{N}]$, assumed to be $\mathbf{u}=(u_1,\ldots,u_{\mathsf{N}})$, where $u_1=i$. We can construct a genie-aided super-user with the cache

$$\cup_{p\in[\mathsf{N}]}Z_2^{(i,u_p)}\backslash\big(F_{u_1}\cup\cdots\cup F_{u_{p-1}}\cup Z_2^{(i,u_1)}\cup\cdots\cup Z_2^{(i,u_{p-1})}\big). \tag{51}$$

The genie-aided super-user can successively decode the whole library from its cache and X_1. More precisely, it can first decode F_{u_1} from $(X_1,Z_2^{(i,u_1)})$. From $(X_1,F_{u_1},Z_2^{(i,u_1)},Z_2^{(i,u_2)}\backslash(F_{u_1}\cup Z_2^{(i,u_1)})$, then it can decode F_{u_2}. By this way, the genie-aided super-user can decode the whole library. Hence, we have

$$\begin{aligned}
&H(F_1,\ldots,F_{\mathsf{N}})\\
&\le H(X_1)+H\Big(\cup_{p\in[\mathsf{N}]}Z_2^{(i,u_p)}\setminus\big(F_{u_1}\cup\cdots\cup F_{u_{p-1}}\\
&\cup Z_2^{(i,u_1)}\cup\cdots\cup Z_2^{(i,u_{p-1})}\big)\Big) &\text{(52a)}\\
&\le H(X_1)+H(Z_2^{(i,u_1)})+H(Z_2^{(i,u_2)}|F_{u_1},Z_2^{(i,u_1)})+\cdots\\
&+H(Z_2^{(i,u_{\mathsf{N}})}|F_{u_1},\ldots,F_{u_{\mathsf{N}-1}},Z_2^{(i,u_1)},\ldots,Z_2^{(i,u_{\mathsf{N}-1})}) &\text{(52b)}\\
&=H(X_1)+H(Z_2^{(i,i)})+H(Z_2^{(i,u_2)}|F_i,Z_2^{(i,i)})+\cdots+\\
&H(Z_2^{(i,u_{\mathsf{N}})}|F_i,F_{u_2},\ldots,F_{u_{\mathsf{N}-1}},Z_2^{(i,i)},Z_2^{(i,u_2)},\ldots,Z_2^{(i,u_{\mathsf{N}-1})}) &\text{(52c)}\\
&=H(X_1)+H(Z_2^{(i,i)})\\
&+\Big(H(Z_{2,u_2}^{(i,u_2)}|Z_{2,u_2}^{(i,i)})+\cdots+H(Z_{2,u_{\mathsf{N}}}^{(i,u_{\mathsf{N}})}|Z_{2,u_{\mathsf{N}}}^{(i,i)})\Big)+\cdots\\
&+\Big(H(Z_{2,u_{\mathsf{N}}}^{(i,u_{\mathsf{N}})}|Z_{2,u_{\mathsf{N}}}^{(i,i)},Z_{2,u_{\mathsf{N}}}^{(i,u_2)},\ldots,Z_{2,u_{\mathsf{N}}}^{(i,u_{\mathsf{N}-1})})\Big) &\text{(52d)}\\
&=H(X_1)+H(Z_2^{(i,i)})+H(Z_{2,u_2}^{(i,u_2)}|Z_{2,u_2}^{(i,i)})+H(Z_{2,u_3}^{(i,u_2)},\\
&Z_{2,u_3}^{(i,u_3)}|Z_{2,u_3}^{(i,i)})+\cdots+H(Z_{2,u_{\mathsf{N}}}^{(i,u_2)},\ldots,Z_{2,u_{\mathsf{N}}}^{(i,u_{\mathsf{N}})}|Z_{2,u_{\mathsf{N}}}^{(i,i)}), &\text{(52e)}
\end{aligned}$$

where (52c) follows since $u_1=i$, (52d) follows since all bits in the library are independent, (52e) comes from the chain rule of the entropy.

From (52e), it will be proved in Appendix A-A and Appendix A-B that (recall $y=\mathsf{M}-\mathsf{N}/2$),

$$H(X_1)\ge\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}; \tag{53}$$

$$H(X_1)\ge\mathsf{B}-\frac{4y\mathsf{B}}{\mathsf{N}}+q_{1,u_2}^i. \tag{54}$$

In addition, by considering all permutations of $[\mathsf{N}]$ where the first element is i, we can list all $(\mathsf{N}-1)!$ inequalities as in (52e). By summing all these $(\mathsf{N}-1)!$ inequalities, we can obtain the following inequality, which will be proved in Appendix A-C,

$$\begin{aligned}
H(X_1)\ge{}&\frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{4(\mathsf{N}-1)y\mathsf{B}}{(h+2)\mathsf{N}}+\frac{2}{h+2}\sum_{p\in[\mathsf{N}]\setminus\{i\}}q_{1,p}^i\\
&-\sum_{p\in[\mathsf{N}]\setminus\{i\}}\left\{\frac{\mathsf{N}-2}{(h+1)(h+2)}\left(\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,p}^i\right)\right.\\
&\left.+\frac{h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\},\ \forall h\in[0:\mathsf{N}-3].
\end{aligned}\tag{55}$$

By considering all $i\in[\mathsf{N}]$, we can list all N inequalities as in (55). By summing all these N inequalities, we obtain

$$\begin{aligned}
H(X_1)\ge{}&\frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{4(\mathsf{N}-1)y\mathsf{B}}{(h+2)\mathsf{N}}\\
&+\frac{2}{(h+2)\mathsf{N}}\sum_{i\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{i\}}q_{1,p}^i\\
&-\sum_{i\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{i\}}\left\{\frac{\mathsf{N}-2}{(h+1)(h+2)\mathsf{N}}\left(\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,p}^i\right)\right.\\
&\left.+\frac{h}{(h+2)\mathsf{N}}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\},\ \forall h\in[0:\mathsf{N}-3].
\end{aligned}\tag{56a}$$

We now consider Z_2^j where $j\in[\mathsf{N}]$. By the similar step as above to derive (56a), we obtain

$$\begin{aligned}
H(X_2)\ge{}&\frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{4(\mathsf{N}-1)y\mathsf{B}}{(h+2)\mathsf{N}}+\\
&\frac{2}{(h+2)\mathsf{N}}\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}q_{2,p}^j\\
&-\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}\left\{\frac{\mathsf{N}-2}{(h+1)(h+2)\mathsf{N}}\left(\frac{2y\mathsf{B}}{\mathsf{N}}-q_{2,p}^j\right)\right.\\
&\left.+\frac{h}{(h+2)\mathsf{N}}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\},\ \forall h\in[0:\mathsf{N}-3].
\end{aligned}\tag{57}$$

By summing (56a) and (57), we obtain

$$\mathsf{R}_{\mathrm{u}}^{\star}\mathsf{B}\ge H(X_1)+H(X_2) \tag{58a}$$

$$\begin{aligned}
\ge{}&\mathsf{NB}-2y\mathsf{B}-\frac{8(\mathsf{N}-1)y\mathsf{B}}{(h+2)\mathsf{N}}-\frac{\mathsf{N}-2}{(h+1)(h+2)\mathsf{N}}4y(\mathsf{N}-1)\mathsf{B}\\
&-\frac{h(\mathsf{N}-1)}{(h+2)}\left(\mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}\right)+\left(\frac{2}{(h+2)\mathsf{N}}+\frac{\mathsf{N}-2}{(h+1)(h+2)\mathsf{N}}\right)\\
&\left(\sum_{i\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{i\}}q_{1,p}^i+\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}q_{2,p}^j\right),\forall h\in[0:\mathsf{N}-3].
\end{aligned}\tag{58b}$$

Second approach: We then use the second approach to construct genie-aided super-users. We first consider X_2. By the construction, from (X_2,Z_1^i) where $i\in[\mathsf{N}]$, we can reconstruct F_i.

Now we fix an integer $i\in[\mathsf{N}]$. We pick a permutation of $[\mathsf{N}]$, assumed to be $\mathbf{u}=(u_1,\ldots,u_{\mathsf{N}})$, where $u_1=i$. We can construct a genie-aided super-user with the cache

$$\cup_{p\in[\mathsf{N}]}\left(Z_{1,u_p}^{u_p}\cup\mathcal{Q}_{1,u_p}^{u_1}\cup\cdots\cup\mathcal{Q}_{1,u_p}^{u_{p-1}}\right). \tag{59}$$

Now we prove that the genie-aided super-user can successively decode the whole library from its cache and X_2. Note that from $(Z_1^{u_1},X_2)$, we can reconstruct F_{u_1}. Furthermore, for each file F_{p_1} where $p_1\in[\mathsf{N}]\setminus\{u_1\}$, by recalling the definition of $\mathcal{Q}_{1,p_1}^{u_1}$ in (48), it can be seen that the bits in $Z_{1,p_1}^{u_1}\setminus\mathcal{Q}_{1,p_1}^{u_1}$ are independent of X_2. Hence, it is enough to reconstruct F_{u_1} from $(X_2,Z_{1,u_1}^{u_1},\mathcal{Q}_{1,u_2}^{u_1},\ldots,\mathcal{Q}_{1,u_{\mathsf{N}}}^{u_1})$, and thus the super-user can reconstruct F_{u_1}. After recovering F_{u_1}, the super-user can reconstruct F_{u_2} from $(X_2,F_{u_1},Z_{1,u_2}^{u_2},\mathcal{Q}_{1,u_3}^{u_2},\ldots,\mathcal{Q}_{1,u_{\mathsf{N}}}^{u_2})$. By this way, the genie-aided super-user can decode the whole library. Hence, we have

$$H(X_2)\ge H(F_1,\ldots,F_{\mathsf{N}})$$

$$- H\left(\cup_{p\in[\mathsf{N}]}\left(Z_{1,u_p}^{u_p}\cup\mathcal{Q}_{1,u_p}^{u_1}\cup\cdots\cup\mathcal{Q}_{1,u_p}^{u_{p-1}}\right)\right) \quad (60a)$$

$$\geq \left(H(F_{u_1})-H(Z_{1,u_1}^{u_1})\right)$$
$$+\left(H(F_{u_2})-H(Z_{1,u_2}^{u_2},\mathcal{Q}_{1,u_2}^{u_1})\right)+\cdots$$
$$+\left(H(F_{u_\mathsf{N}})-H(Z_{1,u_\mathsf{N}}^{u_\mathsf{N}},\mathcal{Q}_{1,u_\mathsf{N}}^{u_1},\ldots,\mathcal{Q}_{1,u_\mathsf{N}}^{u_{\mathsf{N}-1}})\right). \quad (60b)$$

From (60b), it will be proved in Appendix B-A and Appendix B-B that,

$$H(X_2)\geq \frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}; \quad (61)$$

$$H(X_2)\geq \mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,u_2}^i. \quad (62)$$

By letting the two permutations to derive (52e) and (60b) be the same, we now sum (53) and (61) to obtain

$$\mathsf{R}_{\mathrm{u}}^{\star}\mathsf{B}\geq H(X_1)+H(X_2)\geq \mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}, \quad (63)$$

which coincides with the proposed converse bound in (13). Similarly, by summing (54) and (62), we obtain

$$\mathsf{R}_{\mathrm{u}}^{\star}\mathsf{B}\geq H(X_1)+H(X_2)\geq 2\mathsf{B}-\frac{6y\mathsf{B}}{\mathsf{N}}, \quad (64)$$

which coincides with the proposed converse bound in (12).

In addition, by considering all permutations of $[\mathsf{N}]$ where the first element is i, we can list all $(\mathsf{N}-1)!$ inequalities as in (60b). By summing all these $(\mathsf{N}-1)!$ inequalities, we can obtain the following inequalities, which will be proved in Appendix B-C,

$$H(X_2)\geq \frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{2}{h+2}\sum_{p\in[\mathsf{N}]\setminus\{i\}}q_{1,p}^i$$
$$-\sum_{p\in[\mathsf{N}]\setminus\{i\}}\left\{\frac{\sum_{n\in[\mathsf{N}]\setminus\{i,p\}}q_{1,p}^n}{(h+1)(h+2)}+\frac{h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\},$$
$$\forall h\in[0:\mathsf{N}-3]. \quad (65)$$

By considering all $i\in[\mathsf{N}]$, we can list all N inequalities as in (65). By summing all these N inequalities, we obtain

$$H(X_2)\geq \frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{2}{(h+2)\mathsf{N}}\sum_{i\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{i\}}q_{1,p}^i$$
$$-\sum_{i\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{i\}}\left\{\frac{\sum_{n\in[\mathsf{N}]\setminus\{i,p\}}q_{1,p}^n}{(h+1)(h+2)\mathsf{N}}\right.$$
$$\left.+\frac{h}{(h+2)\mathsf{N}}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\},\ \forall h\in[0:\mathsf{N}-3]. \quad (66)$$

We now consider X_1. By the similar steps as above to derive (66), we obtain

$$H(X_1)\geq \frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{2}{(h+2)\mathsf{N}}\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}q_{2,p}^j$$
$$-\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}\left\{\frac{\sum_{n\in[\mathsf{N}]\setminus\{j,p\}}q_{2,p}^n}{(h+1)(h+2)\mathsf{N}}\right.$$
$$\left.+\frac{h}{(h+2)\mathsf{N}}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\},\ \forall h\in[0:\mathsf{N}-3]. \quad (67)$$

By summing (66) and (67), we obtain

$$\mathsf{R}_{\mathrm{u}}^{\star}\mathsf{B}\geq H(X_1)+H(X_2) \quad (68a)$$

$$\geq \mathsf{NB}-2y\mathsf{B}$$
$$-\frac{2}{(h+2)\mathsf{N}}\left(\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}q_{2,p}^j+\sum_{i\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{i\}}q_{1,p}^i\right)$$
$$-\frac{1}{(h+1)(h+2)\mathsf{N}}\left(\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}\sum_{n\in[\mathsf{N}]\setminus\{j,p\}}q_{2,p}^n\right.$$
$$\left.+\sum_{i\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{i\}}\sum_{n\in[\mathsf{N}]\setminus\{i,p\}}q_{1,p}^n\right)-\frac{h(\mathsf{N}-1)}{(h+2)}\left(\mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}\right) \quad (68b)$$

$$= \mathsf{NB}-2y\mathsf{B}$$
$$-\frac{2}{(h+2)\mathsf{N}}\left(\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}q_{2,p}^j+\sum_{i\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{i\}}q_{1,p}^i\right)$$
$$-\frac{1}{(h+1)(h+2)\mathsf{N}}\left((\mathsf{N}-2)\sum_{j_1\in[\mathsf{N}]}\sum_{p_1\in[\mathsf{N}]\setminus\{j_1\}}q_{2,p_1}^{j_1}\right.$$
$$\left.+(\mathsf{N}-2)\sum_{i_2\in[\mathsf{N}]}\sum_{p_2\in[\mathsf{N}]\setminus\{i_2\}}q_{1,p_2}^{i_2}\right)$$
$$-\frac{h(\mathsf{N}-1)}{(h+2)}\left(\mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}\right),\ \forall h\in[0:\mathsf{N}-3], \quad (68c)$$

where (68c) follows since[11]

$$\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}\sum_{n\in[\mathsf{N}]\setminus\{j,p\}}q_{2,p}^n=\sum_{j_1\in[\mathsf{N}]}\sum_{p_1\in[\mathsf{N}]\setminus\{j_1\}}q_{2,p_1}^{j_1},$$

and

$$\sum_{i\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{i\}}\sum_{n\in[\mathsf{N}]\setminus\{i,p\}}q_{1,p}^n=(\mathsf{N}-2)\sum_{i_2\in[\mathsf{N}]}\sum_{p_2\in[\mathsf{N}]\setminus\{i_2\}}q_{1,p_2}^{i_2}.$$

Finally, by summing (58b) and (68c), we obtain $\forall h\in[0:\mathsf{N}-3]$,

$$\mathsf{R}_{\mathrm{u}}^{\star}\geq\frac{1}{2}\left\{\mathsf{N}-2y-\frac{8(\mathsf{N}-1)y}{(h+2)\mathsf{N}}-\frac{(\mathsf{N}-2)(\mathsf{N}-1)4y}{(h+1)(h+2)\mathsf{N}}-\right.$$
$$\left.\frac{h(\mathsf{N}-1)}{(h+2)}\left(1-\frac{2y}{\mathsf{N}}\right)\right\}$$
$$+\frac{1}{2}\left\{\mathsf{N}-2y\mathsf{N}-\frac{h(\mathsf{N}-1)}{(h+2)}\left(1-\frac{2y}{\mathsf{N}}\right)\right\} \quad (69a)$$

$$=\mathsf{N}-2y-\frac{4y+(\mathsf{N}-1)h}{h+2}$$
$$+\frac{h^2(n-1)-\mathsf{N}(\mathsf{N}-3)+h(\mathsf{N}+1)}{(h+1)(h+2)}\frac{2y}{\mathsf{N}}, \quad (69b)$$

which coincides with the proposed converse bound in (11).

C. Proof of Theorem 6: K-user System

We extend the proposed converse bound for the two-user system to K-user system and consider the privacy constraint against colluding users in (7). In the following, we consider

[11] In the sum $\sum_{j\in[\mathsf{N}]}\sum_{p\in[\mathsf{N}]\setminus\{j\}}\sum_{n\in[\mathsf{N}]\setminus\{j,p\}}q_{2,p}^n$, let us compute the coefficient of term $q_{2,p_1}^{j_1}$ where $j_1\neq p_1$. $q_{2,p_1}^{j_1}$ appears in the sum when $p=p_1$ and $n=j_1$. Hence, there are $\mathsf{N}-2$ possibilities of j, which are $[\mathsf{N}]\setminus\{p_1,j_1\}$. So the coefficient of $q_{2,p_1}^{j_1}$ in the sum is $\mathsf{N}-2$.

the case where $\mathsf{K}/2$ is an integer and $2\mathsf{N}/\mathsf{K}$ is also an integer. In Appendix C we generalize the proof to any K and N.

Let $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{2y}{\mathsf{K}}$, where $y \in \left[0, \frac{\mathsf{N}}{2}\right]$. We use a genie-aided proof by generating two aggregate users, denoted by k_1 and k_2. We assume that the cache size of each aggregate user is $\mathsf{MB} \times \frac{\mathsf{K}}{2} = \frac{\mathsf{NB}}{2} + y\mathsf{B}$, i.e., the cache size of each aggregate user is the total cache size of $\mathsf{K}/2$ users. In addition, the demanded files of aggregate users k_1 and k_2 are the union sets of the demanded files of users in $[\mathsf{K}/2]$ and of users in $[\mathsf{K}/2+1:\mathsf{K}]$, respectively. The objective is to design a two-user D2D private caching scheme with minimum load $\mathsf{R}_{\mathrm{g}}^{\star}$, such that each aggregate user can decode its demanded files without knowing anything about the demand of the other aggregate user.

