New Results on Optimal Schemes for Coded Caching

Vijith Kumar K P

New Results on Optimal Schemes for Coded Caching

A

Thesis Submitted

in Partial Fulfilment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

By

Vijith Kumar K P

DEPARTMENT OF ELECTRONICS AND ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781 039, ASSAM, INDIA

May, 2021

Declaration

I hereby declare that the thesis entitled "**New Results on Optimal Schemes for Coded Caching**", submitted in the *Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Assam, India*, for the award of the degree of **Doctor of Philosophy**, has been carried out by me under the supervision and guidance of Dr. Tony Jacob. The results embodied in this thesis are original and have not been submitted to any other University or Institute for the award of any degree or diploma.

Dated:

Place: Guwahati

Vijith Kumar K P

Research Scholar

Dept. of Electronics and Electrical Engineering

Indian Institute of Technology Guwahati

Guwahati - 781039, Assam, India.

Certificate

This is to certify that the thesis entitled "**New Results on Optimal Schemes for Coded Caching**", submitted by **Vijith Kumar K P** (146102006), a research scholar in the *Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati*, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by him under our supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and in my opinion has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dated:

Guwahati.

Dr. Tony Jacob

Associate Professor

Dept. of Electronics and Electrical Engg.

Indian Institute of Technology Guwahati

Guwahati - 781 039, Assam, India.

I would like to dedicate this thesis to my loving parents ...

Acknowledgements

First and foremost, I wish to express my deepest gratitude to my thesis supervisor Dr. Tony Jacob for their excellent guidance throughout my Ph.D. tenure. I also wish to express my appreciation to Dr. Brijesh Kumar Rai for his valuable advice. Their work ethic and attention to detail have been a source of great inspiration to me. My heartfelt thanks to them for their support and guidance during my difficult periods. They were always available during my moments of distress and helped me find solutions to it. I can not thank them enough for the patience shown by them while listening and clarifying my silliest of doubts. They have made an everlasting impression on my way of looking at problems, both professional and personal.

I would like to thank my doctoral committee members: Prof. Ratnajit Bhattacharjee, Dr. A Rajesh, and Dr. Benny George K, for sparing time out of their busy schedule to evaluate my progress and enrich this work with their valuable suggestions and feedback. I sincerely thank Dr. Ganesh Natarajan, Dr. Vibin Ramakrishnan, and Dr. Suresh A Kartha for their friendship, suggestions, and motivations.

During my life in IIT Guwahati, I made many friends with whom I shared several precious memories that will stay with me forever. The comradery shared between me, Mathew, D. P. Nair, Achayan, Chitra, Ashmil, Jiss Joseph, Christy, Hrishi, Vasudevan, Oppol, Riya, Jith, Kunju, Thorappan, Ashwani, Amru, Vivek, Alex, Kaushik, and Sishir is unbelievable, and I thank them all for their friendship. The encouragement and guidance received from seniors Prateek, Pawan, and Anoop were immensely helpful in writing this thesis.

Lastly, I extend my sincere thanks to all the staff members from the EEE office and Academic office for helping me out in all sorts of ways during my stay at IITG.

Vijith Kumar K P

Abstract

The idea of coded caching for content distribution networks was introduced by Maddah-Ali and Niesen, who considered an (N, K) cache network in which a server with N files, each of size F bits, is connected via a shared link with K users, each equipped with an isolated cache of size MF bits. The network operates in two phases. In the first phase of coded caching, called the placement phase, the server fills the users caches with functions of its files, without any knowledge of the files that will be required by each user. In the second phase, called the delivery phase, after knowing the demands of all the users, the server broadcasts a set of packets over the shared link. Each user recovers the file it has requested from the broadcast packets, aided by its cache contents. The problem of coded caching is to determine how to fill each user's cache in the placement phase so that the rate required in the delivery phase is minimized.