Obviously, for any K-user D2D private caching satisfying the encoding (2), decoding (3), and privacy constraints (7), it must be an achievable scheme for the above genie system. In other words, $\mathsf{R}_{\mathrm{u,c}}^{\star} \geq \mathsf{R}_{\mathrm{g}}^{\star}$. Hence, in the following we characterize a converse bound for $\mathsf{R}_{\mathrm{g}}^{\star}$, which is also a converse bound for $\mathsf{R}_{\mathrm{u,c}}^{\star}$.

We partition the N files into $2\mathsf{N}/\mathsf{K}$ equal-size groups, each of which contains $\mathsf{K}/2$ files. Each aggregate user demands one group of files. Hence, it is equivalent to the two-user D2D private caching problem with $2\mathsf{N}/\mathsf{K}$ files, each of which has $\mathsf{KB}/2$ bits, and each of the two users caches $\left(\frac{\mathsf{NB}}{2} + y\mathsf{B}\right)$ bits in its cache and demands one file.

We assume the caches of aggregate users k_1 and k_2 are A_1^1 and A_2^1. The transmitted packets by aggregate users k_1 and k_2 are denoted by X_1' and X_2', such that from (X_2', A_1^1) aggregate user k_1 can decode the files in group 1 and from (X_1', A_2^1) aggregate user k_2 can also decode the files in group 1. We then also construct the cache configurations of aggregate users k_1 and k_2 by a $2\mathsf{N}/\mathsf{K}$-ary tree, as we did in Section V-B.

By the first approach of constructing converse bound described in Section V-B, when we consider A_1^i where $i \in \left[\frac{2\mathsf{N}}{\mathsf{K}}\right]$ (cache of aggregate user k_1 from which and X_2', the files in group i can be reconstructed), with a permutation of $[2\mathsf{N}/\mathsf{K}]$ denoted by $\mathbf{u} = (u_1, \ldots, u_{2\mathsf{N}/\mathsf{K}})$ where $u_1 = i$, we obtain (by the similar derivations of (53) and (54)),

$$H(X_1') \geq \left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}}\right)\frac{\mathsf{K}}{2}; \tag{70}$$

$$H(X_1') \geq \frac{\mathsf{K}}{2}\mathsf{B} - \frac{\mathsf{K}}{2}\frac{4y\mathsf{B}}{\mathsf{N}} + q_{1,u_2}^i, \tag{71}$$

where q_{1,u_2}^i represent the number of bits in $A_1^i \cap A_2^{(i,1)} \cap \cdots \cap A_2^{(i,2\mathsf{N}/\mathsf{K})}$, which are from the files in group u_2.

By considering all permutations of $[2\mathsf{N}/\mathsf{K}]$ whose first element is i, we obtain (by the similar derivation of (55)),

$$\begin{aligned} H(X_1') \geq{}& \frac{\mathsf{NB}}{2} - y\mathsf{B} - \frac{2}{h+2}\left\{\left(\frac{2\mathsf{N}}{\mathsf{K}} - 1\right)\frac{2y\mathsf{B}}{\mathsf{N}}\frac{\mathsf{K}}{2}\right\} \\ &+ \frac{2}{h+2}\sum_{p\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]\setminus\{i\}} q_{1,p}^i \\ &- \sum_{p\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]\setminus\{i\}}\left\{\frac{\frac{2\mathsf{N}}{\mathsf{K}}-2}{(h+1)(h+2)}\left(\frac{2y\mathsf{B}}{\mathsf{N}}\frac{\mathsf{K}}{2} - q_{1,p}^i\right)\right. \\ &\left. + \frac{h}{h+2}\left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}}\right)\frac{\mathsf{K}}{2}\right\},\ \forall h \in \left[0 : \frac{2\mathsf{N}}{\mathsf{K}} - 3\right]. \end{aligned} \tag{72}$$

By considering all $i \in \left[\frac{2\mathsf{N}}{\mathsf{K}}\right]$ to bound $H(X_1')$ and all $j \in \left[\frac{2\mathsf{N}}{\mathsf{K}}\right]$ to bound $H(X_2')$, we sum all inequalities as in (72) to obtain (by the similar derivation of (58b)),

$$\begin{aligned} \mathsf{R}_{\mathrm{g}}^{\star}\mathsf{B} \geq{}& \mathsf{NB} - 2y\mathsf{B} - \frac{4}{h+2}\left\{\left(\frac{2\mathsf{N}}{\mathsf{K}} - 1\right)\frac{2y\mathsf{B}}{\mathsf{N}}\frac{\mathsf{K}}{2}\right\} \\ &- \frac{\frac{2\mathsf{N}}{\mathsf{K}}-2}{(h+1)(h+2)}\frac{4y(\frac{2\mathsf{N}}{\mathsf{K}}-1)\mathsf{B}}{\mathsf{N}}\frac{\mathsf{K}}{2} \\ &- \frac{h\left(\frac{2\mathsf{N}}{\mathsf{K}}-1\right)}{(h+2)}\left(\mathsf{B} - \frac{2y\mathsf{B}}{\mathsf{N}}\right)\frac{\mathsf{K}}{2} \\ &+ \left(\frac{2}{(h+2)\frac{2\mathsf{N}}{\mathsf{K}}} + \frac{\frac{2\mathsf{N}}{\mathsf{K}}-2}{(h+1)(h+2)(2\mathsf{N}/\mathsf{K})}\right) \\ &\left(\sum_{i\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]}\sum_{p\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]\setminus\{i\}} q_{1,p}^i + \sum_{j\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]}\sum_{p\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]\setminus\{j\}} q_{2,p}^j\right), \\ &\forall h \in \left[0 : \frac{2\mathsf{N}}{\mathsf{K}} - 3\right]. \end{aligned} \tag{73}$$

Similarly, by the second approach of constructing converse bound described in Section V-B, when we consider X_2' and the same permutation as the one to derive (70) and (71), we obtain (by the similar derivations of (61) and (62)),

$$H(X_2') \geq \left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}}\right)\frac{\mathsf{K}}{2}; \tag{74}$$

$$H(X_2') \geq \frac{\mathsf{K}}{2}\mathsf{B} - \frac{\mathsf{K}}{2}\frac{2y\mathsf{B}}{\mathsf{N}} - q_{1,u_2}^i. \tag{75}$$

By summing (70) and (74), we prove (16). By summing (71) and (75), we prove (15).

In addition, by the second approach of constructing converse bound described in Section V-B, after considering all permutations to bound $H(X_1')$ and all permutations to bound $H(X_2')$, we obtain (by the similar derivation of (68c)),

$$\begin{aligned} \mathsf{R}_{\mathrm{g}}^{\star}\mathsf{B} \geq{}& \mathsf{NB} - 2y\mathsf{B} - \frac{h\left(\frac{2\mathsf{N}}{\mathsf{K}}-1\right)}{(h+2)}\left(\mathsf{B} - \frac{2y\mathsf{B}}{\mathsf{N}}\right)\frac{\mathsf{K}}{2} \\ &- \left(\frac{2}{(h+2)\frac{2\mathsf{N}}{\mathsf{K}}} + \frac{2\mathsf{N}/\mathsf{K}-2}{(h+1)(h+2)\frac{2\mathsf{N}}{\mathsf{K}}}\right) \\ &\left(\sum_{i\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]}\sum_{p\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]\setminus\{i\}} q_{1,p}^i + \sum_{j\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]}\sum_{p\in\left[\frac{2\mathsf{N}}{\mathsf{K}}\right]\setminus\{j\}} q_{2,p}^j\right), \\ &\forall h \in \left[0 : \frac{2\mathsf{N}}{\mathsf{K}} - 3\right]. \end{aligned} \tag{76}$$

By summing (73) and (76), we prove (14).

VI. Conclusions

We introduced a new D2D private caching model, which aims to preserve the privacy of the users' demands. We proposed new D2D private coded caching schemes, which were proved to be order optimal by matching a new converse bound under the constraint of uncoded cache placement and

user collusion to within a constant gap. Further works include proving new converse bounds for any cache placement, and investigating the decentralized D2D private coded caching problem.

APPENDIX A
PROOFS OF (53), (54), AND (55)

Recall that by considering a permutation of $[\mathsf{N}]$, assumed to be $\mathbf{u}=(u_1,\ldots,u_{\mathsf{N}})$, where $u_1=i$, we can derive (52e),

$$H(F_1,\ldots,F_{\mathsf{N}}) \le H(X_1)+H(Z_2^{(i,i)}) + \sum_{p\in[2:\mathsf{N}]} H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right). \tag{77}$$

For each $p\in[2:\mathsf{N}]$, since $|Z_{2,u_p}^{(i,i)}|=\frac{\mathsf{B}}{2}+\frac{y\mathsf{B}}{\mathsf{N}}$, we have

$$H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right) \le H(F_p|Z_{2,u_p}^{(i,i)}) = \frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}. \tag{78}$$

A. Proof of (53)

Now we bound each term $H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right)$ where $p\in[2:\mathsf{N}]$ in (77) by $\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}$, to obtain

$$\begin{aligned} &H(F_1,\ldots,F_{\mathsf{N}}) \le H(X_1)+H(Z_2^{(i,i)}) \\ &+\sum_{p\in[2:\mathsf{N}]} H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right) && \text{(79a)}\\ &\le H(X_1)+H(Z_2^{(i,i)})+(\mathsf{N}-1)\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right) && \text{(79b)}\\ &= H(X_1)+\frac{\mathsf{NB}}{2}+y\mathsf{B}+(\mathsf{N}-1)\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right). && \text{(79c)} \end{aligned}$$

Hence, we have

$$H(X_1) \ge \frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}, \tag{80}$$

which proves (53).

B. Proof of (54)

We first prove for each $i\in[\mathsf{N}]$ and $n,p\in[\mathsf{N}]\setminus\{i\}$, we have

$$\begin{aligned} &H\left(Z_{2,p}^{(i,n)}|Z_{2,p}^{(i,i)}\right) = H\left(Z_{2,p}^{(i,n)}|Z_{2,p}^{(i,i)},F_p\setminus Z_{1,p}^i\right) && \text{(81a)}\\ &= H\left(Z_{2,p}^{(i,n)}\cap Z_{1,p}^i|Z_{2,p}^{(i,i)},F_p\setminus Z_{1,p}^i\right) && \text{(81b)}\\ &= H\left(Z_{2,p}^{(i,n)}\cap Z_{1,p}^i|Z_{2,p}^{(i,i)}\right) && \text{(81c)}\\ &\le H\left(Z_{2,p}^{(i,n)}\cap Z_{1,p}^i\right)-q_{1,p}^i && \text{(81d)}\\ &= \frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,p}^i, && \text{(81e)} \end{aligned}$$

where (81a) follows since $Z_{2,p}^{(i,i)}\cup Z_{1,p}^i = Z_{2,p}^{(i,n)}\cup Z_{1,p}^i = F_p$ and thus $(F_p\setminus Z_{1,p}^i)\subseteq Z_{2,p}^{(i,i)}$, (81b) and (81c) follow since all bits in the library are independent, (81d) comes from (48), (81e) comes from (47c).

Now we bound each term $H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right)$ where $p\in[3:\mathsf{N}]$ in (77) by $\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}$, to obtain

$$\begin{aligned} &H(F_1,\ldots,F_{\mathsf{N}}) \le H(X_1)+H(Z_2^{(i,i)}) \\ &+\sum_{p\in[2:\mathsf{N}]} H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right) && \text{(82a)}\\ &\le H(X_1)+\frac{\mathsf{NB}}{2}+y\mathsf{B}+H\left(Z_{2,u_2}^{(i,u_2)}|Z_{2,u_2}^{(i,i)}\right) \\ &+(\mathsf{N}-2)\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right) && \text{(82b)}\\ &\le H(X_1)+\frac{\mathsf{NB}}{2}+y\mathsf{B}+\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,u_2}^i \\ &+(\mathsf{N}-2)\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right), && \text{(82c)} \end{aligned}$$

where (82c) comes from (81e).

Hence, we have

$$\begin{aligned} &H(X_1) \ge \frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}+q_{1,u_2}^i-(\mathsf{N}-2)\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right) && \text{(83a)}\\ &= \mathsf{B}-\frac{4y\mathsf{B}}{\mathsf{N}}+q_{1,u_2}^i, && \text{(83b)} \end{aligned}$$

which proves (54).

C. Proof of (55)

Note that when $\mathsf{N}=2$, (55) does not exist. Hence, in the following we consider $\mathsf{N}\ge 3$ to prove (55).

From (77), we have

$$\begin{aligned} &H(F_1,\ldots,F_{\mathsf{N}}) \le H(X_1)+H(Z_2^{(i,i)}) \\ &+\sum_{p\in[2:\mathsf{N}]} H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right) && \text{(84a)}\\ &= H(X_1)+H(Z_2^{(i,i)})+\sum_{p\in[2:\mathsf{N}]}\Big\{H\left(Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right) \\ &+H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_{p-1})}|Z_{2,u_p}^{(i,i)},Z_{2,u_p}^{(i,u_p)}\right)\Big\} && \text{(84b)}\\ &= H(X_1)+\frac{\mathsf{NB}}{2}+y\mathsf{B}+\sum_{p\in[2:\mathsf{N}]} H\left(Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right) \\ &+\sum_{p\in[2:\mathsf{N}]} H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_{p-1})}|Z_{2,u_p}^{(i,i)},Z_{2,u_p}^{(i,u_p)}\right). && \text{(84c)} \end{aligned}$$

By considering all permutations of $[\mathsf{N}]$ where the first element is i and summing all inequalities as (84c), we have

$$\begin{aligned} &H(X_1) \ge \frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{1}{(\mathsf{N}-1)!}\sum_{\mathbf{u}:u_1=i}\sum_{p\in[2:\mathsf{N}]} \\ &H\left(Z_{2,u_p}^{(i,u_p)}|Z_{2,u_p}^{(i,i)}\right)-\frac{1}{(\mathsf{N}-1)!}\sum_{\mathbf{u}:u_1=i}\sum_{p\in[2:\mathsf{N}]} \\ &H\left(Z_{2,u_p}^{(i,u_2)},\ldots,Z_{2,u_p}^{(i,u_{p-1})}|Z_{2,u_p}^{(i,i)},Z_{2,u_p}^{(i,u_p)}\right) && \text{(85a)}\\ &= \frac{\mathsf{NB}}{2}-y\mathsf{B}-\sum_{p\in[\mathsf{N}]\setminus\{i\}} H\left(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)}\right)-\frac{1}{(\mathsf{N}-1)!}\sum_{p\in[\mathsf{N}]\setminus\{i\}} \\ &\sum_{r\in[2:\mathsf{N}]}\sum_{\mathbf{u}:u_1=i,u_r=p} H\left(Z_{2,p}^{(i,u_2)},\ldots,Z_{2,p}^{(i,u_{r-1})}|Z_{2,p}^{(i,i)},Z_{2,p}^{(i,p)}\right), && \text{(85b)} \end{aligned}$$

where (85b) comes from the re-arrangements on the summations.

To bound the last term in (85b), we now focus on one file F_p where $p \in [\mathsf{N}] \setminus \{i\}$ and bound the following term

$$\sum_{r\in[2:\mathsf{N}]} \sum_{\mathbf{u}:u_1=i,u_r=p} H\left(Z_{2,p}^{(i,u_2)},\ldots,Z_{2,p}^{(i,u_{r-1})}|Z_{2,p}^{(i,i)},Z_{2,p}^{(i,p)}\right). \tag{86}$$

Note that the conditional entropies in (86) are conditioned on the same term, which is $Z_{2,p}^{(i,i)} \cup Z_{2,p}^{(i,p)}$. In addition, for any $n \in [\mathsf{N}] \setminus \{i,p\}$, we have

$$Z_{2,p}^{(i,n)} \setminus (Z_{2,p}^{(i,i)} \cup Z_{2,p}^{(i,p)}) \subseteq F_p \setminus (Z_{2,p}^{(i,i)} \cup Z_{2,p}^{(i,p)}).$$

Hence, we divide the bits in $F_p \setminus (Z_{2,p}^{(i,i)} \cup Z_{2,p}^{(i,p)})$ into sub-pieces, and denote (with a slight abuse of notation)

$$F_p \setminus (Z_{2,p}^{(i,i)} \cup Z_{2,p}^{(i,p)}) = \{\mathcal{F}_{p,\mathcal{S}} : \mathcal{S} \subseteq ([\mathsf{N}] \setminus \{i,p\})\}, \tag{87}$$

where

$$\mathcal{F}_{p,\mathcal{S}} = \left(F_p \setminus (Z_{2,p}^{(i,i)} \cup Z_{2,p}^{(i,p)})\right) \cap \left(\cap_{n\in\mathcal{S}} Z_{2,p}^{(i,n)}\right) \setminus \left(\cup_{n_1\notin\mathcal{S}} Z_{2,p}^{(i,n_1)}\right). \tag{88}$$

In other words, $\mathcal{F}_{p,\mathcal{S}}$ represents the bits in $F_p \setminus (Z_{2,p}^{(i,i)} \cup Z_{2,p}^{(i,p)})$ which are exclusively in $Z_{2,p}^{(i,n)}$ where $n \in \mathcal{S}$.

We then define

$$f_t := \sum_{\mathcal{S}\subseteq([\mathsf{N}]\setminus\{i,p\}):|\mathcal{S}|=t} |\mathcal{F}_{p,\mathcal{S}}|, \ \forall t \in [0:\mathsf{N}-2], \tag{89}$$

as the total length of sub-pieces $\mathcal{F}_{p,\mathcal{S}}$ where $|\mathcal{S}| = t$.

In (81e), we proved that for each $n \in [\mathsf{N}] \setminus \{i,p\}$, we have $H(Z_{2,p}^{(i,n)}|Z_{2,p}^{(i,i)}) \le \frac{2y\mathsf{B}}{\mathsf{N}} - q_{1,p}^i$. Hence, we also have $H(Z_{2,p}^{(i,n)}|Z_{2,p}^{(i,i)},Z_{2,p}^{(i,p)}) \le H(Z_{2,p}^{(i,n)}|Z_{2,p}^{(i,i)}) \le \frac{2y\mathsf{B}}{\mathsf{N}} - q_{1,p}^i$. In other words,

$$\sum_{\mathcal{S}\subseteq[\mathsf{N}]\setminus\{i,p\}:n\in\mathcal{S}} |\mathcal{F}_{p,\mathcal{S}}| \le \frac{2y\mathsf{B}}{\mathsf{N}} - q_{1,p}^i. \tag{90}$$

By summing (90) over all $n \in [\mathsf{N}] \setminus \{i,p\}$, we have

$$\sum_{t\in[0:\mathsf{N}-2]} t f_t = \sum_{n\in[\mathsf{N}]\setminus\{i,p\}} \sum_{\mathcal{S}\subseteq[\mathsf{N}]\setminus\{i,p\}:n\in\mathcal{S}} |\mathcal{F}_{p,\mathcal{S}}| \tag{91a}$$

$$\le (\mathsf{N}-2)\left(\frac{2y\mathsf{B}}{\mathsf{N}} - q_{1,p}^i\right). \tag{91b}$$

In addition, since $F_p \setminus (Z_{2,p}^{(i,i)} \cup Z_{2,p}^{(i,p)}) = (F_p \setminus Z_{2,p}^{(i,i)}) \setminus (Z_{2,p}^{(i,p)} \setminus Z_{2,p}^{(i,i)})$, we have

$$|F_p \setminus (Z_{2,p}^{(i,i)} \cup Z_{2,p}^{(i,p)})| = |F_p \setminus Z_{2,p}^{(i,i)}| - |Z_{2,p}^{(i,p)} \setminus Z_{2,p}^{(i,i)}| \tag{92a}$$

$$= \frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}} - H(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)}). \tag{92b}$$

Hence, we have

$$\sum_{t\in[0:\mathsf{N}-2]} f_t = \sum_{\mathcal{S}\subseteq[\mathsf{N}]\setminus\{i,p\}} |\mathcal{F}_{p,\mathcal{S}}| \tag{93a}$$

$$= \frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}} - H(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)}). \tag{93b}$$

From the above definitions, we can re-write (86) as follows,

$$\sum_{r\in[2:\mathsf{N}]} \sum_{\substack{\mathbf{u}:\\ u_1=i,u_r=p}} H\left(Z_{2,p}^{(i,u_2)},\ldots,Z_{2,p}^{(i,u_{r-1})}|Z_{2,p}^{(i,i)},Z_{2,p}^{(i,p)}\right) = \sum_{r\in[2:\mathsf{N}]} \sum_{\substack{\mathbf{u}:\\ u_1=i,u_r=p}} \sum_{\substack{\mathcal{S}\subseteq([\mathsf{N}]\setminus\{i,p\}):\\ \mathcal{S}\cap\{u_2,\ldots,u_{r-1}\}\neq\emptyset}} |\mathcal{F}_{p,\mathcal{S}}|. \tag{94}$$

In (94), for each $r \in [2:\mathsf{N}]$, we can compute

$$\sum_{\substack{\mathbf{u}:\\ u_1=i,u_r=p}} \sum_{\substack{\mathcal{S}\subseteq([\mathsf{N}]\setminus\{i,p\}):\\ \mathcal{S}\cap\{u_2,\ldots,u_{r-1}\}\neq\emptyset}} |\mathcal{F}_{p,\mathcal{S}}| = \sum_{t\in[0:\mathsf{N}-2]} (\mathsf{N}-2)! \frac{\binom{\mathsf{N}-2}{t} - \binom{\mathsf{N}-r-1}{t}}{\binom{\mathsf{N}-2}{t}} f_t. \tag{95}$$

This is because in $\sum_{\substack{\mathcal{S}\subseteq([\mathsf{N}]\setminus\{i,p\}):\\ \mathcal{S}\cap\{u_2,\ldots,u_{r-1}\}\neq\emptyset}} |\mathcal{F}_{p,\mathcal{S}}|$, there are $\binom{\mathsf{N}-2}{t} - \binom{\mathsf{N}-2-(r-1)}{t}$ sub-pieces whose $\mathcal{S}$ has t elements. Considering all permutations $\mathbf{u}$ where $u_1 = i$ and $u_r = p$, by the symmetry, the coefficient of each $|\mathcal{F}_{p,\mathcal{S}}|$ where $\mathcal{S} = t$ should be the same. In addition, there are in total $\binom{\mathsf{N}-2}{t}$ sub-pieces whose $\mathcal{S}$ has t elements. Hence, we obtain (95).

Considering all $r \in [2:\mathsf{N}-2]$, from (95) we have

$$\sum_{r\in[2:\mathsf{N}]} \sum_{\substack{\mathbf{u}:\\ u_1=i,u_r=p}} H\left(Z_{2,p}^{(i,u_2)},\ldots,Z_{2,p}^{(i,u_{r-1})}|Z_{2,p}^{(i,i)},Z_{2,p}^{(i,p)}\right)$$

$$= \sum_{r\in[2:\mathsf{N}]} \sum_{t\in[0:\mathsf{N}-2]} (\mathsf{N}-2)! \frac{\binom{\mathsf{N}-2}{t} - \binom{\mathsf{N}-r-1}{t}}{\binom{\mathsf{N}-2}{t}} f_t \tag{96a}$$

$$= (\mathsf{N}-2)! \sum_{t\in[0:\mathsf{N}-2]} \sum_{r\in[2:\mathsf{N}]} \frac{\binom{\mathsf{N}-2}{t} - \binom{\mathsf{N}-r-1}{t}}{\binom{\mathsf{N}-2}{t}} f_t \tag{96b}$$

$$= (\mathsf{N}-2)! \sum_{t\in[0:\mathsf{N}-2]} \left(\frac{(\mathsf{N}-2)\binom{\mathsf{N}-2}{t} - \binom{\mathsf{N}-2}{t+1}}{\binom{\mathsf{N}-2}{t}}\right) f_t \tag{96c}$$

$$= (\mathsf{N}-1)! \sum_{t\in[0:\mathsf{N}-2]} \frac{t}{t+1} f_t, \tag{96d}$$

where (96c) comes from the Pascal's Triangle, $\binom{\mathsf{N}-3}{t} + \cdots + \binom{t}{t} = \binom{\mathsf{N}-2}{t+1}$.

The next step is to use Fourier-Motzkin elimination on f_t where $t \in [0:\mathsf{N}-2]$ in (96d) (as we did in [11]) with the help of (91b) and (93b). More precisely, we fix one integer $h \in [0:\mathsf{N}-3]$. We multiply (91b) by $\frac{(\mathsf{N}-1)!}{(h+1)(h+2)}$ and multiply (93b) by $\frac{(\mathsf{N}-1)!h}{h+2}$, and sum them to obtain

$$\sum_{t\in[0:\mathsf{N}-2]} \left(t\frac{(\mathsf{N}-1)!}{(h+1)(h+2)} + \frac{(\mathsf{N}-1)!h}{h+2}\right) f_t \le \frac{(\mathsf{N}-1)!(\mathsf{N}-2)}{(h+1)(h+2)}\left(\frac{2y\mathsf{B}}{\mathsf{N}} - q_{1,p}^i\right) + \frac{(\mathsf{N}-1)!h}{h+2}\left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}} - H(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)})\right). \tag{97}$$

From (97), we have

$$\frac{(\mathsf{N}-1)!h}{h+1} f_h + \frac{(\mathsf{N}-1)!(h+1)}{h+2} f_{h+1} \le \frac{(\mathsf{N}-1)!(\mathsf{N}-2)}{(h+1)(h+2)}\left(\frac{2y\mathsf{B}}{\mathsf{N}} - q_p^i\right)$$

$$+\frac{(\mathsf{N}-1)!h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}-H(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)})\right)$$
$$-\sum_{t\in[0:\mathsf{N}-2]:t\notin\{h,h+1\}}\left(t\frac{(\mathsf{N}-1)!}{(h+1)(h+2)}+\frac{(\mathsf{N}-1)!h}{h+2}\right)f_t. \tag{98}$$

We then take (98) into (96d) to obtain,

$$\sum_{r\in[2:\mathsf{N}]}\sum_{\substack{\mathbf{u}:\\ u_1=i,u_r=p}}H\left(Z_{2,p}^{(i,u_2)},\ldots,Z_{2,p}^{(i,u_{r-1})}|Z_{2,p}^{(i,i)},Z_{2,p}^{(i,p)}\right)$$
$$\le\frac{(\mathsf{N}-1)!(\mathsf{N}-2)}{(h+1)(h+2)}\left(\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,p}^{i}\right)$$
$$+\frac{(\mathsf{N}-1)!h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}-H(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)})\right)$$
$$-\sum_{t\in[0:\mathsf{N}-2]}(\mathsf{N}-1)!\frac{(h-t)(h+1-t)}{(h+1)(h+2)(t+1)}f_t \tag{99a}$$
$$\le\frac{(\mathsf{N}-1)!(\mathsf{N}-2)}{(h+1)(h+2)}\left(\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,p}^{i}\right)$$
$$+\frac{(\mathsf{N}-1)!h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}-H(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)})\right). \tag{99b}$$

Finally, we take (99b) into (85b) to obtain, for each $h\in[0:\mathsf{N}-3]$,

$$H(X_1)\ge\frac{\mathsf{NB}}{2}-y\mathsf{B}-\sum_{p\in[\mathsf{N}]\setminus\{i\}}H\left(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)}\right)$$
$$-\frac{1}{(\mathsf{N}-1)!}\sum_{p\in[\mathsf{N}]\setminus\{i\}}\sum_{r\in[2:\mathsf{N}]}\sum_{\substack{\mathbf{u}:\\ u_1=i,u_r=p}}$$
$$H\left(Z_{2,p}^{(i,u_2)},\ldots,Z_{2,p}^{(i,u_{r-1})}|Z_{2,p}^{(i,i)},Z_{2,p}^{(i,p)}\right) \tag{100a}$$
$$\ge\frac{\mathsf{NB}}{2}-y\mathsf{B}-\sum_{p\in[\mathsf{N}]\setminus\{i\}}H\left(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)}\right)$$
$$-\frac{1}{(\mathsf{N}-1)!}\sum_{p\in[\mathsf{N}]\setminus\{i\}}\left\{\frac{(\mathsf{N}-1)!(\mathsf{N}-2)}{(h+1)(h+2)}\left(\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,p}^{i}\right)\right.$$
$$\left.+\frac{(\mathsf{N}-1)!h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}-H(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)})\right)\right\} \tag{100b}$$
$$=\frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{2}{h+2}\sum_{p\in[\mathsf{N}]\setminus\{i\}}H\left(Z_{2,p}^{(i,p)}|Z_{2,p}^{(i,i)}\right)-\sum_{p\in[\mathsf{N}]\setminus\{i\}}$$
$$\left\{\frac{(\mathsf{N}-2)}{(h+1)(h+2)}\left(\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,p}^{i}\right)+\frac{h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\} \tag{100c}$$
$$\ge\frac{\mathsf{NB}}{2}-y\mathsf{B}-\frac{2}{h+2}\sum_{p\in[\mathsf{N}]\setminus\{i\}}\left(\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,p}^{i}\right)-\sum_{p\in[\mathsf{N}]\setminus\{i\}}$$
$$\left\{\frac{(\mathsf{N}-2)}{(h+1)(h+2)}\left(\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,p}^{i}\right)+\frac{h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\}, \tag{100d}$$

where (100d) comes from (81e). Hence, we prove (55).

APPENDIX B
PROOFS OF (61), (62), AND (65)

The proofs of (61) (62) (65) come from a similar strategy used in Appendix A. Hence, in the following, we briefly describe the proofs of (61) (62) (65).

Recall from (60b) that by considering a permutation of $[\mathsf{N}]$, assumed to be $\mathbf{u}=(u_1,\ldots,u_{\mathsf{N}})$, where $u_1=i$, we can derive

$$H(X_2)\ge\big(H(F_i)-H(Z_{1,i}^{i})\big)+$$
$$\sum_{p\in[2:\mathsf{N}]}\Big(H(F_{u_p})-H(Z_{1,u_p}^{u_p},\mathcal{Q}_{1,u_p}^{i},\mathcal{Q}_{1,u_p}^{u_2},\ldots,\mathcal{Q}_{1,u_p}^{u_{p-1}})\Big). \tag{101}$$

For each $p\in[2:\mathsf{N}]$, we have

$$H(F_{u_p})-H(Z_{1,u_p}^{u_p},\mathcal{Q}_{1,u_p}^{i},\mathcal{Q}_{1,u_p}^{u_2},\ldots,\mathcal{Q}_{1,u_p}^{u_{p-1}})\ge 0. \tag{102}$$

A. Proof of (61)

Now we bound each term $H(F_{u_p})-H(Z_{1,u_p}^{u_p},\mathcal{Q}_{1,u_p}^{i},\mathcal{Q}_{1,u_p}^{u_2},\ldots,\mathcal{Q}_{1,u_p}^{u_{p-1}})$ where $p\in[2:\mathsf{N}]$ in (101) by 0, to obtain

$$H(X_2)\ge H(F_i)-H(Z_{1,i}^{i})=\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}, \tag{103}$$

which proves (61).

B. Proof of (62)

Now we bound each term $H(F_{u_p})-H(Z_{1,u_p}^{u_p},\mathcal{Q}_{1,u_p}^{i},\mathcal{Q}_{1,u_p}^{u_2},\ldots,\mathcal{Q}_{1,u_p}^{u_{p-1}})$ where $p\in[3:\mathsf{N}]$ in (101) by 0, to obtain

$$H(X_2)\ge\big(H(F_i)-H(Z_{1,i}^{i})\big)$$
$$+\sum_{p\in[2:\mathsf{N}]}\Big(H(F_{u_p})-H(Z_{1,u_p}^{u_p},\mathcal{Q}_{1,u_p}^{i},\mathcal{Q}_{1,u_p}^{u_2},\ldots,\mathcal{Q}_{1,u_p}^{u_{p-1}})\Big) \tag{104a}$$
$$\ge\big(H(F_i)-H(Z_{1,i}^{i})\big)+\big(H(F_{u_2})-H(Z_{1,u_2}^{u_2},\mathcal{Q}_{1,u_2}^{i})\big) \tag{104b}$$
$$\ge H(F_i)-H(Z_{1,i}^{i}+H(F_{u_2})-H(Z_{1,u_2}^{u_2})-H(\mathcal{Q}_{1,u_2}^{i}) \tag{104c}$$
$$=\mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}-q_{1,u_2}^{i}. \tag{104d}$$

which proves (62).

C. Proof of (65)

Note that when $\mathsf{N}=2$, (65) does not exist. Hence, in the following we consider $\mathsf{N}\ge 3$ to prove (65).

From (101), we have

$$H(X_2)\ge\big(H(F_i)-H(Z_{1,i}^{i})\big)$$
$$+\sum_{p\in[2:\mathsf{N}]}\Big(H(F_{u_p})-H(Z_{1,u_p}^{u_p},\mathcal{Q}_{1,u_p}^{i},\mathcal{Q}_{1,u_p}^{u_2},\ldots,\mathcal{Q}_{1,u_p}^{u_{p-1}})\Big) \tag{105a}$$
$$=\big(H(F_i)-H(Z_{1,i}^{i})\big)+\sum_{p\in[2:\mathsf{N}]}\Big(H(F_{u_p})-H(Z_{1,u_p}^{u_p})$$
$$-H(\mathcal{Q}_{1,u_p}^{i}|Z_{1,u_p}^{u_p})-H(\mathcal{Q}_{1,u_p}^{u_2},\ldots,\mathcal{Q}_{1,u_p}^{u_{p-1}}|Z_{1,u_p}^{u_p},\mathcal{Q}_{1,u_p}^{i})\Big) \tag{105b}$$
$$=\mathsf{N}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)-\sum_{p\in[2:\mathsf{N}]}H(\mathcal{Q}_{1,u_p}^{i}|Z_{1,u_p}^{u_p})$$
$$-\sum_{p\in[2:\mathsf{N}]}H(\mathcal{Q}_{1,u_p}^{u_2},\ldots,\mathcal{Q}_{1,u_p}^{u_{p-1}}|Z_{1,u_p}^{u_p},\mathcal{Q}_{1,u_p}^{i}). \tag{105c}$$

By considering all permutations of $[\mathsf{N}]$ where the first element is i and summing all inequalities as (105c), we can obtain

$$\begin{aligned}H(X_2) &\geq \mathsf{N}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\\ &-\frac{1}{(\mathsf{N}-1)!}\sum_{\mathbf{u}:u_1=i}\sum_{p\in[2:\mathsf{N}]}H(\mathcal{Q}^i_{1,u_p}|Z^{u_p}_{1,u_p})\\ &-\frac{1}{(\mathsf{N}-1)!}\sum_{\mathbf{u}:u_1=i}\sum_{p\in[2:\mathsf{N}]}H(\mathcal{Q}^{u_2}_{1,u_p},\ldots,\mathcal{Q}^{u_{p-1}}_{1,u_p}|Z^{u_p}_{1,u_p},\mathcal{Q}^i_{1,u_p}) \qquad (106a)\\ &=\mathsf{N}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)-\sum_{p\in[\mathsf{N}]\setminus\{i\}}H(\mathcal{Q}^i_{1,p}|Z^p_{1,p})-\frac{1}{(\mathsf{N}-1)!}\\ &\sum_{p\in[\mathsf{N}]\setminus\{i\}}\sum_{r\in[2:\mathsf{N}]}\sum_{\mathbf{u}:u_1=i,u_r=p}H(\mathcal{Q}^{u_2}_{1,p},\ldots,\mathcal{Q}^{u_{r-1}}_{1,p}|Z^p_{1,p},\mathcal{Q}^i_{1,p}), \qquad (106b)\end{aligned}$$

where (106b) comes from the re-arrangements on the summations.