Maddah-Ali and Niesen, in their seminal work, considered a collection of demands where each file in the server is requested by at least one user. For this set of demands, they introduced a caching scheme in which each user's cache is filled with uncoded file fragments and use multicast coding in the delivery phase. With the help of cut set arguments, they showed that rate of the proposed scheme is within a multiplicative gap of 12 from the optimal rate. Though, this gap was further reduced by proposing new schemes and deriving tighter lower bounds, the optimal rate memory tradeoff for this situation was not known except when $M \leq \frac{1}{K}$ and $M \geq \frac{N}{K}(K-1)$. For the case $\lceil \frac{K+1}{2} \rceil \leq N \leq K$, we derive new lower bounds and propose a new coded caching scheme which leads to a characterizations of the optimal rate memory tradeoff when $M \leq \frac{1}{K} + \frac{1}{K(N-1)}$ and $M \geq \frac{N}{K}(K-1) - \frac{N-1}{K(K-1)}$. For the case $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$, we derive new lower bounds which leads to a characterization of the optimal rate memory tradeoff for $M \geq \frac{N}{K}(K-2)$.

The problem of coded caching is fundamentally a multi-objective optimization problem as noted by Tian, who initiated a study of the optimal rate memory tradeoff for each

demand type. Yu, Maddah-Ali and Avestimehr introduced a new scheme (referred to as the YMA scheme in this thesis) and in a surprising result demonstrated that it was simultaneously optimal for all demand types among caching schemes where the placement phase is restricted to be uncoded. We investigate the possibility of such a universal scheme in the general case where coding is also permitted in the placement phase. We derive new lower bounds which characterize the constraints among different demand types and use it to prove the non-existence of a universal scheme. Inspired by this result, we initiate a study of Pareto optimal schemes in coded caching and establish the Pareto optimality of the YMA scheme.

Contents

List of Figures

List of Tables

1

Introduction

Contents

1.1 Introduction to Caching

Over the past three decades, we have witnessed an exponential growth in data traffic over the internet which has lead to parts of the network being highly congested during peak hours and resources being abundant during off-peak hours. To alleviate this problem, caching techniques place some portions of files in the server in caches distributed across network during off-peak traffic time. When the server receives requests for the files, it broadcasts a set of packets with which the users, aided by the caches, can obtain their requested files. The first phase, where the server copies a portion of its contents to the caches, is called the placement phase. The latter phase, where each user recovers its requested file using the broadcast packets and cache contents, is called the delivery phase. A fundamental issue in a caching problem is to decide what to place in each user's cache during the placement phase and what to broadcast such that the network experiences the least load during the delivery phase. There have been several proposals for web caching [2–6] primarily based on the characteristics of users' demands, such as file popularity. These techniques mainly concentrate on what to store in the caches and not how to deliver. During the delivery phase, the server transmits parts of the requested file, missing from the caches available to each user. These traditional caching schemes fail to exploit the multicast opportunities available in networks with multiple caches. Maddah-Ali and Niesen, in their seminal work [7], introduced the notion of coded caching to exploit this opportunity and demonstrated that coding indeed helps in reducing peak data traffic over uncoded caching schemes. In this thesis, we undertake a study of coded caching schemes and present new schemes along with new lower bounds.

1.2 Motivation for Coded Caching

Consider a cache network where two users, U_1 and U_2, are connected to a server with two files, A and B (each of size F bits), through a common shared error-free link as shown in Figure 1.1, where each user U_k is equipped with an isolated cache Z_k of size F bits. As the users' future demands are not known, the server splits both the files into two non-overlapping subfiles of size $\frac{1}{2}F$ bits, where A_1, A_2 are the subfiles of the file A and B_1, B_2 are the subfiles of file B. One way to fill the caches is to copy the same subfiles A_1 and B_1 into both the caches.