To bound the last term in (106b), we now focus on one file F_p where $p\in[\mathsf{N}]\setminus\{i\}$ and bound the following term

$$\sum_{r\in[2:\mathsf{N}]}\sum_{\mathbf{u}:u_1=i,u_r=p}H(\mathcal{Q}^{u_2}_{1,p},\ldots,\mathcal{Q}^{u_{r-1}}_{1,p}|Z^p_{1,p},\mathcal{Q}^i_{1,p}). \qquad (107)$$

We divide the bits in $F_p\setminus(Z^p_{1,p}\cup\mathcal{Q}^i_{1,p})$ into sub-pieces, and denote

$$F_p\setminus(Z^p_{1,p}\cup\mathcal{Q}^i_{1,p})=\{\mathcal{G}_{p,\mathcal{S}}:\mathcal{S}\subseteq([\mathsf{N}]\setminus\{i,p\})\}, \qquad (108)$$

where

$$\mathcal{G}_{p,\mathcal{S}}=\left(F_p\setminus(Z^p_{1,p}\cup\mathcal{Q}^i_{1,p})\right)\cap\left(\cap_{n\in\mathcal{S}}\mathcal{Q}^n_{1,p}\right)\setminus\left(\cup_{n_1\notin\mathcal{S}}\mathcal{Q}^{n_1}_{1,p}\right). \qquad (109)$$

We then define

$$g_t:=\sum_{\mathcal{S}\subseteq([\mathsf{N}]\setminus\{i,p\}):|\mathcal{S}|=t}|\mathcal{G}_{p,\mathcal{S}}|,\ \forall t\in[0:\mathsf{N}-2]. \qquad (110)$$

For each $n\in[\mathsf{N}]\setminus\{i,p\}$, we have $H(\mathcal{Q}^n_{1,p}|Z^p_{1,p},\mathcal{Q}^i_{1,p})\leq H(\mathcal{Q}^n_{1,p})$. Hence, we have

$$\sum_{t\in[0:\mathsf{N}-2]}tg_t\leq\sum_{n\in[\mathsf{N}]\setminus\{i,p\}}q^n_{1,p}. \qquad (111)$$

In addition, since $F_p\setminus(Z^p_{1,p}\cup\mathcal{Q}^i_{1,p})=(F_p\setminus Z^p_{1,p})\setminus(\mathcal{Q}^i_{1,p}\setminus Z^p_{1,p})$, we have

$$\sum_{t\in[0:\mathsf{N}-2]}g_t=\sum_{\mathcal{S}\subseteq[\mathsf{N}]\setminus\{i,p\}}|\mathcal{G}_{p,\mathcal{S}}|=\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}-H(\mathcal{Q}^i_{1,p}|Z^p_{1,p}). \qquad (112)$$

From the above definitions, we can re-write (106b) (as we did to obtain (96d)),

$$\begin{aligned}&\sum_{r\in[2:\mathsf{N}]}\sum_{\substack{\mathbf{u}:\\u_1=i,u_r=p}}H(\mathcal{Q}^{u_2}_{1,p},\ldots,\mathcal{Q}^{u_{r-1}}_{1,p}|Z^p_{1,p},\mathcal{Q}^i_{1,p})\\ &=\sum_{r\in[2:\mathsf{N}]}\sum_{\substack{\mathbf{u}:\\u_1=i,u_r=p}}\sum_{\substack{\mathcal{S}\subseteq([\mathsf{N}]\setminus\{i,p\}):\\\mathcal{S}\cap\{u_2,\ldots,u_{r-1}\}\neq\emptyset}}|\mathcal{G}_{p,\mathcal{S}}| \qquad (113a)\\ &=\sum_{r\in[2:\mathsf{N}]}\sum_{t\in[0:\mathsf{N}-2]}(\mathsf{N}-2)!\frac{\binom{\mathsf{N}-2}{t}-\binom{\mathsf{N}-r-1}{t}}{\binom{\mathsf{N}-2}{t}}g_t \qquad (113b)\\ &=(\mathsf{N}-1)!\sum_{t\in[0:\mathsf{N}-2]}\frac{t}{t+1}g_t. \qquad (113c)\end{aligned}$$

By Fourier-Motzkin elimination on g_t where $t\in[0:\mathsf{N}-2]$ in (113c) with the help of (111) and (112), we obtain for each $h\in[0:\mathsf{N}-3]$,

$$\begin{aligned}&\sum_{r\in[2:\mathsf{N}]}\sum_{\substack{\mathbf{u}:\\u_1=i,u_r=p}}H(\mathcal{Q}^{u_2}_{1,p},\ldots,\mathcal{Q}^{u_{r-1}}_{1,p}|Z^p_{1,p},\mathcal{Q}^i_{1,p})\\ &\leq\frac{(\mathsf{N}-1)!}{(h+1)(h+2)}\sum_{n\in[\mathsf{N}]\setminus\{p,i\}}q^n_{1,p}\\ &+\frac{(\mathsf{N}-1)!h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}-H(\mathcal{Q}^i_{1,p}|Z^p_{1,p})\right). \qquad (114)\end{aligned}$$

Finally, by taking (114) into (106b), we obtain for each $h\in[0:\mathsf{N}-3]$,

$$\begin{aligned}H(X_2)&\geq\mathsf{N}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)-\frac{2}{h+2}\sum_{p\in[\mathsf{N}]\setminus\{i\}}H(\mathcal{Q}^i_{1,p}|Z^p_{1,p})\\ &-\sum_{p\in[\mathsf{N}]\setminus\{i\}}\left\{\frac{\sum_{n\in[\mathsf{N}]\setminus\{p,i\}}q^n_{1,p}}{(h+1)(h+2)}-\frac{h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\} \qquad (115a)\\ &\geq\mathsf{N}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)-\frac{2}{h+2}\sum_{p\in[\mathsf{N}]\setminus\{i\}}q^i_{1,p}\\ &-\sum_{p\in[\mathsf{N}]\setminus\{i\}}\left\{\frac{\sum_{n\in[\mathsf{N}]\setminus\{p,i\}}q^n_{1,p}}{(h+1)(h+2)}-\frac{h}{h+2}\left(\frac{\mathsf{B}}{2}-\frac{y\mathsf{B}}{\mathsf{N}}\right)\right\}, \qquad (115b)\end{aligned}$$

where (115b) follows since $H(\mathcal{Q}^i_{1,p}|Z^p_{1,p})\leq H(\mathcal{Q}^i_{1,p})=q^i_{1,p}$. Hence, we prove (65).

Appendix C
Generalization of the Proof in Section V-C

In Section V-C, we prove Theorem 6 for the case where $\mathsf{K}/2$ is an integer and $2\mathsf{N}/\mathsf{K}$ is also an integer. In the following, we only consider the case where $\mathsf{K}/2$ is not integer and $\frac{\mathsf{N}}{\lfloor\mathsf{K}/2\rfloor}$ is not an integer neither. The proof for the case where $\mathsf{K}/2$ is an integer and $2\mathsf{N}/\mathsf{K}$ is not an integer, or $\mathsf{K}/2$ is not an integer and $\frac{\mathsf{N}}{\lfloor\mathsf{K}/2\rfloor}$ is an integer, can be directly derived from the following proof.

Recall $\mathsf{M}=\frac{\mathsf{N}}{\mathsf{K}}+\frac{2y}{\mathsf{K}}$, where $y\in\left[0,\frac{\mathsf{N}}{2}\right]$. We first fix one user $k\in[\mathsf{K}]$ (assuming now $k=\mathsf{K}$). We can divide the users in $[\mathsf{K}]\setminus\{k\}$ into two groups, and generate an aggregate user for each group. Denoted the two aggregate users by k_1 and k_2, respectively. The cache size of each aggregate user is $\mathsf{M}\mathsf{B}\times\frac{\mathsf{K}-1}{2}$. In addition, the demanded files of aggregate users k_1 and k_2 are the union sets of the demanded files of users in $[(\mathsf{K}-1)/2]$ and of users in $[(\mathsf{K}+1)/2:\mathsf{K}-1]$, respectively.

By denoting $\mathsf{N}_1:=\lfloor 2\mathsf{N}/\mathsf{K}\rfloor\lfloor\mathsf{K}/2\rfloor$, we divide files in $[\mathsf{N}_1]$ into $\lfloor 2\mathsf{N}/\mathsf{K}\rfloor$ non-overlapping groups, each of which contains $\lfloor\mathsf{K}/2\rfloor$ files. Each aggregate user demands one group of files.

We assume that the caches of aggregate users k_1 and k_2 are A^1_1 and A^1_2. The transmitted packets by aggregate users k_1 and k_2 are denoted by X'_1 and X'_2, and the transmitted packets by user $k=\mathsf{K}$ are denoted by X_k, such that from (X'_2,X_k,A^1_1) aggregate user k_1 can decode the files in group

1 and from (X_1', X_k, A_2^1) aggregate user k_2 can also decode the files in group 1. We then construct the cache configurations of aggregate users k_1 and k_2 by a $\lfloor 2\mathsf{N}/\mathsf{K} \rfloor$-ary tree, as we did in Section V-B.

In the first approach, when we consider A_1^i where $i \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor]$ (cache of aggregate user k_1 where from (X_2', X_k, A_1^i), the files in group i can be decoded), by constructing a genie-aided super-user as in (51) (the cache of this super-user is denoted by A), by Fano's inequality,

$$\begin{aligned} &H(F_1, \ldots, F_{\mathsf{N}} | \{F_\ell : \ell \in [\mathsf{N}_1 + 1 : \mathsf{N}]\}) \\ &\leq H(X_1') + H(X_k) + H(A | \{F_\ell : \ell \in [\mathsf{N}_1 + 1 : \mathsf{N}]\}). \end{aligned} \tag{116}$$

By considering one permutation of $[\lfloor 2\mathsf{N}/\mathsf{K} \rfloor]$, denoted by $\mathbf{u} = (u_1, \ldots, u_{\lfloor 2\mathsf{N}/\mathsf{K} \rfloor})$ where $u_1 = i$, by the similar derivations of (70) and (71), we obtain

$$H(X_1') + H(X_k) \geq \left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}} \right) \lfloor \mathsf{K}/2 \rfloor ; \tag{117}$$

$$H(X_1') + H(X_k) \geq \lfloor \mathsf{K}/2 \rfloor \, \mathsf{B} - \lfloor \mathsf{K}/2 \rfloor \, \frac{4y\mathsf{B}}{\mathsf{N}} + q_{1,u_2}^i. \tag{118}$$

By considering all permutations of $[\lfloor 2\mathsf{N}/\mathsf{K} \rfloor]$ where the first element is i to develop (116) as we did in (52e), and by the similar derivation of (72), we obtain

$$\begin{aligned} &H(X_1') + H(X_k) \\ &\geq \left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}} \right) \mathsf{N}_1 - \frac{2}{h+2} \left((\lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 1) \frac{2y\mathsf{B}}{\mathsf{N}} \lfloor \mathsf{K}/2 \rfloor \right) \\ &+ \frac{2}{h+2} \sum_{p \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor] \setminus \{i\}} q_{1,p}^i \\ &- \frac{(\lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 1)(\lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 2)}{(h+1)(h+2)} \frac{2y\mathsf{B}}{\mathsf{N}} \lfloor \mathsf{K}/2 \rfloor \\ &- \frac{(\lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 1)h}{h+2} \left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}} \right) \lfloor \mathsf{K}/2 \rfloor \\ &+ \frac{\lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 2}{(h+1)(h+2)} \sum_{p \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor] \setminus \{i\}} q_{1,p}^i \\ &\geq \frac{\mathsf{N}_1}{\mathsf{N}} \left\{ \left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}} \right) \mathsf{N} - \frac{2}{h+2} \left((2\mathsf{N}/\mathsf{K} - 1) \frac{2y\mathsf{B}}{\mathsf{N}} \frac{\mathsf{K}}{2} \right) \right. \\ &- \frac{(2\mathsf{N}/\mathsf{K} - 1)(2\mathsf{N}/\mathsf{K} - 2)}{(h+1)(h+2)} \frac{2y\mathsf{B}}{\mathsf{N}} \frac{\mathsf{K}}{2} - \frac{(2\mathsf{N}/\mathsf{K} - 1)h}{h+2} \\ &\left. \left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}} \right) \frac{\mathsf{K}}{2} \right\} + \left(\frac{2}{h+2} + \frac{\lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 2}{(h+1)(h+2)} \right) \\ &\sum_{p \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor] \setminus \{i\}} q_{1,p}^i, \ \forall h \in [0 : \lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 3] , \end{aligned} \tag{119a}$$

where (119a) follows since

$$\frac{\mathsf{N}}{\mathsf{N}_1} (\lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 1) \lfloor \mathsf{K}/2 \rfloor = \mathsf{N} - \frac{\mathsf{N}}{\lfloor 2\mathsf{N}/\mathsf{K} \rfloor} \leq (2\mathsf{N}/\mathsf{K} - 1) \frac{\mathsf{K}}{2}. \tag{120}$$

By considering all $i \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor]$ to bound $H(X_1') + H(X_k)$, and all $j \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor]$ to bound $H(X_2') + H(X_k)$, we sum all inequalities as (119a) to obtain (by the similar derivation of (73)),

$$\begin{aligned} \mathsf{R}_{\mathrm{u,c}}^\star \mathsf{B} + H(X_k) &\geq \frac{\mathsf{N}_1}{\mathsf{N}} \left\{ \left(\mathsf{B} - \frac{2y\mathsf{B}}{\mathsf{N}} \right) \mathsf{N} - \frac{4}{h+2} \right. \\ &\left((2\mathsf{N}/\mathsf{K} - 1) \frac{2y\mathsf{B}}{\mathsf{N}} \frac{\mathsf{K}}{2} \right) - \frac{(2\mathsf{N}/\mathsf{K} - 1)(2\mathsf{N}/\mathsf{K} - 2)}{(h+1)(h+2)} \frac{4y\mathsf{B}}{\mathsf{N}} \frac{\mathsf{K}}{2} \\ &- \frac{h(2\mathsf{N}/\mathsf{K} - 1)}{(h+2)} \left(\mathsf{B} - \frac{2y\mathsf{B}}{\mathsf{N}} \right) \frac{\mathsf{K}}{2} \Bigg\} + \Bigg(\frac{2}{(h+2) \lfloor 2\mathsf{N}/\mathsf{K} \rfloor} \\ &+ \frac{\lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 2}{(\lfloor 2\mathsf{N}/\mathsf{K} \rfloor)(h+1)(h+2)} \Bigg) \Bigg(\sum_{i \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor]} \sum_{p \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor] \setminus \{i\}} q_{1,p}^i \\ &+ \sum_{j \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor]} \sum_{p \in [\lfloor 2\mathsf{N}/\mathsf{K} \rfloor] \setminus \{j\}} q_{2,p}^j \Bigg) , \ \forall h \in [0 : \lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 3] . \end{aligned} \tag{121}$$

Similarly, in the second approach, when we consider (X_2', X_k) and the same permutation as the one to derive (117) and (118), by constructing a genie-aided super-user as in (59) (the cache of this super-user is denoted by A'), by Fano's inequality,

$$\begin{aligned} &H(F_1, \ldots, F_{\mathsf{N}} | \{F_\ell : \ell \in [\mathsf{N}_1 + 1 : \mathsf{N}]\}) \\ &\leq H(X_2') + H(X_k) + H(A' | \{F_\ell : \ell \in [\mathsf{N}_1 + 1 : \mathsf{N}]\}). \end{aligned} \tag{122}$$

By the similar derivations of (74) and (75), we obtain

$$H(X_2') + H(X_k) \geq \left(\frac{\mathsf{B}}{2} - \frac{y\mathsf{B}}{\mathsf{N}} \right) \lfloor \mathsf{K}/2 \rfloor ; \tag{123}$$

$$H(X_2') + H(X_k) \geq \lfloor \mathsf{K}/2 \rfloor \, \mathsf{B} - \lfloor \mathsf{K}/2 \rfloor \, \frac{2y\mathsf{B}}{\mathsf{N}} - q_{1,u_2}^i. \tag{124}$$

In addition, by considering all permutations to bound $H(X_1') + H(X_k)$ and all permutations to bound $H(X_2') + H(X_k)$, we sum all inequalities to obtain (by the similar derivation of (76)),

$$\begin{aligned} &\mathsf{R}_{\mathrm{u,c}}^\star \mathsf{B} + H(X_k) \\ &\geq \frac{\mathsf{N}_1}{\mathsf{N}} \left\{ \mathsf{N}\mathsf{B} - 2y\mathsf{B} - \frac{h \left(\frac{2\mathsf{N}}{\mathsf{K}} - 1 \right)}{(h+2)} \left(\mathsf{B} - \frac{2y\mathsf{B}}{\mathsf{N}} \right) \frac{\mathsf{K}}{2} \right\} \\ &- \left(\frac{2}{(h+2) \lfloor 2\mathsf{N}/\mathsf{K} \rfloor} + \frac{\lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 2}{(h+1)(h+2) \lfloor 2\mathsf{N}/\mathsf{K} \rfloor} \right) \\ &\left(\sum_{j \in \left[\frac{2\mathsf{N}}{\mathsf{K}} \right]} \sum_{p \in \left[\frac{2\mathsf{N}}{\mathsf{K}} \right] \setminus \{j\}} q_{2,p}^j + \sum_{i \in \left[\frac{2\mathsf{N}}{\mathsf{K}} \right]} \sum_{p \in \left[\frac{2\mathsf{N}}{\mathsf{K}} \right] \setminus \{i\}} q_{1,p}^i \right) , \\ &\forall h \in [0 : \lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 3] . \end{aligned} \tag{125}$$

By summing (117) and (123), summing (118) and (124), and summing (121) and (125), we obtain

$$\mathsf{R}_{\mathrm{u,c}}^\star \mathsf{B} + H(X_k) \geq \left(\mathsf{B} - \frac{2y\mathsf{B}}{\mathsf{N}} \right) \lfloor \mathsf{K}/2 \rfloor ; \tag{126a}$$

$$\mathsf{R}_{\mathrm{u,c}}^\star \mathsf{B} + H(X_k) \geq \left(2\mathsf{B} - \frac{6y\mathsf{B}}{\mathsf{N}} \right) \lfloor \mathsf{K}/2 \rfloor ; \tag{126b}$$

$$\begin{aligned} &\mathsf{R}_{\mathrm{u,c}}^\star \mathsf{B} + H(X_k) \\ &\geq \frac{\mathsf{N}_1}{\mathsf{N}} \left\{ \mathsf{N}\mathsf{B} - 2y\mathsf{B} - \frac{2}{h+2} \left((2\mathsf{N}/\mathsf{K} - 1) \frac{2y\mathsf{B}}{\mathsf{N}} \frac{\mathsf{K}}{2} \right) \right. \\ &- \frac{(2\mathsf{N}/\mathsf{K} - 1)(2\mathsf{N}/\mathsf{K} - 2)}{(h+1)(h+2)} \frac{2y\mathsf{B}}{\mathsf{N}} \frac{\mathsf{K}}{2} \\ &\left. - \frac{h(2\mathsf{N}/\mathsf{K} - 1)}{(h+2)} \left(\mathsf{B} - \frac{2y\mathsf{B}}{\mathsf{N}} \right) \frac{\mathsf{K}}{2} \right\} , \ \forall h \in [0 : \lfloor 2\mathsf{N}/\mathsf{K} \rfloor - 3] . \end{aligned} \tag{126c}$$

Finally we consider all $k \in [\mathsf{K}]$ and sum inequalities as (126), to obtain (recall that $\mathsf{R}^{\star}_{\mathrm{u,c}}\mathsf{B} \geq \sum_{k\in[\mathsf{K}]} H(X_k)$),

$$\mathsf{R}^{\star}_{\mathrm{u,c}}\mathsf{B} \geq \frac{\mathsf{K}}{2\lceil \mathsf{K}/2\rceil}\left(\mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}\right)\lfloor \mathsf{K}/2\rfloor$$
$$= \frac{\lfloor \mathsf{K}/2\rfloor}{\lceil \mathsf{K}/2\rceil}\left(\mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}\right)\frac{\mathsf{K}}{2}; \tag{127a}$$
$$\mathsf{R}^{\star}_{\mathrm{u,c}}\mathsf{B} \geq \frac{\mathsf{K}}{2\lceil \mathsf{K}/2\rceil}\left(2\mathsf{B}-\frac{6y\mathsf{B}}{\mathsf{N}}\right)\lfloor \mathsf{K}/2\rfloor$$
$$= \frac{\lfloor \mathsf{K}/2\rfloor}{\lceil \mathsf{K}/2\rceil}\left(2\mathsf{B}-\frac{6y\mathsf{B}}{\mathsf{N}}\right)\frac{\mathsf{K}}{2}; \tag{127b}$$
$$\mathsf{R}^{\star}_{\mathrm{u,c}}\mathsf{B} \geq \frac{\mathsf{K}}{2\lceil \mathsf{K}/2\rceil}\frac{\mathsf{N}_1}{\mathsf{N}}\left\{\mathsf{N}\mathsf{B}-2y\mathsf{B}-\frac{2}{h+2}\right.$$
$$\left((2\mathsf{N}/\mathsf{K}-1)\frac{2y\mathsf{B}}{\mathsf{N}}\frac{\mathsf{K}}{2}\right)-\frac{(2\mathsf{N}/\mathsf{K}-1)(2\mathsf{N}/\mathsf{K}-2)}{(h+1)(h+2)}\frac{2y\mathsf{B}}{\mathsf{N}}\frac{\mathsf{K}}{2}$$
$$\left.-\frac{h(2\mathsf{N}/\mathsf{K}-1)}{(h+2)}\left(\mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}\right)\frac{\mathsf{K}}{2}\right\}$$
$$= \frac{\lfloor \mathsf{K}/2\rfloor}{\lceil \mathsf{K}/2\rceil}\frac{\lfloor 2\mathsf{N}/\mathsf{K}\rfloor}{2\mathsf{N}/\mathsf{K}}\left\{\mathsf{N}\mathsf{B}-2y\mathsf{B}-\frac{2}{h+2}\right.$$
$$\left((2\mathsf{N}/\mathsf{K}-1)\frac{2y\mathsf{B}}{\mathsf{N}}\frac{\mathsf{K}}{2}\right)-\frac{(2\mathsf{N}/\mathsf{K}-1)(2\mathsf{N}/\mathsf{K}-2)}{(h+1)(h+2)}\frac{2y\mathsf{B}}{\mathsf{N}}\frac{\mathsf{K}}{2}$$
$$\left.-\frac{h(2\mathsf{N}/\mathsf{K}-1)}{(h+2)}\left(\mathsf{B}-\frac{2y\mathsf{B}}{\mathsf{N}}\right)\frac{\mathsf{K}}{2}\right\},\ \forall h\in[0:\lfloor 2\mathsf{N}/\mathsf{K}\rfloor-3], \tag{127c}$$

where (127c) comes from (recall that $\mathsf{N}_1 := \lfloor 2\mathsf{N}/\mathsf{K}\rfloor\lfloor \mathsf{K}/2\rfloor$),

$$\frac{\mathsf{K}}{2\lceil \mathsf{K}/2\rceil}\frac{\mathsf{N}_1}{\mathsf{N}} = \frac{\mathsf{K}}{2\lceil \mathsf{K}/2\rceil}\frac{\lfloor 2\mathsf{N}/\mathsf{K}\rfloor\lfloor \mathsf{K}/2\rfloor}{\mathsf{N}} = \frac{\lfloor \mathsf{K}/2\rfloor}{\lceil \mathsf{K}/2\rceil}\frac{\lfloor 2\mathsf{N}/\mathsf{K}\rfloor}{2\mathsf{N}/\mathsf{K}}. \tag{128}$$

Hence, we prove Theorem 6.

Appendix D
Proof of Theorem 2

We first provide a direct upper bound of the achieved load of Scheme A in Theorem 1, since $\frac{\binom{\mathsf{U}}{t}-\binom{\mathsf{U}-\mathsf{N}}{t}}{\binom{\mathsf{U}}{t-1}} \leq \frac{\binom{\mathsf{U}}{t}}{\binom{\mathsf{U}}{t-1}} = \frac{\mathsf{U}-t+1}{t}$.

Lemma 3. *The achieved load of Scheme A in Theorem 1 is upper bound by the lower convex envelop of* $(\mathsf{N}/\mathsf{K},\mathsf{N})$ *and*

$$\left(\frac{\mathsf{N}+t-1}{\mathsf{K}},\frac{\mathsf{U}-t+1}{t}\right),\ \forall t\in[\mathsf{U}+1]. \tag{129}$$

We then introduce the following lemma, whose proof is in Appendix E.

Lemma 4. *The multiplicative gap between the lower convex envelop of the memory-load tradeoff* $\left(\frac{\mathsf{N}+t_1-1}{\mathsf{K}},\frac{\mathsf{U}-t_1+1}{t_1}\right)$ *where* $t_1\in[\mathsf{U}]$*, and the lower convex envelop of the memory-load tradeoff* $\left(\frac{\mathsf{N}t}{\mathsf{K}},\frac{\mathsf{K}-t}{t+1}\right)$ *where* $t\in[2:\mathsf{K}]$*, is at most* 3 *when* $\mathsf{M}\geq\frac{2\mathsf{N}}{\mathsf{K}}$.

We then prove the two cases in Theorem 2, where $\mathsf{N}\geq\mathsf{K}$ and $\mathsf{N}<\mathsf{K}$.

A. $\mathsf{N}\geq\mathsf{K}$

Converse. It was proved in [9] that for the shared-link caching model with $\mathsf{N}\geq\mathsf{K}$, the lower convex envelope of the corner points $\left(\frac{\mathsf{N}t}{\mathsf{K}},\frac{\mathsf{K}-t}{t+1}\right)$, where $t\in[0:\mathsf{K}]$, achieved by the MAN caching scheme in [4] is order optimal to within a factor of 2. In addition, it was proved in [11] that these corner points are successively convex. Hence, when $\mathsf{M}\geq 2\mathsf{N}/\mathsf{K}$, the lower convex envelop of $\left(\frac{\mathsf{N}t}{\mathsf{K}},\frac{\mathsf{K}-t}{t+1}\right)$, where $t\in[2:\mathsf{K}]$ is order optimal to within a factor of 2. We will also use this converse in our model. Hence, for $\mathsf{M}\in[2\mathsf{N}/\mathsf{K},\mathsf{N}]$, $\mathsf{R}^{\star}$ is lower bounded by the lower convex envelope $\left(\frac{\mathsf{N}t}{\mathsf{K}},\frac{\mathsf{K}-t}{2(t+1)}\right)$, where $t\in[2:\mathsf{K}]$.

Achievability. From Lemma 4, it can be seen that from the proposed scheme in Theorem 1, we can achieve the lower convex envelop of the memory-load tradeoff $\left(\frac{\mathsf{N}t}{\mathsf{K}},\frac{3(\mathsf{K}-t)}{t+1}\right)$ where $t\in[2:\mathsf{K}]$.

As a result, the proposed scheme in Theorem 1 is order optimal to within a factor of 6 when $\mathsf{N}\geq\mathsf{K}$ and $\mathsf{M}\geq\frac{2\mathsf{N}}{\mathsf{K}}$.

B. $\mathsf{N}<\mathsf{K}$

Converse. It was proved in [36] that for the shared-link caching model with $\mathsf{N}<\mathsf{K}$, the lower convex envelope of the corner points $(0,\mathsf{N})$ and $\left(\frac{\mathsf{N}t}{\mathsf{K}},\frac{\mathsf{K}-t}{t+1}\right)$, where $t\in[\mathsf{K}]$, achieved by the MAN caching scheme in [4] is order optimal to within a factor of 4.

Since the corner points $\left(\frac{\mathsf{N}t}{\mathsf{K}},\frac{\mathsf{K}-t}{t+1}\right)$ where $t\in[\mathsf{K}]$, are successively convex, the lower convex envelop of the MAN caching scheme for $\mathsf{N}<\mathsf{K}$ is as follows. There exists one $t_2\in[\mathsf{K}]$, such that the lower convex envelop of the MAN caching scheme for $\mathsf{M}\in[0,\mathsf{N}t_2/\mathsf{K}]$ is the memory-sharing between $(0,\mathsf{N})$ and $\left(\frac{\mathsf{N}t_2}{\mathsf{K}},\frac{\mathsf{K}-t_2}{t_2+1}\right)$, while the lower convex envelop for $\mathsf{M}\in[\mathsf{N}t_2/\mathsf{K},\mathsf{N}]$ is the lower convex envelop of the successive corner points $\left(\frac{\mathsf{N}t}{\mathsf{K}},\frac{\mathsf{K}-t}{t+1}\right)$ where $t\in[t_2:\mathsf{K}]$. In addition, it is obvious that t_2 is the maximum value among $x\in[\mathsf{K}]$ such that the memory-sharing between $(0,\mathsf{N})$ and $\left(\frac{\mathsf{N}x}{\mathsf{K}},\frac{\mathsf{K}-x}{x+1}\right)$ at the memory $\mathsf{M}'=\frac{\mathsf{N}(x-1)}{\mathsf{K}}$ leads to a lower load than $\frac{\mathsf{K}-x+1}{x}$. More precisely, if we interpolate $(0,\mathsf{N})$ and $\left(\frac{\mathsf{N}x}{\mathsf{K}},\frac{\mathsf{K}-x}{x+1}\right)$ where $x\in[\mathsf{K}]$ to match $\mathsf{M}'=\frac{\mathsf{N}(x-1)}{\mathsf{K}}$, the achieved load is

$$-\frac{\mathsf{N}-\frac{\mathsf{K}-x}{x+1}}{\frac{\mathsf{N}x}{\mathsf{K}}}\frac{\mathsf{N}(x-1)}{\mathsf{K}}+\mathsf{N} = \frac{(\mathsf{K}-x)(x-1)}{x(x+1)}+\frac{\mathsf{N}}{x}.$$

Hence, we have

$$t_2 := \arg\max_{x\in[\mathsf{K}]}\left\{\frac{(\mathsf{K}-x)(x-1)}{x(x+1)}+\frac{\mathsf{N}}{x}\leq\frac{\mathsf{K}-x+1}{x}\right\} \tag{130a}$$
$$= \left\lfloor\frac{2\mathsf{K}-\mathsf{N}+1}{\mathsf{N}+1}\right\rfloor. \tag{130b}$$

We then interpolate $(0,\mathsf{N})$ and $\left(\frac{\mathsf{N}t_2}{\mathsf{K}},\frac{\mathsf{K}-t_2}{t_2+1}\right)$ to match $\mathsf{M}_1=\mathsf{N}/\mathsf{K}$, to get the memory-load tradeoff

$$(\mathsf{M}_1,\mathsf{R}_1) = \left(\frac{\mathsf{N}}{\mathsf{K}},\mathsf{N}-\frac{\mathsf{N}-\frac{\mathsf{K}-t_2}{t_2+1}}{t_2}\right). \tag{131}$$

Hence, it is equivalent to say that the lower convex envelop of the achieved memory-load tradeoffs by the MAN caching scheme for $\mathsf{M} \geq \mathsf{N}/\mathsf{K}$ also has two regimes.

1) $\mathsf{M} \in \left[\frac{\mathsf{N}}{\mathsf{K}}, \frac{\mathsf{N}t_2}{\mathsf{K}}\right]$. The lower convex envelop is the memory-sharing between $(\mathsf{M}_1, \mathsf{R}_1)$ and $\left(\frac{\mathsf{N}t_2}{\mathsf{K}}, \frac{\mathsf{K}-t_2}{t_2+1}\right)$.
2) $\mathsf{M} \in \left[\frac{\mathsf{N}t_2}{\mathsf{K}}, \mathsf{N}\right]$. The lower convex envelop of the MAN scheme is the lower convex envelop of the corner points $\left(\frac{\mathsf{N}t}{\mathsf{K}}, \frac{\mathsf{K}-t}{t+1}\right)$, where $t \in [t_2 : \mathsf{K}]$.

Since the MAN scheme is order optimal to within a factor of 4, $\mathsf{R}^\star$ is lower bounded by the lower convex envelope of the corner points $\left(\mathsf{M}_1, \frac{\mathsf{R}_1}{4}\right)$ and $\left(\frac{\mathsf{N}t}{\mathsf{K}}, \frac{\mathsf{K}-t}{4(t+1)}\right)$, where $t \in [t_2 : \mathsf{K}]$.

Achievability. Let us first focus on $\mathsf{M} = \mathsf{N}/\mathsf{K}$. The achieved load by the proposed scheme in Theorem 1 is N. In the following, we will prove $\mathsf{N} \leq 2\mathsf{R}_1$. More precisely,

$$\begin{aligned}
\mathsf{N} - 2\mathsf{R}_1 &= 2\frac{\mathsf{N} - \frac{\mathsf{K}-t_2}{t_2+1}}{t_2} - \mathsf{N} \\
&= \frac{2\mathsf{N}(t_2+1) - 2(\mathsf{K}-t_2) - \mathsf{N}t_2(t_2+1)}{t_2(t_2+1)} \\
&= \frac{-\mathsf{N}t_2^2 + (\mathsf{N}+2)t_2 - 2(\mathsf{K}-\mathsf{N})}{t_2(t_2+1)} \\
&= \frac{-t_2(\mathsf{N}t_2 - \mathsf{N} - 2) - 2(\mathsf{K}-\mathsf{N})}{t_2(t_2+1)} \\
&= \frac{-(\mathsf{N}t_2 - \mathsf{N} - 2) - \frac{2(\mathsf{K}-\mathsf{N})}{t_2}}{(t_2+1)}. \qquad (132)
\end{aligned}$$

We consider the following two cases.

1) $t_2 = 1$. From (132), we have

$$\mathsf{N} - 2\mathsf{R}_1 = \frac{2 - 2(\mathsf{K}-\mathsf{N})}{2} \leq 0, \qquad (133)$$

which follows $\mathsf{K} > \mathsf{N}$.

2) $t_2 > 1$. From (132), we have

$$\mathsf{N} - 2\mathsf{R}_1 \leq \frac{-(2\mathsf{N} - \mathsf{N} - 2) - \frac{2(\mathsf{K}-\mathsf{N})}{t_2}}{t_2+1} < 0, \qquad (134)$$

which follows $\mathsf{N} \geq 2$ and $\mathsf{K} > \mathsf{N}$.

Hence, from the proposed scheme in Theorem 1, we can achieve $(\mathsf{M}_1, 2\mathsf{R}_1)$. In addition, from Lemma 4, it can be seen that from the proposed scheme in Theorem 1, we can achieve the lower convex envelop of the memory-load tradeoff $\left(\frac{\mathsf{N}t}{\mathsf{K}}, \frac{3(\mathsf{K}-t)}{t+1}\right)$ where $t \in [t_2 : \mathsf{K}]$.

As a result, the proposed scheme in Theorem 1 is order optimal to within a factor of 12 when $\mathsf{N} < \mathsf{K}$.

Appendix E
Proof of Lemma 4

It was proved in [11] that the corner points $\left(\frac{\mathsf{N}t}{\mathsf{K}}, \frac{\mathsf{K}-t}{t+1}\right)$ where $t \in [0 : \mathsf{K}]$ are successively convex, i.e., for each memory size $\mathsf{M} \in \left[\frac{\mathsf{N}t}{\mathsf{K}}, \frac{\mathsf{N}(t+1)}{\mathsf{K}}\right]$ where $t \in [0 : \mathsf{K}-1]$, the lower convex envelop is obtained by memory-sharing between $\left(\frac{\mathsf{N}t}{\mathsf{K}}, \frac{\mathsf{K}-t}{t+1}\right)$ and $\left(\frac{\mathsf{N}(t+1)}{\mathsf{K}}, \frac{\mathsf{K}-t-1}{t+2}\right)$. Hence, in order to prove Lemma 4, in the following we prove from $\left(\frac{\mathsf{N}+t_1-1}{\mathsf{K}}, \frac{\mathsf{U}-t_1+1}{t_1}\right)$ where $t_1 \in [\mathsf{U}]$, we can achieve $\left(\frac{\mathsf{N}t}{\mathsf{K}}, 3\frac{\mathsf{K}-t}{(t+1)}\right)$ for each $t \in [2 : \mathsf{K}]$.

We now focus on one $t \in [2 : \mathsf{K}]$. We let $t_1 = \mathsf{N}(t-1)+1$ such that the memory size is

$$\frac{\mathsf{N}+t_1-1}{\mathsf{K}} = \frac{\mathsf{N}+\mathsf{N}(t-1)+1-1}{\mathsf{K}} = \frac{\mathsf{N}t}{\mathsf{K}}. \qquad (135)$$

The achieved load is

$$\begin{aligned}
\frac{\mathsf{U}-t_1+1}{t_1} &= \frac{\mathsf{U} - \frac{\mathsf{U}(t-1)}{\mathsf{K}-1}}{\frac{\mathsf{U}(t-1)}{\mathsf{K}-1}+1} \\
&= \frac{\mathsf{U}(\mathsf{K}-1) - \mathsf{U}(t-1)}{\mathsf{U}(t-1) + (\mathsf{K}-1)} \\
&= \frac{\mathsf{K}-t}{t-1+\frac{\mathsf{K}-1}{\mathsf{N}}} \\
&\leq \frac{\mathsf{K}-t}{t-1} \\
&\leq 3\frac{\mathsf{K}-t}{t+1}, \qquad (136)
\end{aligned}$$

where (136) comes from $t \geq 2$. Hence, we prove the proof of Lemma 4.

Appendix F
Proof of Corollary 1

Recall that for the two-user system, the achieved corner points of Scheme A are $\left(\frac{\mathsf{N}+t-1}{2}, \frac{\mathsf{N}-t+1}{t}\right)$, where $t \in [\mathsf{N}+1]$. The achieved corner points of Scheme B are $\left(\frac{\mathsf{N}}{2} + \frac{\mathsf{N}t'}{2(\mathsf{N}+t'-1)}, \frac{\mathsf{N}(\mathsf{N}-1)}{(t'+1)(\mathsf{N}+t'-1)}\right)$ and $(\mathsf{N}, 0)$, where $t' \in [0 : \mathsf{N}-1]$.

To prove Scheme B is better than Scheme A for the two-user system, we prove that for each $t \in [\mathsf{N}]$, by memory-sharing between $\left(\frac{\mathsf{N}}{2} + \frac{\mathsf{N}t'}{2(\mathsf{N}+t'-1)}, \frac{\mathsf{N}(\mathsf{N}-1)}{(t'+1)(\mathsf{N}+t'-1)}\right)$ and $(\mathsf{N}, 0)$, where $t' = t-1$, we can obtain $\left(\frac{\mathsf{N}+t-1}{2}, \frac{\mathsf{N}-t+1}{t}\right)$. More precisely, we let $\alpha = \frac{(\mathsf{N}+t'-1)(\mathsf{N}-t')}{\mathsf{N}(\mathsf{N}-1)}$. We have

$$\begin{aligned}
&\alpha\left(\frac{\mathsf{N}}{2} + \frac{\mathsf{N}t'}{2(\mathsf{N}+t'-1)}\right) + (1-\alpha)\mathsf{N} \\
&= \frac{(\mathsf{N}+t'-1)(\mathsf{N}-t')}{\mathsf{N}(\mathsf{N}-1)} \frac{\mathsf{N}(\mathsf{N}+2t'-1)}{2(\mathsf{N}+t'-1)} + \frac{t'(t'-1)}{\mathsf{N}(\mathsf{N}-1)}\mathsf{N} \\
&= \frac{(\mathsf{N}+2t'-1)(\mathsf{N}-t')}{2(\mathsf{N}-1)} + \frac{t'(t'-1)}{\mathsf{N}-1} \\
&= \frac{(\mathsf{N}-1)(\mathsf{N}-t')}{2(\mathsf{N}-1)} \\
&= \frac{\mathsf{N}-t+1}{2}; \qquad (137)
\end{aligned}$$

$$\alpha\frac{\mathsf{N}(\mathsf{N}-1)}{(t'+1)(\mathsf{N}+t'-1)} + (1-\alpha)\times 0 = \frac{\mathsf{N}-t'}{t'+1} = \frac{\mathsf{N}-t+1}{t}. \qquad (138)$$

Appendix G
Proof of Theorem 5

A. Optimality in Theorem 5

When $\mathsf{N} = 2$, it can be easily checked that the converse bound in Theorem 4 is a piecewise curve with corner points $\left(\frac{\mathsf{N}}{2}, \mathsf{N}\right)$, $\left(\frac{3\mathsf{N}}{4}, \frac{1}{2}\right)$, and $(\mathsf{N}, 0)$, which can be achieved

by Scheme B in (10). Hence, in the following, we focus on $\mathsf{N} > 2$.