$$Z_1 = A_1, B_1,$$

$$Z_2 = A_1, B_1.$$

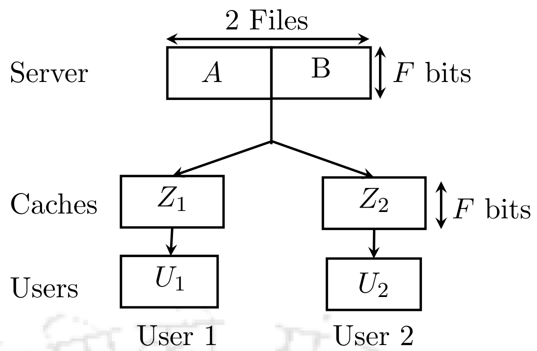

Figure 1.1: The $(2, 2)$ cache network

Consider the demand where user U_1 requests file A and user U_2 requests file B. Since U_1 has the subfile A_1 in its cache Z_1, the server needs to broadcast the subfile A_2 to fulfill its request. Similarly, since U_2 has the subfile B_1 in its cache Z_2, the server needs to broadcast the subfile B_2 to fulfill its request. Thus, the server broadcasts two subfiles each of size $\frac{1}{2}F$ bits separately and the total load experienced by the network corresponding to this demand is F bits. Another way to fill the caches is to copy the subfiles A_1 and B_1 in Z_1 and the subfiles A_2 and B_2 in Z_2.

$$Z_1 = A_1, B_1,$$

$$Z_2 = A_2, B_2.$$

Again, consider the same demand where U_1 requests file A and U_2 requests file B. Since U_1 has subfile A_1 in Z_1, the server needs to broadcast the subfile A_2 to fulfill its request. Similarly, since U_2 has the subfile B_2 in Z_2, the server needs to broadcast the subfile B_1 to fulfill its request. If the two subfiles each of size $\frac{1}{2}F$ bits are broadcast separately, the load experienced by the network corresponding to this demand is F bits. Note that with this cache configuration, the subfile B_1 is available to U_1 from its cache Z_1 and the subfile A_2 is available to U_2 from its cache Z_2. Thus, the server can broadcast a coded packet $A_2 + B_1$ and U_1 can decode A_2 as it has B_1 (from cache Z_1) and U_2 can decode B_1 as it has A_2 (from cache Z_2). As the size of the coded packet is only $\frac{1}{2}F$ bits, the load experienced by the network is $\frac{1}{2}F$ bits. This example demonstrates that coding can indeed help in reducing the load experienced by the network over uncoded caching schemes.

1.3 The Problem of Coded Caching

The problem of coded caching was introduced by Maddah-Ali and Niesen in [7] where they considered the canonical (N, K) cache network where K users $\{U_1, \ldots, U_K\}$ are connected to a server with N files $\{W_1, \ldots, W_N\}$ (each of size F bits) through a common shared error-free link. Each user is equipped with an isolated cache of size MF bits, where $M \in [0, N]$, as shown in Figure 1.2. In the