Recall that $\mathsf{M} = \frac{\mathsf{N}}{2} + y$. For $0 \le y \le \frac{1}{2}$, from the converse bound in (11) with $h = 0$, we have

$$\begin{aligned}\mathsf{R}_{\mathrm{u}}^{\star} &\ge \mathsf{N} - 2y - \frac{4y + (\mathsf{N}-1)h}{h+2} \\ &\quad + \frac{h^2(\mathsf{N}-1) - \mathsf{N}(\mathsf{N}-3) + h(\mathsf{N}+1)}{(h+1)(h+2)} \frac{2y}{\mathsf{N}} \\ &= \mathsf{N} - 2y - 2y - y(\mathsf{N}-3) \\ &= \mathsf{N} - y(\mathsf{N}+1). \end{aligned} \tag{139}$$

In other words, when $\frac{\mathsf{N}}{2} \le \mathsf{M} \le \frac{\mathsf{N}+1}{2}$, the converse bound on $\mathsf{R}_{\mathrm{u}}^{\star}$ in (139) is a straight line between $\left(\frac{\mathsf{N}}{2}, \mathsf{N}\right)$ and $\left(\frac{\mathsf{N}+1}{2}, \frac{\mathsf{N}-1}{2}\right)$. In addition, Scheme B in (10) achieves $\left(\frac{\mathsf{N}}{2}, \mathsf{N}\right)$ with $t' = 0$, and $\left(\frac{\mathsf{N}+1}{2}, \frac{\mathsf{N}-1}{2}\right)$ with $t' = 1$. Hence, we prove Scheme B is optimal under the constraint of uncoded cache placement when $\frac{\mathsf{N}}{2} \le \mathsf{M} \le \frac{\mathsf{N}+1}{2}$.

For $\frac{2\mathsf{N}}{3} \le \mathsf{M} \le \frac{3\mathsf{N}}{4}$ (i.e., $\frac{\mathsf{N}}{6} \le y \le \frac{\mathsf{N}}{4}$), from the converse bound in (12)

$$\mathsf{R}_{\mathrm{u}}^{\star} \ge 2 - \frac{6y}{\mathsf{N}} = 5 - \frac{6\mathsf{M}}{\mathsf{N}}. \tag{140}$$

By noticing that $\frac{\mathsf{N}(3\mathsf{N}-5)}{2(2\mathsf{N}-3)} \ge \frac{2\mathsf{N}}{3}$ when $\mathsf{N} \ge 3$, from (140), it can be seen that when $\mathsf{M} = \frac{\mathsf{N}(3\mathsf{N}-5)}{2(2\mathsf{N}-3)}$, $\mathsf{R}_{\mathrm{u}}^{\star} \ge \frac{\mathsf{N}}{2\mathsf{N}-3}$, coinciding with Scheme B in (10) with $t' = \mathsf{N} - 2$. When $\mathsf{M} = \frac{3\mathsf{N}}{4}$, $\mathsf{R}_{\mathrm{u}}^{\star} \ge \frac{1}{2}$, coinciding with Scheme B in (10) with $t' = \mathsf{N} - 1$. Hence, we prove that Scheme B is optimal under the constraint of uncoded cache placement when $\frac{\mathsf{N}(3\mathsf{N}-5)}{2(2\mathsf{N}-3)} \le \mathsf{M} \le \frac{3\mathsf{N}}{4}$.

Finally, for $\frac{3\mathsf{N}}{4} \le \mathsf{M} \le \mathsf{N}$ (i.e., $\frac{\mathsf{N}}{4} \le y \le \frac{\mathsf{N}}{2}$), from the converse bound in (13), we have

$$\mathsf{R}_{\mathrm{u}}^{\star} \ge 1 - \frac{2y}{\mathsf{N}} = 2 - \frac{2\mathsf{M}}{\mathsf{N}}. \tag{141}$$

From (141), it can be seen that when $\mathsf{M} = \frac{3\mathsf{N}}{4}$, $\mathsf{R}_{\mathrm{u}}^{\star} \ge \frac{1}{2}$, coinciding with Scheme B in (10) with $t' = \mathsf{N} - 1$. When $\mathsf{M} = \mathsf{N}$, $\mathsf{R}_{\mathrm{u}}^{\star} \ge 0$, which can be also achieved by Scheme B. Hence, we prove that Scheme B is optimal under the constraint of uncoded cache placement when $\frac{3\mathsf{N}}{4} \le \mathsf{M} \le \mathsf{N}$.

B. Order optimality in Theorem 5

From Theorem 4, we can compute that the proposed converse bound is a piecewise curve with the corner points

$$\left(\frac{\mathsf{N}}{2} + \frac{\mathsf{N}h'}{2(\mathsf{N}+2h'-2)}, \frac{(h'-1)(\mathsf{N}+h') + (\mathsf{N}-1)\mathsf{N}}{(h'+1)(\mathsf{N}+2h'-2)}\right), \quad \forall h' \in [0 : \mathsf{N}-2], \tag{142}$$

$\left(\frac{3\mathsf{N}}{4}, \frac{1}{2}\right)$, and $(\mathsf{N}, 0)$.[12] Note that the proposed converse bound is a piecewise linear curve with the above corner points, and that the straight line in the memory-load tradeoff between two achievable points is also achievable by memory-sharing. Hence, in the following, we focus on each corner point of the converse bound, and characterize the multiplicative gap between Scheme B and the converse bound.

Note that in (142), when $h' = 0$, we have $\left(\frac{\mathsf{N}}{2}, \mathsf{N}\right)$; when $h' = 1$, we have $\left(\frac{\mathsf{N}+1}{2}, \frac{\mathsf{N}-1}{2}\right)$; when $h' = \mathsf{N} - 2$, we have $\left(\frac{2\mathsf{N}}{3}, 1\right)$. In addition, in Appendix G-A, we proved the optimality of Scheme B under the constraint of uncoded cache placement when $\mathsf{M} \le \frac{\mathsf{N}+1}{2}$ or when $\mathsf{M} \ge \frac{3\mathsf{N}}{4}$. Hence, in the following, we only need to compare Scheme B and the corner points in (142) where $h' \in [2 : \mathsf{N}-2]$ and $\mathsf{N} \ge 4$.

In Corollary 1, we show that Scheme B is better than Scheme A. We will prove the multiplicative gap between Scheme A and the corner points in (142) where $h' \in [2 : \mathsf{N}-2]$ and $\mathsf{N} \ge 4$, is no more than 3.

Recall that the achieved points of Scheme A for the two-user system are

$$\left(\frac{\mathsf{N}+t-1}{2}, \frac{\mathsf{N}-t+1}{t}\right), \forall t \in [\mathsf{N}+1]. \tag{143}$$

We want to interpolate the achieved points of Scheme A to match the converse bound at the memory size $\mathsf{M} = \frac{\mathsf{N}}{2} + \frac{\mathsf{N}h'}{2(\mathsf{N}+2h'-2)}$ where $h' \in [2 : \mathsf{N}-2]$. By computing

$$\begin{aligned}\frac{\mathsf{N}+t-1}{2} &= \frac{\mathsf{N}}{2} + \frac{\mathsf{N}h'}{2(\mathsf{N}+2h'-2)} \\ \iff t &= \frac{\mathsf{N}h'}{\mathsf{N}+2h'-2} + 1, \end{aligned} \tag{144}$$

and observing $\frac{\mathsf{N}-t+1}{t}$ is non-increasing with t, it can be seen that the achieved load of Scheme A at $\mathsf{M} = \frac{\mathsf{N}}{2} + \frac{\mathsf{N}h'}{2(\mathsf{N}+2h'-2)}$ is lower than

$$\mathsf{R}' = \frac{\mathsf{N} - \frac{\mathsf{N}h'}{\mathsf{N}+2h'-2} + 1}{\frac{\mathsf{N}h}{\mathsf{N}+2h'-2}} = \frac{\mathsf{N}^2 + (\mathsf{N}+2)(h'-1)}{\mathsf{N}h'}. \tag{145}$$

By comparing R' and $\frac{(h'-1)(\mathsf{N}+h')+(\mathsf{N}-1)\mathsf{N}}{(h'+1)(\mathsf{N}+2h'-2)}$, we have

$$\begin{aligned}&\frac{\mathsf{R}'}{\frac{(h'-1)(\mathsf{N}+h')+(\mathsf{N}-1)\mathsf{N}}{(h'+1)(\mathsf{N}+2h'-2)}} \\ &= \frac{\big(\mathsf{N}^2 + (\mathsf{N}+2)(h'-1)\big)(h'+1)(\mathsf{N}+2h'-2)}{\mathsf{N}h'\big((h'-1)(\mathsf{N}+h') + (\mathsf{N}-1)\mathsf{N}\big)}. \end{aligned} \tag{146}$$

In addition, we compute

$$\begin{aligned}&3\mathsf{N}h'\big((h'-1)(\mathsf{N}+h') + (\mathsf{N}-1)\mathsf{N}\big) \\ &- \big(\mathsf{N}^2 + (\mathsf{N}+2)(h'-1)\big)(h'+1)(\mathsf{N}+2h'-2) \\ &= 2\mathsf{N}^3h' - \mathsf{N}^3 - 6\mathsf{N}^2h' - 3\mathsf{N}h'^2 + (\mathsf{N}-4)h'^3 \\ &+ 3\mathsf{N}^2 + 2\mathsf{N}h' + 4h'(h'+1) - 4. \end{aligned} \tag{147}$$

Now we want to prove the RHS of (147) is larger than 0 for $\mathsf{N} \ge 4$ and $h' \in [2 : \mathsf{N}-2]$. More precisely, when $\mathsf{N} = 4$ and $h' = 2$, we can compute the RHS of (147) is equal to 36; when $\mathsf{N} = 5$ and $h' = 2$, the RHS of (147) is equal to 138; when $\mathsf{N} = 5$ and $h' = 3$, the RHS of (147) is equal to 216. Now we only need to consider $\mathsf{N} \ge 6$ and $h' \in [2 : \mathsf{N}-2]$.

When $\mathsf{N} \ge 6$ and $h' \in [2 : \mathsf{N}-2]$, we have

$$2\mathsf{N}^3h' - \mathsf{N}^3 - 6\mathsf{N}^2h' - 3\mathsf{N}h'^2 + (\mathsf{N}-4)h'^3 + 3\mathsf{N}^2$$

[12] The first corner point in (142) is $\left(\frac{\mathsf{N}}{2}, \mathsf{N}\right)$ with $h' = 0$, and the last corner point is $(\mathsf{N}, 0)$. For each $h' \in [\mathsf{N}-3]$, we obtain the corner point in (142) by taking the intersection between the converse bounds in (11) with $h = h' - 1$ and $h = h'$. The corner point in (142) with $h' = \mathsf{N}-2$, is obtained by taking the intersection between the converse bounds in (11) with $h = \mathsf{N}-3$ and the converse bound in (12). The corner point $\left(\frac{3\mathsf{N}}{4}, \frac{1}{2}\right)$ is obtained by taking the intersection between the converse bounds in (12) and (13).

$$
\begin{aligned}
&+ 2\mathsf{N}h' + 4h'(h'+1) - 4\\
&> 2\mathsf{N}^3h' - \mathsf{N}^3 - 6\mathsf{N}^2h' - 3\mathsf{N}h'^2\\
&= (\mathsf{N}^3h' - 6\mathsf{N}^2h') + (0.5\mathsf{N}^3h' - 3\mathsf{N}h'^2) + (0.5\mathsf{N}^3h' - \mathsf{N}^3)\\
&\geq 0. \qquad (148)
\end{aligned}
$$

Hence, we prove

$$
\begin{aligned}
&3\mathsf{N}h'\big((h'-1)(\mathsf{N}+h') + (\mathsf{N}-1)\mathsf{N}\big)\\
&- \big(\mathsf{N}^2 + (\mathsf{N}+2)(h'-1)\big)(h'+1)(\mathsf{N}+2h'-2) > 0. \qquad (149)
\end{aligned}
$$

By taking (149) into (146), we prove that the multiplicative gap between Scheme A and the corner points in (142) where $h' \in [2:\mathsf{N}-2]$ and $\mathsf{N} \geq 4$, is no more than 3.

In conclusion, we prove that Scheme B is order optimal under the constraint of uncoded cache placement to within a factor of 3.

Appendix H
Proof of Theorem 7

In this proof, for the achievability, we consider the load in Lemma 3, which is an upper bound of the achieved load of Scheme A.

We first focus on the case where $\mathsf{N} \leq 6\mathsf{K}$, and compare Scheme A with the shared-link caching converse bound under the constraint of uncoded cache placement (without privacy) in [11]. Recall that when $\mathsf{M} \in \left[\frac{\mathsf{N}}{\mathsf{K}}, \mathsf{N}\right]$, the converse bound in [11] is a piecewise curve with corner points $\left(\frac{\mathsf{N}t}{\mathsf{K}}, \frac{\mathsf{K}-t}{t+1}\right)$, where $t \in [\mathsf{K}]$. It was proved in Appendix D-A that Scheme A can achieve the corner points $\left(\frac{\mathsf{N}t}{\mathsf{K}}, 3\frac{\mathsf{K}-t}{t+1}\right)$, where $t \in [2:\mathsf{K}]$. In addition, when $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}}$, the converse bound in [11] is $\mathsf{R}^\star_{\mathrm{u}} \geq \frac{\mathsf{K}-1}{2}$, while the achieved load of Scheme A is

$$\mathsf{N} \leq 6\mathsf{K} \leq 9(\mathsf{K}-1), \quad \text{when } \mathsf{K} \geq 3.$$

Hence, the multiplicative gap between Scheme A and the converse bound in [11] at $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}}$ is no more than 18. So we prove that $\mathsf{N} \leq 6\mathsf{K}$, Scheme A is order optimal under the constraint of uncoded cache placement within a factor of 18.

In the rest of the proof, we focus on the case where $\mathsf{N} > 6\mathsf{K}$. It was proved in Theorem 2 that when $\mathsf{N} \geq \mathsf{K}$ and $\mathsf{M} \geq \frac{2\mathsf{N}}{\mathsf{K}}$, Scheme A is order optimal to within a factor of 6. Hence, in the following we consider $\frac{\mathsf{N}}{\mathsf{K}} \leq \mathsf{M} \leq \frac{2\mathsf{N}}{\mathsf{K}}$, which is then divided into three memory size regimes, and prove the order optimality of Scheme A separately,

$$
\begin{aligned}
&\text{Regime 1}: \frac{\mathsf{N}}{\mathsf{K}} \leq \mathsf{M} \leq \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_1}{2(\mathsf{N}+\mathsf{K}h_1-\mathsf{K})},\\
&\text{where } h_1 := \left\lfloor \frac{4(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})}{\mathsf{K}(\mathsf{N}-4\mathsf{K}+8)} \right\rfloor; \qquad (150a)\\
&\text{Regime 2}: \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_1}{2(\mathsf{N}+\mathsf{K}h_1-\mathsf{K})} \leq \mathsf{M} \leq\\
&\frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})}, \text{ where } h_2 := \left\lfloor \frac{2\mathsf{N}}{\mathsf{K}} - 2 \right\rfloor; \qquad (150b)\\
&\text{Regime 3}: \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})} \leq \mathsf{M} \leq \frac{2\mathsf{N}}{\mathsf{K}}. \qquad (150c)
\end{aligned}
$$

Note that when $\mathsf{N} > 6\mathsf{K}$, we have $h_1 := \left\lfloor \frac{4(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})}{\mathsf{K}(\mathsf{N}-4\mathsf{K}+8)} \right\rfloor < 10$ and $h_2 := \left\lfloor \frac{2\mathsf{N}}{\mathsf{K}} - 2 \right\rfloor \geq 10$. Thus we have $h_1 < h_2$. In addition, we have

$$
\begin{aligned}
\frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})} &\leq \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}\frac{2\mathsf{N}}{\mathsf{K}} - 2}{2\left(\mathsf{N}+\mathsf{K}\frac{2\mathsf{N}}{\mathsf{K}} - 2\mathsf{K} - \mathsf{K}\right)}\\
&= \frac{4\mathsf{N}}{3\mathsf{K}}. \qquad (151)
\end{aligned}
$$

Hence, the above memory regime division is possible.

From the converse bound in (14), for each $h \in [0:\lfloor 2\mathsf{N}/\mathsf{K} - 3\rfloor]$ we have,

$$
\begin{aligned}
\mathsf{R}^\star_{\mathrm{u,c}} &\geq \frac{\lfloor \mathsf{K}/2 \rfloor}{\lceil \mathsf{K}/2 \rceil} \frac{\lfloor 2\mathsf{N}/\mathsf{K} \rfloor}{2\mathsf{N}/\mathsf{K}} \left\{ \mathsf{N} - 2y - \frac{8y + h(2\mathsf{N}-\mathsf{K})}{2h+4} \right.\\
&\left. + \frac{h^2\mathsf{K}(2\mathsf{N}-\mathsf{K}) - 2\mathsf{N}(2\mathsf{N}-3\mathsf{K}) + h\mathsf{K}(\mathsf{K}+2\mathsf{N})}{(h+1)(h+2)\mathsf{K}\mathsf{N}} y \right\}\\
&\geq \frac{6}{13} \left\{ \mathsf{N} - 2y - \frac{8y + h(2\mathsf{N}-\mathsf{K})}{2h+4} \right.\\
&\left. + \frac{h^2\mathsf{K}(2\mathsf{N}-\mathsf{K}) - 2\mathsf{N}(2\mathsf{N}-3\mathsf{K}) + h\mathsf{K}(\mathsf{K}+2\mathsf{N})}{(h+1)(h+2)\mathsf{K}\mathsf{N}} y \right\}, \qquad (152)
\end{aligned}
$$

where (152) follows since $\mathsf{K} \geq 3$ and $\mathsf{N} > 6\mathsf{K}$.

In Regimes 1 and 2, we will use (152) as the converse bound. In Regime 3, we use the shared-link caching converse bound under the constraint of uncoded cache placement in [11].

A. Regime 1

It can be computed that the converse bound in (152) for $\frac{\mathsf{N}}{\mathsf{K}} \leq \mathsf{M} \leq \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_1}{2(\mathsf{N}+\mathsf{K}h_1-\mathsf{K})}$ is a piecewise curve with the corner points

$$
\begin{aligned}
&\left(\frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h'}{2(\mathsf{N}+\mathsf{K}h'-\mathsf{K})}, \right.\\
&\left. \frac{6}{13} \frac{\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h') + 2\mathsf{N}(2\mathsf{N}-\mathsf{K})}{4(h'+1)(\mathsf{N}+\mathsf{K}h'-\mathsf{K})} \right), \ \forall h' \in [0:h_1], \qquad (153)
\end{aligned}
$$

where $h' = 0$ represents the first corner point where $\mathsf{M} = \mathsf{N}/2$, and each corner point in (153) with h' is obtained by taking the intersection of the converse bounds in (152) between $h = h'-1$ and $h = h'$.

For the achievability, we take the memory-sharing between $\left(\frac{\mathsf{N}}{\mathsf{K}}, \mathsf{N}\right)$ and $\left(\frac{\mathsf{N}+t_3-1}{\mathsf{K}}, \frac{\mathsf{U}-t_3+1}{t_3}\right)$, where $t_3 = 2\mathsf{K}-3$. Notice that

$$\frac{\mathsf{N}+t_3-1}{\mathsf{K}} = \frac{\mathsf{N}+2\mathsf{K}-4}{\mathsf{K}} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{2\mathsf{K}-4}{\mathsf{K}}. \qquad (154)$$

In addition, we have

$$
\begin{aligned}
&\frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_1}{2(\mathsf{N}+\mathsf{K}h_1-\mathsf{K})} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_1}{2(\mathsf{N}+\mathsf{K}h_1-\mathsf{K})} \qquad (155a)\\
&\leq \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}\frac{4(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})}{\mathsf{K}(\mathsf{N}-4\mathsf{K}+8)}}{2(\mathsf{N}+\mathsf{K}\frac{4(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})}{\mathsf{K}(\mathsf{N}-4\mathsf{K}+8)} - \mathsf{K})} \qquad (155b)\\
&= \frac{\mathsf{N}}{\mathsf{K}} + \frac{4\mathsf{N}(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})}{2\big((\mathsf{N}-\mathsf{K})\mathsf{K}(\mathsf{N}-4\mathsf{K}+8) + 4\mathsf{K}(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})\big)} \qquad (155c)
\end{aligned}
$$

$$= \frac{\mathsf{N}}{\mathsf{K}} + \frac{4\mathsf{N}(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})}{2\mathsf{K}\mathsf{N}(\mathsf{N}-\mathsf{K})} \tag{155d}$$

$$= \frac{\mathsf{N}}{\mathsf{K}} + \frac{2\mathsf{K}-4}{\mathsf{K}}, \tag{155e}$$

where (155b) comes from $\frac{\mathsf{N}h_1}{2(\mathsf{N}+\mathsf{K}h_1-\mathsf{K})}$ is increasing with h_1 and $h_1 \le \frac{4(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})}{\mathsf{K}(\mathsf{N}-4\mathsf{K}+8)}$. From (154) and (155e), we can see that this memory-sharing can cover all memory sizes in regime 1.

When $h'=0$, we have the corner point in (153) is $\left(\frac{\mathsf{N}}{2}, \frac{6\mathsf{N}}{13}\right)$, while Scheme A achieves $\left(\frac{\mathsf{N}}{2}, \mathsf{N}\right)$. Hence, the multiplicative gap between Scheme A and the converse is $\frac{13}{6}$.

For each $h' \in [h_1]$, we now interpolate Scheme A between $(\mathsf{M}_1, \mathsf{R}_1) = \left(\frac{\mathsf{N}}{\mathsf{K}}, \mathsf{N}\right)$ and $(\mathsf{M}_2, \mathsf{R}_2) = \left(\frac{\mathsf{N}+t_3-1}{\mathsf{K}}, \frac{\mathsf{U}-t_3+1}{t_3}\right)$ to match the corner point in the converse bound $(\mathsf{M}_3, \mathsf{R}_3) = \left(\frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h'}{2(\mathsf{N}+\mathsf{K}h'-\mathsf{K})}, \frac{6}{13}\frac{\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})}{4(h'+1)(\mathsf{N}+\mathsf{K}h'-\mathsf{K})}\right)$. More precisely, by memory-sharing between $(\mathsf{M}_1, \mathsf{R}_1)$ and $(\mathsf{M}_2, \mathsf{R}_2)$ with coefficient

$$\alpha = \frac{\mathsf{M}_2 - \mathsf{M}_3}{\mathsf{M}_2 - \mathsf{M}} = \frac{\mathsf{N}(4\mathsf{K}-h'\mathsf{K}-8)+4\mathsf{K}(h'-1)(\mathsf{K}-2)}{4(\mathsf{K}-2)(\mathsf{N}+h'\mathsf{K}-\mathsf{K})} \tag{156}$$

such that $\alpha\mathsf{M}_1 + (1-\alpha)\mathsf{M}_2 = \mathsf{M}_3$, we get at M_3 Scheme A can achieve,

$$\begin{aligned}\mathsf{R}' &= \alpha\mathsf{R}_1 + (1-\alpha)\mathsf{R}_2 \\ &= \mathsf{N}\frac{-12\mathsf{N}+8\mathsf{K}^2(h'-1)+\mathsf{K}\big(\mathsf{N}(8-h')-14h'+12\big)}{4(2\mathsf{K}-3)(\mathsf{N}+h'\mathsf{K}-\mathsf{K})}.\end{aligned} \tag{157}$$

In the following, we compare R' and R_3 to obtain

$$\begin{aligned}\frac{\mathsf{R}'}{\mathsf{R}_3} = &\frac{13\mathsf{N}(h'+1)}{6(2\mathsf{K}-3)\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big)} \\ &\big(-12\mathsf{N}+8\mathsf{K}^2(h'-1)+\mathsf{K}\big(\mathsf{N}(8-h')-14h'+12\big)\big).\end{aligned} \tag{158}$$

Finally, we will prove

$$\begin{aligned}\frac{6\mathsf{R}'}{13\mathsf{R}_3} = &\frac{\mathsf{N}(h'+1)}{(2\mathsf{K}-3)\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big)} \\ &\big(-12\mathsf{N}+8\mathsf{K}^2(h'-1)+\mathsf{K}\big(\mathsf{N}(8-h')-14h'+12\big)\big) \\ &< 8.\end{aligned} \tag{159}$$

We can compute that

$$\begin{aligned}&8(2\mathsf{K}-3)\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big)-\mathsf{N}(h'+1) \\ &\big(-12\mathsf{N}+8\mathsf{K}^2(h'-1)+\mathsf{K}\big(\mathsf{N}(8-h')-14h'+12\big)\big) \\ &\ge 8(2\mathsf{K}-3)\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big) \\ &-\mathsf{N}(h'+1)\left(-12\mathsf{N}+8\mathsf{K}^2(h'-1)+\mathsf{K}\mathsf{N}(8-h')\right) &\text{(160a)} \\ &= \big(32(2\mathsf{K}-3)+12(h'+1)-\mathsf{K}(8-h')(h'+1)\big)\mathsf{N}^2 \\ &-\big(8\mathsf{K}(h'+1)(h'-1)-16(2\mathsf{K}-3)(h'-2)\big)\mathsf{K}\mathsf{N} \\ &+8(2\mathsf{K}-3)\mathsf{K}^2h'(h'-1) &\text{(160b)} \\ &\ge \big(32(2\mathsf{K}-3)+12(h'+1)-\mathsf{K}(8-h')(h'+1)\big)\mathsf{N}^2 \\ &-\big(8(h'+1)(h'-1)-16(h'-2)\big)\mathsf{K}^2\mathsf{N} \\ &+8(2\mathsf{K}-3)\mathsf{K}^2h'(h'-1), &\text{(160c)}\end{aligned}$$

where (160a) comes from $h' \ge 1$ and (160b) comes from $\mathsf{K} \ge 3$.

Recall that $\mathsf{N} > 6\mathsf{K}$, and that $h' \le h_1 = \left\lfloor \frac{4(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})}{\mathsf{K}(\mathsf{N}-4\mathsf{K}+8)} \right\rfloor < 10$.

We first focus on $h'=9$. If $h'=9$, it can be seen that $6\mathsf{K} < \mathsf{N} < \frac{32}{5}\mathsf{K}$. Hence, we have

$$\begin{aligned}&8(2\mathsf{K}-3)\mathsf{K}^2h'(h'-1) > \frac{5}{4}(2\mathsf{K}-3)\mathsf{K}\mathsf{N}h'(h'-1) \\ &\ge \frac{5}{4}\mathsf{K}^2\mathsf{N}h'(h'-1) = 90\mathsf{K}^2\mathsf{N}.\end{aligned} \tag{161}$$

We take $h'=9$ and (161) into (160c) to obtain

$$\begin{aligned}&8(2\mathsf{K}-3)\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big)-\mathsf{N}(h'+1) \\ &\big(-12\mathsf{N}+8\mathsf{K}^2(h'-1)+\mathsf{K}\big(\mathsf{N}(8-h')-14h'+12\big)\big) \\ &> (74\mathsf{K}+24)\mathsf{N}^2-(640-112-90)\mathsf{K}^2\mathsf{N} &\text{(162a)} \\ &> 74\mathsf{K}\mathsf{N}^2-438\mathsf{K}^2\mathsf{N} &\text{(162b)} \\ &> 0, &\text{(162c)}\end{aligned}$$

where (162c) comes from $\mathsf{N} > 6\mathsf{K}$.

We then focus on $h'=8$. If $\mathsf{K}=3$, from (160c), we have the RHS of (160c) becomes $204\mathsf{N}(\mathsf{N}-18)+12096$, which is larger than 0 since $\mathsf{N} > 6\mathsf{K} \ge 18$. Now we consider $\mathsf{K} \ge 4$. From (160b), we have

$$\begin{aligned}&8(2\mathsf{K}-3)\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big)-\mathsf{N}(h'+1) \\ &\big(-12\mathsf{N}+8\mathsf{K}^2(h'-1)+\mathsf{K}\big(\mathsf{N}(8-h')-14h'+12\big)\big) \\ &> \big(32(2\mathsf{K}-3)+12(h'+1)-\mathsf{K}(8-h')(h'+1)\big)\mathsf{N}^2 \\ &-\big(8\mathsf{K}(h'+1)(h'-1)-16(2\mathsf{K}-3)(h'-2)\big)\mathsf{K}\mathsf{N} &\text{(163a)} \\ &\ge \big(32(2\mathsf{K}-3)+12(h'+1)-\mathsf{K}(8-h')(h'+1)\big)\mathsf{N}^2 \\ &-\big(8\mathsf{K}(h'+1)(h'-1)-20\mathsf{K}(h'-2)\big)\mathsf{K}\mathsf{N} &\text{(163b)} \\ &= \big((56+h'^2-7h')\mathsf{K}+12h'-84\big)\mathsf{N}^2 \\ &-(32+8h'^2-20h')\mathsf{K}^2\mathsf{N} &\text{(163c)} \\ &\ge (56+h'^2-7h')\mathsf{K}\mathsf{N}^2-(32+8h'^2-20h')\mathsf{K}^2\mathsf{N} &\text{(163d)} \\ &> 6(56+h'^2-7h')\mathsf{K}^2\mathsf{N}-(32+8h'^2-20h')\mathsf{K}^2\mathsf{N} &\text{(163e)} \\ &= 0, &\text{(163f)}\end{aligned}$$

where (163b) comes from $\mathsf{K} \ge 4$ and thus $\frac{2\mathsf{K}-3}{\mathsf{K}} \ge \frac{5}{4}$, and (163e) comes from $\mathsf{N} > 6\mathsf{K}$.

Lastly, we consider $h' \in [7]$. From (160c), we have

$$\begin{aligned}&8(2\mathsf{K}-3)\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big)-\mathsf{N}(h'+1) \\ &\big(-12\mathsf{N}+8\mathsf{K}^2(h'-1)+\mathsf{K}\big(\mathsf{N}(8-h')-14h'+12\big)\big) \\ &> \big(32(2\mathsf{K}-3)+12(h'+1)-\mathsf{K}(8-h')(h'+1)\big)\mathsf{N}^2 \\ &-\big(8(h'+1)(h'-1)-16(h'-2)\big)\mathsf{K}^2\mathsf{N} &\text{(164a)} \\ &= \big((56+h'^2-7h')\mathsf{K}+12h'-84\big)\mathsf{N}^2 \\ &-(24+8h'^2-16h')\mathsf{K}^2\mathsf{N} &\text{(164b)} \\ &\ge (56+h'^2-7h'+4h'-28)\mathsf{K}\mathsf{N}^2 \\ &-(24+8h'^2-16h')\mathsf{K}^2\mathsf{N} &\text{(164c)} \\ &> 6(28+h'^2-3h')\mathsf{K}^2\mathsf{N}-(24+8h'^2-16h')\mathsf{K}^2\mathsf{N} &\text{(164d)} \\ &= (144-2h'^2-2h')\mathsf{K}^2\mathsf{N} &\text{(164e)} \\ &> 0 &\text{(164f)}\end{aligned}$$

where (164c) comes from $h' \le 7$ and $\mathsf{K} \ge 3$, which lead to $12h' - 84 \ge (4h' - 28)\mathsf{K}$, and (164d) comes from $\mathsf{N} > 6\mathsf{K}$, and (164f) comes from $h' \in [7]$.

In conclusion, we prove (159). In other words, under the constraint of uncoded cache placement and user collusion, Scheme A is order optimal to within a factor of $\frac{13}{6} \times 8 < 18$ for the memory size Regime 1.

B. Regime 2

Similar to the converse bound for Regime 1, it can be computed that the converse bound in (152) for $\frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_1}{2(\mathsf{N}+\mathsf{K}h_1-\mathsf{K})} \le \mathsf{M} \le \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})}$ is a piecewise curve with the corner points

$$\left(\frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h'}{2(\mathsf{N}+\mathsf{K}h'-\mathsf{K})}, \right. \\ \left. \frac{6}{13}\frac{\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h') + 2\mathsf{N}(2\mathsf{N}-\mathsf{K})}{4(h'+1)(\mathsf{N}+\mathsf{K}h'-\mathsf{K})}\right), \ \forall h' \in [h_1 : h_2]. \tag{165}$$

For the achievability, we take the memory-sharing among the achieved points in (129), $\left(\frac{\mathsf{N}+t-1}{\mathsf{K}}, \frac{\mathsf{U}-t+1}{t}\right)$, where $t \in [\mathsf{U}+1]$. We want to interpolate the achieved points of Scheme A to match the converse bound at the memory size $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h'}{2(\mathsf{N}+\mathsf{K}h'-\mathsf{K})}$ where $h' \in [h_1 : h_2]$. By computing

$$\frac{\mathsf{N}+t-1}{\mathsf{K}} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h'}{2(\mathsf{N}+\mathsf{K}h'-\mathsf{K})} \\ \iff t = \frac{\mathsf{N}h'\mathsf{K}}{2(\mathsf{N}+\mathsf{K}h'-\mathsf{K})} + 1, \tag{166}$$

and observing $\frac{\mathsf{U}-t+1}{t}$ is non-increasing with t, it can be seen that the achieved load of Scheme A at $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h'}{2(\mathsf{N}+\mathsf{K}h'-\mathsf{K})}$ is lower than

$$\mathsf{R}' = \frac{\mathsf{U} - \frac{\mathsf{N}h'\mathsf{K}}{2(\mathsf{N}+\mathsf{K}h'-\mathsf{K})} + 1}{\frac{\mathsf{N}h'\mathsf{K}}{2(\mathsf{N}+\mathsf{K}h'-\mathsf{K})}}. \tag{167}$$

By comparing R' and $\frac{6}{13}\frac{\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})}{4(h'+1)(\mathsf{N}+\mathsf{K}h'-\mathsf{K})}$, we have (168) (at the top of the next page).

Since $\mathsf{K} \ge 3$, we have

$$h' \ge h_1 = \left\lfloor \frac{4(\mathsf{K}-2)(\mathsf{N}-\mathsf{K})}{\mathsf{K}(\mathsf{N}-4\mathsf{K}+8)} \right\rfloor \ge \left\lfloor \frac{2(\mathsf{N}-\mathsf{K})}{\mathsf{N}-4\mathsf{K}+8} \right\rfloor > 2; \tag{169a}$$

$$h' \le h_2 = \left\lfloor \frac{2\mathsf{N}}{\mathsf{K}} - 2 \right\rfloor < \frac{2\mathsf{N}}{\mathsf{K}}. \tag{169b}$$

In the following, we will use (169) and $\mathsf{N} > 6\mathsf{K} \ge 18$ to prove (170) (at the top of the next page).

We can compute that

$$\begin{aligned} &8\mathsf{K}\mathsf{N}h'\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h') + 2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big) \\ &- 4(\mathsf{N}+\mathsf{K}h'-\mathsf{K})(h'+1)\big(2\mathsf{K}^2\mathsf{N}(h'-1) \\ &+ \mathsf{K}(2\mathsf{N}^2+2\mathsf{N}+2h'-3\mathsf{N}h'-2) - 2\mathsf{N}(\mathsf{N}-1)\big) \\ &\ge 8\mathsf{K}\mathsf{N}h'\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h') + 2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big) \\ &- 4(\mathsf{N}+\mathsf{K}h'-\mathsf{K})(h'+1)\big(2\mathsf{K}^2\mathsf{N}(h'-1) \\ &+ \mathsf{K}(2\mathsf{N}^2+2\mathsf{N}+2h'-3\mathsf{N}h'-2)\big) \end{aligned} \tag{171a}$$

$$\begin{aligned} &= 8\mathsf{K}(\mathsf{N}-\mathsf{K}) + 8\mathsf{K}^3\mathsf{N}(h'-1) + 4\mathsf{K}\mathsf{N}^2(h'-2) \\ &+ 8\mathsf{K}\mathsf{N}(3\mathsf{N}^2h' - 4\mathsf{K}\mathsf{N}h' - \mathsf{N}^2) \\ &+ 4\mathsf{K}h'(3\mathsf{K}\mathsf{N}h'^2 - 3\mathsf{K}\mathsf{N} - 2\mathsf{K}h'^2) + 4\mathsf{K}\mathsf{N}h'^2(3\mathsf{N}-2\mathsf{K}-2) \\ &+ 8\mathsf{K}^2\mathsf{N} + 16\mathsf{K}^2\mathsf{N}^2 + 8\mathsf{K}^2h' + 8\mathsf{K}^2h'^2 \end{aligned} \tag{171b}$$

$$\begin{aligned} &> 8\mathsf{K}(\mathsf{N}-\mathsf{K}) + 8\mathsf{K}^3\mathsf{N}(h'-1) + 4\mathsf{K}\mathsf{N}^2(h'-2) \\ &+ 8\mathsf{K}\mathsf{N}(3\mathsf{N}^2h' - 4\mathsf{K}\mathsf{N}h' - \mathsf{N}^2) \\ &+ 4\mathsf{K}h'(3\mathsf{K}\mathsf{N}h'^2 - 3\mathsf{K}\mathsf{N} - 2\mathsf{K}h'^2) + 4\mathsf{K}\mathsf{N}h'^2(3\mathsf{N}-2\mathsf{K}-2) \end{aligned} \tag{171c}$$

$$\begin{aligned} &> 8\mathsf{K}\mathsf{N}(3\mathsf{N}^2h' - 4\mathsf{K}\mathsf{N}h' - \mathsf{N}^2) \\ &+ 4\mathsf{K}h'(3\mathsf{K}\mathsf{N}h'^2 - 3\mathsf{K}\mathsf{N} - 2\mathsf{K}h'^2) \end{aligned} \tag{171d}$$

$$\begin{aligned} &= 8\mathsf{K}\mathsf{N}(\mathsf{N}^2h' - \mathsf{N}^2) + 8\mathsf{K}\mathsf{N}(2\mathsf{N}^2h' - 4\mathsf{K}\mathsf{N}h') \\ &+ 4\mathsf{K}h'(\mathsf{K}\mathsf{N}h'^2 - 3\mathsf{K}\mathsf{N}) + 4\mathsf{K}h'(2\mathsf{K}\mathsf{N}h'^2 - 2\mathsf{K}h'^2) \end{aligned} \tag{171e}$$

$$> 0, \tag{171f}$$

where (171d) and (171f) come from $\mathsf{N} > 6\mathsf{K}$ and $h' > 2$.