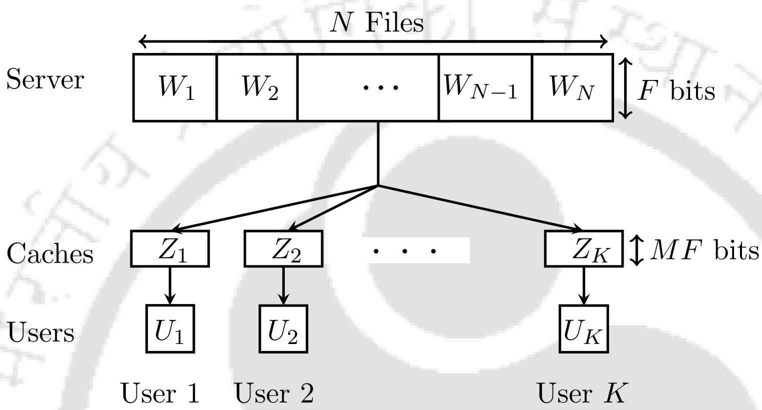

Figure 1.2: The (N, K) cache network

first phase of a coded caching scheme, called the placement phase, the server copies some functions of the files available to it into the caches, without any knowledge of the files that will be required by each user in the future. Let the demands by the users be represented by a vector $\boldsymbol{d} = (W_{d_1}, \ldots, W_{d_K})$, where W_{d_l} is the file requested by user U_l. In the second phase, called the delivery phase, the server broadcasts a set of packets $X_{\boldsymbol{d}}$ of size $R_{\boldsymbol{d}}(M)F$ bits in response to the demand $\boldsymbol{d}$. Each user recovers its required file from the broadcast packets aided by the contents of its isolated cache. The quantity $R_{\boldsymbol{d}}(M)F$ is called the load experienced by the network, and the quantity $R_{\boldsymbol{d}}(M)$ is called the rate. The design of a coded caching scheme involves deciding what to place in the cache attached to each user during the placement phase and what to broadcast for each possible demand such that the shared link experiences the minimum load during the delivery phase. Thus, a caching scheme consists of three sets of functions, namely caching functions, broadcast functions, and recovery functions. The caching functions map the server contents $\{W_1, \ldots, W_N\}$ of size NF bits to each user's cache contents, Z_l, of size MF bits:

$$\alpha_l(W_1, \ldots, W_K) = Z_l \tag{1.1}$$

The broadcast functions map the server contents to broadcast packets $X_{\boldsymbol{d}}$ of size $R_{\boldsymbol{d}}(M)F$ bits:

$$\beta_{\boldsymbol{d}}(W_1, \ldots, W_K) = X_{\boldsymbol{d}} \tag{1.2}$$

The recovery functions extract the file requested by each user, W_{d_l}, from the cache contents and the received packets:

$$\gamma_l(Z_l, X_{\boldsymbol{d}}) = W_{d_l} \tag{1.3}$$

Let $\boldsymbol{D}$ represent the set of all possible demands and let $\boldsymbol{D}_p$ denote the set of all demands that request p distinct files. The corresponding rate is denoted by $R_p(M)$, where

$$R_p(M) = \max\{R_{\boldsymbol{d}}(M) \mid \boldsymbol{d} \in \boldsymbol{D}_p\}. \tag{1.4}$$

For the (N, K) cache network with cache size M, the memory rate pair (M, R_p) is said to be achievable for the demands in $\boldsymbol{D}_p$ if there is a scheme with $R_p(M) \leq R_p$. For a given cache size M, the smallest R_p such that (M, R_p) is achievable is called the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_p$ and it is denoted by

$$R_p^*(M) = \min\{R_p : (M, R_p) \text{ is achievable}\} \tag{1.5}$$

For a scheme that achieves the memory rate pair (M, R_p), we have

$$H(Z_l) \leq M \tag{1.6}$$

$$H(X_{\boldsymbol{d}}) \leq R_p \tag{1.7}$$

$$H(Z_l, X_{\boldsymbol{d}}) = H(W_{d_l}, Z_l, X_{\boldsymbol{d}}), \tag{1.8}$$

$$H(W_1, \ldots, W_N, Z_l, X_{\boldsymbol{d}}) = H(W_1, \ldots, W_N), \tag{1.9}$$

where (1.6) follows from the fact that the size of each cache is M, (1.7) follows from the fact that for any demand in $\boldsymbol{D}_p$ the size of $X_{\boldsymbol{d}}$ is at most $R_p(M) \leq R_p$, (1.8) follows from the fact that the file $W_{\boldsymbol{d}_l}$ can be computed from $X_{\boldsymbol{d}}$ and Z_l by the user U_l, and (1.9) follows from the fact that Z_l and $X_{\boldsymbol{d}}$ are functions of files $\{W_1, \ldots, W_N\}$. From (1.6) and (1.7), we have

$$M + R_p \geq H(Z_l) + H(X_{\boldsymbol{d}_p}) \geq H(Z_l, X_{\boldsymbol{d}_p}) \tag{1.10}$$

Throughout this thesis we use $[L]$ to represent the set $\{1, 2, \ldots, L\}$, $Z_{[L]}$ to represent the set $\{Z_1, \ldots, Z_L\}$

and $W_{[L]}$ to represent the set $\{W_1, W_2, \ldots, W_L\}$.