In conclusion, we prove (170). In other words, under the constraint of uncoded cache placement and user collusion, Scheme A is order optimal to within a factor of $\frac{13}{6} \times 8 < 18$ for the memory size Regime 2.

C. Regime 3

When $\frac{\mathsf{N}}{\mathsf{K}} \le \mathsf{M} \le \frac{2\mathsf{N}}{\mathsf{K}}$, the converse bound in [11] is a straight line between $\left(\frac{\mathsf{N}}{\mathsf{K}}, \frac{\mathsf{K}-1}{2}\right)$ and $\left(\frac{2\mathsf{N}}{\mathsf{K}}, \frac{\mathsf{K}-2}{3}\right)$, which is denoted by $\mathsf{R}_{[11]}(\mathsf{M})$. Hence, the converse bound in [11] for Regime 3 where $\frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})} \le \mathsf{M} \le \frac{2\mathsf{N}}{\mathsf{K}}$ is a straight line. When $\mathsf{M} = \frac{2\mathsf{N}}{\mathsf{K}}$, we proved in Appendix D-A that the multiplicative gap between Scheme A and the converse bound in [11] is no more than 6. Hence, in the rest of this proof, we focus on the memory size $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})} \le \mathsf{M} \le \frac{2\mathsf{N}}{\mathsf{K}}$.

Recall that $h_2 := \left\lfloor \frac{2\mathsf{N}}{\mathsf{K}} - 2 \right\rfloor \le \frac{2\mathsf{N}}{\mathsf{K}} - 2$, we note that

$$\frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})} \le \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}\left(\frac{2\mathsf{N}}{\mathsf{K}}-2\right)}{2\{\mathsf{N}+\mathsf{K}\left(\frac{2\mathsf{N}}{\mathsf{K}}-2\right)-\mathsf{K}\}} \\ = \frac{4\mathsf{N}}{3\mathsf{K}}. \tag{172}$$

Hence, the load of the converse bound in [11] at $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})}$ is strictly higher than the one at $\mathsf{M}' = \frac{4\mathsf{N}}{3\mathsf{K}}$. By computing the converse bound in [11] at $\mathsf{M}' = \frac{4\mathsf{N}}{3\mathsf{K}}$ is

$$\mathsf{R}_{[11]}(\mathsf{M}') = \frac{2}{3}\frac{\mathsf{K}-1}{2} + \frac{1}{3}\frac{\mathsf{K}-2}{3} = \frac{4\mathsf{K}-5}{9}, \tag{173}$$

at $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})}$, we have

$$\mathsf{R}^\star_{\mathrm{u,c}} \ge \mathsf{R}_{[11]}(\mathsf{M}) > \mathsf{R}_{[11]}(\mathsf{M}') = \frac{4\mathsf{K}-5}{9}. \tag{174}$$

For the achievability, it was proved in (167) that the achieved load of Scheme A at $\mathsf{M} = \frac{\mathsf{N}}{\mathsf{K}} + \frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})}$ is lower than

$$\mathsf{R}' = \frac{\mathsf{U} - \frac{\mathsf{N}h_2\mathsf{K}}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})} + 1}{\frac{\mathsf{N}h_2\mathsf{K}}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})}} \tag{175a}$$

$$\le \frac{\mathsf{U} - \frac{\mathsf{N}(2\mathsf{N}/\mathsf{K}-3)\mathsf{K}}{2\left(\mathsf{N}+\mathsf{K}(2\mathsf{N}/\mathsf{K}-3)-\mathsf{K}\right)} + 1}{\frac{\mathsf{N}(2\mathsf{N}/\mathsf{K}-3)\mathsf{K}}{2\left(\mathsf{N}+\mathsf{K}(2\mathsf{N}/\mathsf{K}-3)-\mathsf{K}\right)}} \tag{175b}$$

$$\frac{\mathsf{R}'}{\mathsf{R}_3}=\frac{13}{6}\frac{4(\mathsf{N}+\mathsf{K}h'-\mathsf{K})(h'+1)\big(2\mathsf{K}^2\mathsf{N}(h'-1)+\mathsf{K}(2\mathsf{N}^2+2\mathsf{N}+2h'-3\mathsf{N}h'-2)-2\mathsf{N}(\mathsf{N}-1)\big)}{\mathsf{K}\mathsf{N}h'\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big)}. \tag{168}$$

$$\frac{6\mathsf{R}'}{13\mathsf{R}_3}=\frac{4(\mathsf{N}+\mathsf{K}h'-\mathsf{K})(h'+1)\big(2\mathsf{K}^2\mathsf{N}(h'-1)+\mathsf{K}(2\mathsf{N}^2+2\mathsf{N}+2h'-3\mathsf{N}h'-2)-2\mathsf{N}(\mathsf{N}-1)\big)}{\mathsf{K}\mathsf{N}h'\big(\mathsf{K}(h'-1)(2\mathsf{N}+\mathsf{K}h')+2\mathsf{N}(2\mathsf{N}-\mathsf{K})\big)}<8. \tag{170}$$

$$=\frac{(6\mathsf{K}-8)\mathsf{N}^2-(8\mathsf{K}-11)\mathsf{K}\mathsf{N}+6\mathsf{N}-8\mathsf{K}}{2\mathsf{N}^2-3\mathsf{K}\mathsf{N}}, \tag{175c}$$

where (175b) comes that $\frac{\mathsf{U}-t+1}{t}$ is non-increasing with t, and that $h_2\le 2\mathsf{N}/\mathsf{K}-3$.

Finally, we compare R' and $\frac{4\mathsf{K}-5}{9}$ to obtain,

$$\frac{\mathsf{R}'}{\frac{4\mathsf{K}-5}{9}}=9\frac{(6\mathsf{K}-8)\mathsf{N}^2-(8\mathsf{K}-11)\mathsf{K}\mathsf{N}+6\mathsf{N}-8\mathsf{K}}{(2\mathsf{N}^2-3\mathsf{K}\mathsf{N})(4\mathsf{K}-5)}. \tag{176}$$

In addition, we compute

$$\begin{aligned}
&2(2\mathsf{N}^2-3\mathsf{K}\mathsf{N})(4\mathsf{K}-5)\\
&-\big((6\mathsf{K}-8)\mathsf{N}^2-(8\mathsf{K}-11)\mathsf{K}\mathsf{N}+6\mathsf{N}-8\mathsf{K}\big)\\
&=2\mathsf{N}(5\mathsf{K}\mathsf{N}-6\mathsf{N}-8\mathsf{K}^2)+(19\mathsf{K}\mathsf{N}-6\mathsf{N})+8\mathsf{K} &\text{(177a)}\\
&>2\mathsf{N}(5\mathsf{K}\mathsf{N}-6\mathsf{N}-8\mathsf{K}^2) &\text{(177b)}\\
&\ge 2\mathsf{N}(3\mathsf{K}\mathsf{N}-8\mathsf{K}^2) &\text{(177c)}\\
&>0, &\text{(177d)}
\end{aligned}$$

where (177b) and (177c) come from $\mathsf{K}\ge 3$, and (177d) comes from $\mathsf{N}>6\mathsf{K}$. By taking (177d) into (176), it can be seen that the multiplicative gap between Scheme A and the converse bound in [11] at $\mathsf{M}=\frac{\mathsf{N}}{\mathsf{K}}+\frac{\mathsf{N}h_2}{2(\mathsf{N}+\mathsf{K}h_2-\mathsf{K})}$ is less than 18.

In conclusion, we prove that under the constraint of uncoded cache placement and user collusion, Scheme A is order optimal to within a factor of 18 for the memory size Regime 3.

REFERENCES

[1] K. Wan, H. Sun, M. Ji, D. Tuninetti, and G. Caire, "Device-to-Device private caching with trusted server," *in IEEE Intern. Conf. Commun. (ICC)*, Jun. 2020.

[2] ——, "Novel converse for Device-to-Device demand-private caching with a trusted server," *in IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1705–1710, Jun. 2020.

[3] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5g wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[4] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Infor. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[5] Q. Yu, M. A. Maddah-Ali, and S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," *IEEE Trans. Infor. Theory*, vol. 64, no. 2, pp. 1281–1296, Feb. 2018.

[6] K. Wan and G. Caire, "On coded caching with private demands," *IEEE Trans. Inf. Theory*, vol. 67, no. 1, pp. 358–372, Jan. 2021.

[7] S. Kamath, "Demand private coded caching," *arXiv:1909.03324*, Sep. 2019.

[8] C. Yapar, K. Wan, R. F. Schaefer, and G. Caire, "On the optimality of D2D coded caching with uncoded cache placement and one-shot delivery," *IEEE Trans. Communications*, vol. 67, no. 12, pp. 8179–8192, Dec. 2019.

[9] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "Characterizing the rate-memory tradeoff in cache networks within a factor of 2," *IEEE Trans. Infor. Theory*, vol. 65, no. 1, pp. 647–663, Jan. 2019.

[10] K. Wan, D. Tuninetti, and P. Piantanida, "On the optimality of uncoded cache placement," *in IEEE Infor. Theory Workshop (ITW)*, Sep. 2016.

[11] ——, "An index coding approach to caching with uncoded cache placement," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1318–1332, Mar. 2020.

[12] F. Arbabjolfaei, B. Bandemer, Y.-H. Kim, E. Sasoglu, and L. Wang, "On the capacity region for index coding," *in IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013.

[13] C. Tian, "Symmetry, outer bounds, and code constructions: A computer-aided investigation on the fundamental limits of caching," *Entropy 2018, 20, 603*.

[14] F. Engelmann and P. Elia, "A content-delivery protocol, exploiting the privacy benefits of coded caching," *2017 15th Intern. Symp. on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2017.

[15] V. Aravind, P. K. Sarvepalli, and A. Thangaraj, "Coded caching with demand privacy: Constructions for lower subpacketization and generalizations," *arXiv:2007.07475*, Jul. 2020.

[16] Q. Yan and D. Tuninetti, "Fundamental limits of caching for demand privacy against colluding users," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 192–207, Mar. 2021.

[17] K. K. K. Namboodiri and B. S. Rajan, "Optimal demand private coded caching for users with small buffers," *in IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 706–711, Jul. 2021.

[18] A. Gholami, K. Wan, H. Sun, M. Ji, and G. Caire, "Coded caching with private demands and caches," *arXiv:2201.11539*, Jan. 2022.

[19] S. Kamath, J. Ravi, and B. K. Dey, "Demand-private coded caching and the exact trade-off for N=K=2," *National Conference on Communications (NCC)*, Feb. 2020.

[20] M. Ji, G. Caire, and A. Molisch, "Fundamental limits of caching in wireless D2D networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 849–869, 2016.

[21] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," *in Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, pp. 41–50, 1995.

[22] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4075–4088, 2017.

[23] ——, "The capacity of robust private information retrieval with colluding databases," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2361–2370, 2018.

[24] Y. Zhang and G. Ge, "Private information retrieval from MDS coded databases with colluding servers under several variant models," *available at arXiv:1705.03186*, May. 2017.

[25] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, and C. Hollant, "Private information retrieval from coded storage systems with colluding, byzantine, and unresponsive servers," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3898–3906, 2019.

[26] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Infor. Theory*, vol. 23, no. 1, pp. 60–64, 1977.

[27] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, UK: Cambridge University Press, 2011.

[28] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Infor. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[29] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Trans. Infor. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[30] A. Porter and M. Wootters, "Embedded index coding," *IEEE Trans. Inf. Theory*, vol. 67, no. 3, pp. 1461–1477, Mar. 2021.

[31] T. Liu and D. Tuninetti, "Private pliable index coding," *in Proc. IEEE Inf. Theory Workshop (ITW)*, Aug. 2019.

[32] A. Beimel, Y. Ishai, and E. Kushilevitz, "General constructions for information-theoretic private information retrieval," *Journal of Computer and System Sciences*, vol. 71, no. 2, pp. 213–247, 2005.

[33] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, C. Hollanti, and S. El Rouayheb, "Private information retrieval schemes for coded data with arbitrary collusion patterns," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 1908–1912.
[34] K. Banawan and S. Ulukus, "The capacity of private information retrieval from byzantine and colluding databases," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 1206–1219, 2018.
[35] K. Wan, H. Sun, M. Ji, D. Tuninetti, and G. Caire, "On optimal load-memory tradeoff of cache-aided scalar linear function retrieval," *IEEE Trans. Infor. Theory*, vol. 67, no. 6, pp. 4001–4018, Jun. 2021.
[36] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," *IEEE Trans. Infor. Theory*, vol. 63, no. 7, pp. 4388–4413, May 2017.

Kai Wan (S '15 – M '18) received the B.E. degree in Optoelectronics from Huazhong University of Science and Technology, China, in 2012, the M.Sc. and Ph.D. degrees in Communications from Université Paris-Saclay, France, in 2014 and 2018. He is currently a post-doctoral researcher with the Communications and Information Theory Chair (CommIT) at Technische Universität Berlin, Berlin, Germany. His research interests include information theory, coding techniques, and their applications on coded caching, index coding, distributed storage, distributed computing, wireless communications, privacy and security. He has served as an Associate Editor of IEEE Communications Letters from Aug. 2021.

Hua Sun (S '12 – M '17) received the B.E. degree in Communications Engineering from Beijing University of Posts and Telecommunications, China, in 2011, and the M.S. degree in Electrical and Computer Engineering and the Ph.D. degree in Electrical Engineering from University of California Irvine, USA, in 2013 and 2017, respectively. He is an Assistant Professor in the Department of Electrical Engineering at the University of North Texas, USA. His research interests include information theory and its applications to communications, privacy, security, and storage.

Dr. Sun is a recipient of the NSF CAREER award in 2021, and the UNT College of Engineering Distinguished Faculty Fellowship in 2021. His co-authored papers received the IEEE Jack Keil Wolf ISIT Student Paper Award in 2016, and an IEEE GLOBECOM Best Paper Award in 2016.

Mingyue Ji (S '09 – M '15) received the B.E. in Communication Engineering from Beijing University of Posts and Telecommunications (China), in 2006, the M.Sc. degrees in Electrical Engineering from Royal Institute of Technology (Sweden) and from University of California, Santa Cruz, in 2008 and 2010, respectively, and the PhD from the Ming Hsieh Department of Electrical Engineering at University of Southern California in 2015. He subsequently was a Staff II System Design Scientist with Broadcom Corporation (Broadcom Limited) in 2015-2016. He is now an Assistant Professor of Electrical and Computer Engineering Department and an Adjunct Assistant Professor of School of Computing at the University of Utah. He received the NSF CAREER Award in 2022, the IEEE Communications Society Leonard G. Abraham Prize for the best IEEE JSAC paper in 2019, the best paper awards at 2021 IEEE Globecom conference and at 2015 IEEE ICC conference, the best student paper award at 2010 IEEE European Wireless conference and USC Annenberg Fellowship from 2010 to 2014. He has served as an Associate Editor of IEEE Transactions on Communications from 2020. He is interested the broad area of information theory, coding theory, concentration of measure and statistics with the applications of caching networks, wireless communications, distributed storage and computing systems, distributed machine learning, and (statistical) signal processing.

Daniela Tuninetti (M '98 – SM '13 – F '21) is currently a Professor within the Department of Electrical and Computer Engineering at the University of Illinois at Chicago (UIC), which she joined in 2005. Dr. Tuninetti got her Ph.D. in Electrical Engineering in 2002 from ENST/Télécom ParisTech (Paris, France, with work done at the Eurecom Institute in Sophia Antipolis, France), and she was a postdoctoral research associate at the School of Communication and Computer Science at the Swiss Federal Institute of Technology in Lausanne (EPFL, Lausanne, Switzerland) from 2002 to 2004. Dr. Tuninetti is a recipient of a best paper award at the European Wireless Conference in 2002, of an NSF CAREER award in 2007, and named University of Illinois Scholar in 2015. Dr. Tuninetti was the editor-in-chief of the IEEE Information Theory Society Newsletter from 2006 to 2008, an editor for IEEE COMMUNICATION LETTERS from 2006 to 2009, for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2011 to 2014; and for IEEE TRANSACTIONS ON INFORMATION THEORY from 2014 to 2017. She is currently a distinguished lecturer for the Information Theory society. She is also currently an editor for IEEE Transactions on Communications. Dr. Tuninetti's research interests are in the ultimate performance limits of wireless interference networks (with special emphasis on cognition and user cooperation), coexistence between radar and communication systems, multi-relay networks, content-type coding, cache-aided systems and distributed private coded computing.

Giuseppe Caire (S '92 – M '94 – SM '03 – F '05) was born in Torino in 1965. He received the B.Sc. in Electrical Engineering from Politecnico di Torino in 1990, the M.Sc. in Electrical Engineering from Princeton University in 1992, and the Ph.D. from Politecnico di Torino in 1994. He has been a post-doctoral research fellow with the European Space Agency (ESTEC, Noordwijk, The Netherlands) in 1994-1995, Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, a Professor of Electrical Engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles, and he is currently an Alexander von Humboldt Professor with the Faculty of Electrical Engineering and Computer Science at the Technical University of Berlin, Germany.

He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, the IEEE Communications Society and Information Theory Society Joint Paper Award in 2004 and in 2011, the Okawa Research Award in 2006, the Alexander von Humboldt Professorship in 2014, the Vodafone Innovation Prize in 2015, an ERC Advanced Grant in 2018, the Leonard G. Abraham Prize for best IEEE JSAC paper in 2019, the IEEE Communications Society Edwin Howard Armstrong Achievement Award in 2020, and he is a recipient of the 2021 Leibinz Prize of the German National Science Foundation (DFG). Giuseppe Caire is a Fellow of IEEE since 2005. He has served in the Board of Governors of the IEEE Information Theory Society from 2004 to 2007, and as officer from 2008 to 2013. He was President of the IEEE Information Theory Society in 2011. His main research interests are in the field of communications theory, information theory, channel and source coding with particular focus on wireless communications.