1.4 Symmetric Schemes in Coded Caching

The canonical (N, K) cache network is inherently symmetric with all files being of the same size F and all caches being of the same size MF. As the caches are filled in the placement phase without knowing the users' demands, it is natural to consider symmetric caching schemes. Tian [8], showed that for any caching scheme, there exists a symmetric caching scheme that gives the same or better performance. Thus, without loss of generality, we consider only symmetric schemes in this thesis and use their properties repeatedly. Consider a demand $\boldsymbol{d}$, where user U_l requires the file W_{d_l},

$$\boldsymbol{d} = (W_{d_1}, \ldots, W_{d_K}) \tag{1.11}$$

Let $\pi(.)$ be a permutation operation defined over the user index set $[K]$ and let $\pi^{-1}(.)$ be its inverse. The demand $\pi\boldsymbol{d}$, where user $U_{\pi(l)}$ requires the file W_{d_l} is

$$\pi\boldsymbol{d} = (W_{d_{\pi^{-1}(1)}}, \ldots, W_{d_{\pi^{-1}(K)}}). \tag{1.12}$$

In response to the demand $\pi\boldsymbol{d}$, the server broadcasts a set of packets $X_{\pi\boldsymbol{d}}$. As shown in [8], for a symmetric caching scheme, we have

$$H(W_{d_k}, Z_{\pi(k)}, X_{\pi\boldsymbol{d}}) = H(W_{d_k}, Z_k, X_{\boldsymbol{d}}) \tag{1.13}$$

Consider another permutation operation $\phi(.)$, defined over the file index set $[N]$ and let $\phi^{-1}(.)$ be its inverse. The demand $\phi\boldsymbol{d}$, where user U_l requires the file $W_{\phi(d_l)}$ is

$$\phi\boldsymbol{d} = (W_{\phi(d_1)}, \ldots, W_{\phi(d_K)}) \tag{1.14}$$

In response to the demand $\phi\boldsymbol{d}$, the server broadcasts a set of packets $X_{\phi\boldsymbol{d}}$. As shown in [8], for a symmetric caching scheme, we have

$$H(W_{\phi(d_k)}, Z_k, X_{\phi\boldsymbol{d}}) = H(W_{d_k}, Z_k, X_{\boldsymbol{d}}) \tag{1.15}$$

1.5 Summary of Previous Results

Maddah-Ali and Niesen, in introducing the problem of coded caching [7], focused on the demand set $\boldsymbol{D}_N$, where all files are requested by some user. They introduced a coded caching scheme with an uncoded placement phase and a coded delivery phase. For $M = \frac{N}{K}t$, this scheme was shown to achieve the rate

$$R_N^t = \frac{K-t}{1+t} \tag{1.16}$$

With the help of cut-set arguments, the authors showed that the rate achieved the scheme is within a multiplicative gap of 12 from the optimal rate. Several ideas to improve this uncoded placement scheme was pursued in [9–13] culminating in the modification proposed by Yu, Maddah-Ali, and Avestimehr (referred to as the YMA scheme in this thesis). For $M = \frac{N}{K}t$, the YMA scheme achieves the rate

$$R_N^t = \frac{\binom{K}{t+1} - \binom{K-N}{t+1}}{\binom{K}{t}} \tag{1.17}$$

for the demand set $\boldsymbol{D}_N$. The rate achieved by the YMA scheme for the demand set $\boldsymbol{D}_p$ is

$$R_p^t = \frac{\binom{K}{t+1} - \binom{K-p}{t+1}}{\binom{K}{t}} \tag{1.18}$$

This rate achieved by the scheme was shown to be within a multiplicative gap of 2 from the optimal rate. Yu et al. [13] also considered the problem of finding optimal schemes among uncoded placement schemes. In a surprising result, it was shown that the YMA scheme is simultaneously optimal for all the demand sets $\boldsymbol{D}_p$, $1 \leq p \leq N$, and is thus a universal scheme.

A coded placement strategy for the $(2, 2)$ network was presented and shown to be optimal in [7]. This idea was generalized to the (N, K) network by Chen et al. [14] (referred to as the CFL scheme in this thesis) and was shown to be optimal when $M \leq \frac{1}{K}$ for the demand set $\boldsymbol{D}_N$. Several coded placement schemes were presented in [15–20] to improve the rates achieved for the demand set $\boldsymbol{D}_N$. In [18], Gómez-Vilardebó introduced a coded placement scheme for the (N, K) cache network with cache size $M = \frac{N}{Kq}$, for $q \in [N]$, to achieve the rate

$$R_N^q = N - \frac{N(N+1)}{K(q+1)} \tag{1.19}$$

for the demand set $\boldsymbol{D}_N$. This scheme was shown to be optimal when $N = K$ and $M = \frac{1}{N-1}$.

Caching Scheme	Placement	Cache Size	Rate	Condition for Optimality
Maddah-Ali & Niesen [7]	Uncoded	$M = \frac{tN}{K}$ where $t \in [K]$	$\frac{K(1-M/N)}{1+(KM/N)}$	$M \geq \frac{N(K-1)}{K}$
Chen et al. [14]	Coded	$M = \frac{1}{K}$	$N(1 - \frac{1}{K})$	$N \leq K$ and $M \leq \frac{1}{K}$
Sahraei & Gastpar [15]	Coded	$\frac{1}{K} \leq M \leq 1$	$\max_{0<L<K} \left(1 + \frac{\left(\sum_{i=\max(0,m-L+1)}^{j-1} \binom{K-L}{i}\binom{L-1}{m-i}\right)}{\binom{K}{m}} + \frac{\left(\sum_{i=j}^{min(m,K-L-1)} \binom{K-L-1}{i}\binom{L}{m-i}\right)}{\binom{K}{m}} \right)$ where $m = KM$ and $j = \lceil M(K-L) \rceil$	-
Wan et al. [9]	Uncoded	$M = \frac{tN}{K}$ where $t \in [K]$	$\begin{cases} N(1-M), & M = \frac{1}{K} \\ N - M - \left(\frac{t(N-1)(1-1/K)}{(K-1)}\right) & 0 \leq t < t_{th} \\ \frac{K-t}{1+t} & t_{th} \leq t \leq K \end{cases}$ where $t_{th} = \frac{NK-2N+1-\sqrt{1+K(N-2)(K(N-2)+2)}}{2(N-1)}$	-
Amiri et al. [10]	Uncoded	$M = \frac{N}{K}$	$N\left(1 - \frac{N+1}{2K}\right)$	-
Amiri & Gündüz [16]	Coded	$M = \frac{N-1}{K}$	$\mathrm{N}\left(1\text{-}\frac{N}{2K}\right)$	-
Wan et al. [11], [12]	Uncoded	$M = \frac{tN}{K}$	$\frac{\binom{K}{t+1}-\binom{K-\min(K,N)}{t+1}}{\binom{K}{t}}$	-
Yu et al. [13]	Uncoded	$M = \frac{tN}{K}$	$\frac{\binom{K}{t+1}-\binom{K-p}{t+1}}{\binom{K}{t}}$ where $t = \frac{KM}{N}$ and p be the number of distinct files requested in the demand	Optimal among uncoded placement schemes
Tian & Chen [17]	Coded	$M = \frac{N(K-t)}{K}$	$\frac{t\left(t(N-1)+K-N\right)}{K(K-1)}$ where $t \in [K]$	-
Gomez-Vilardebo [18]	Coded	$M = \frac{N}{Kq}$	$N - \frac{N(N+1)}{K(q+1)}$ where $q \in [N]$	$N = K$ and $M = \frac{1}{(N-1)}$
Shao et al. [19] [20]	Coded	$M = \frac{t(N-1)}{K-1} + \frac{t+1}{K}$ where $t \in [K]$	$\frac{K-1-t}{t+1} - S(e,t)$, where $S(e,t) = \frac{K\binom{K-N-1}{t+1}+p_e\binom{K-N-1}{t}}{K\binom{K-1}{t}}$ and p_e is the number of files requested by exactly one user	-

Table 1.1: Previous results in coding schemes

In Table 1.1, we summarise the rates achieved by different coding schemes presented in the literature.

Several attempts were presented in [8, 21–26] to reduce the gap between the achievable and optimal rates by deriving new lower bounds. Sengupta et al. [21] reduced the multiplicative gap to 11 which was further reduced to 4.7 in [22, 23] by Wang et al. Ghasemi et al. [25] proposed an algorithm based on the submodularity property of entropy to derive a lower bound, which reduced the gap to 4. This gap was again reduced to 2.315 by Wang et al. [24], and the current best result of 2 was achieved by Yu et al. [26]. Tian showed in [8] that the scheme presented by Yu et al. [13] characterizes the optimal rate memory tradeoff for the $(N, 2)$ cache network with $N \geq 3$. Yu et al. expanded this result to the (N, K) cache network in [26], where $K \leq 5$ and $N \geq 6$.

Several variants of the problem of coded caching has been introduced to study decentralized cache networks [27, 28], cache network with non-uniform demands [29–31], hierarchical cache networks [32], cache networks with multiple servers [33], coded caching with privacy [34], heterogeneous cache networks [35–41], networks with shared cache [42–45], cache network with asynchronous demands [46–48], cache network with erasure broadcast channel [49, 50], cache network with secure delivery [51], cache aided private information retrieval [52, 53], cache aided D2D networks [54, 55] and data shuffling problems with cache aided worker nodes [56–58].

1.6 Thesis Contributions

For the demand set $\boldsymbol{D}_N$, the optimal rate memory tradeoff was known when $0 \leq M \leq \frac{1}{K}$ and $\frac{N}{K}(K-1) \leq M \leq N$. In this context, the contributions of this thesis are as follows:

- We propose a new caching scheme to achieve the memory rate pair $\left(\left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)}\right), \frac{1}{K-1}\right)$.
- For $\left\lceil \frac{K+1}{2} \right\rceil \leq N \leq K$, we derive a matching lower bound to obtain a characterization of the optimal rate memory tradeoff when $M \geq \frac{N}{K}(K-1) - \frac{N-1}{K(K-1)}$.
- For $\left\lceil \frac{K+1}{2} \right\rceil \leq N \leq K$, we derive a new lower bound matching the scheme proposed by Gómez-Vilardebó [18] to obtain a characterization of the optimal rate memory tradeoff when $M \leq \frac{1}{K} + \frac{1}{K(N-1)}$.
- For $1 \leq N \leq \left\lceil \frac{K+1}{2} \right\rceil$, we derive a new lower bound to match the scheme proposed by Yu et al. [13] to obtain a characterization of the optimal rate memory tradeoff when $M \geq \frac{N}{K}(K-2)$.

- For $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$, we derive a new lower bound which improves upon previously known lower bounds for $M \leq \frac{N}{K}$.

We also investigate the multi-objective nature of the problem of coded caching. In this context, the contributions of the thesis are as follows:

- We demonstrate that there is no universal scheme in the general setting where coding is allowed in both the placement phase and the delivery phase and propose the study of Pareto optimal schemes.

- We demonstrate that the scheme proposed by Chen et al. in [14] operate at the Pareto optimal frontier for $M \leq \frac{1}{K}$.

- We demonstrate that the uncoded placement scheme proposed by Yu et al. is Pareto optimal when $M \in [0, \frac{N}{K}] \cup [\frac{N}{K}(K-2), N]$.

1.7 Thesis Organization

The thesis is organized into chapters with the first chapter providing a brief introduction to the problem of coded caching. The focus of the second chapter is on small caches and we derive new lower bounds which demonstrate the optimality of the scheme proposed by Gómez-Vilardebó. Large caches are studied in chapter 3 by deriving new lower bounds and by proposing a new caching scheme that is shown to be optimal. In the fourth chapter, we prove that universal caching schemes do not exist when coding is permitted in the placement phase. As a result we formulate the notion of Pareto optimal schemes in coded caching and demonstrate the Pareto optimality of the schemes proposed in [13] and [14]. In the final chapter, we note a few problems of interest for further investigations.

2

Small Caches

Contents

2.1 Introduction

In this chapter, we consider the demands where each of the N files is required by at least one user (and hence $N \leq K$). The set of all such demands is denoted by $\boldsymbol{D}_N$ and the corresponding rate is denoted by $R_N(M)$, where

$$R_N(M) = \max\{R_{\boldsymbol{d}}(M) \mid \boldsymbol{d} \in \boldsymbol{D}_N\}. \tag{2.1}$$

Maddah-Ali and Niesen in [7] proposed a coding scheme with an uncoded placement phase and a coded delivery phase for the demands in $\boldsymbol{D}_N$ and demonstrated using cut set arguments that the rate achieved by the proposed scheme is within a multiplicative gap of 12 from the optimal rate. Several caching schemes were proposed in [10, 13, 14, 17–20, 61, 62] to improve upon the rate achieved by the scheme proposed in [7]. Despite several lower bounds on the achievable rates being proposed in [21, 24–26, 64], the nature of the optimal rate memory tradeoff is still elusive, except for the $(N, 2)$ cache network. The schemes proposed in [7], [14] provide a characterization of the optimal rate memory tradeoff when $M \in \left[0, \frac{1}{K}\right] \cup \left[\frac{N(K-1)}{K}, N\right]$. For the special case of $N = K$, the scheme proposed in [18] provides a characterization of the optimal rate memory tradeoff when $M \in \left[\frac{1}{N}, \frac{1}{N(N-1)}\right]$. In a surprising result, Yu et al. [13] showed the existence of a universal scheme among caching schemes with an uncoded placement phase. These results are summarised in TABLE 2.1. In this chapter, we focus on small caches where $M \in \left[\frac{1}{K}, \frac{N}{K}\right]$ and derive a set of new lower bounds for the demands in $\boldsymbol{D}_N$.

Caching Scheme	Cache Size, (M)	Rate Memory Tradeoff	Condition
Chen et al. [14]	$\left[0, \frac{1}{K}\right]$	$R_N^*(M) = N - NM$	$N \leq K$
Gómez-Vilardebó [18]	$\left[\frac{1}{N}, \frac{1}{(N-1)}\right]$	$R_N^*(M) = \dfrac{N^2-1}{N} - (N-1)M$	$K = N$
Maddah-Ali and Niesen [7]	$\left[\frac{N}{K}(K-1), N\right]$,	$R_N^*(M) = 1 - \frac{1}{N}M$	-
Yu et al. [13]	$[0, N]$	$R_N(M) = R(r) + (R(r) - R(r+1))\left(r - \frac{N}{K}M\right)$ where $R(r) = \frac{{}^{K}C_{r+1} - {}^{K-N}C_{r+1}}{{}^{K}C_r}$ and $r = \frac{KM}{N}$	Optimal among uncoded prefetching schemes
This chapter	$\left[\frac{1}{K}, \left(\frac{1}{K} + \frac{1}{K(N-1)}\right)\right]$	$R_N^*(M) = \dfrac{KN-1}{K} - (N-1)M$	$\lceil \frac{K+1}{2} \rceil \leq N \leq K$

Table 2.1: Summary of previous work in coded caching

Rest of this chapter is organized as follows. In Section 3.2, we motivate our new bounds with the help of two examples, the $(3, 4)$ cache network and the $(2, 4)$ cache network. We extend these lower bounds for the (N, K) cache network, where $N \leq K$, in Section 3.3. We compare the new lower bounds we derived with the lower bounds presented in the literature in Section 3.4, and we conclude this chapter in Section 3.5.

2.2 Example Networks

In this section, we consider two examples to motivate the results we present in the paper. The $(3,4)$ network is an example for the case $\lceil \frac{K+1}{2} \rceil \leq N \leq K$ and the $(2,4)$ network is an example for the case $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$.

2.2.1 Case I: The $(3,4)$ Cache Network

Here, users $\{U_1, U_2, U_3, U_4\}$ are connected to a server with three files $\{A, B, C\}$ (each of size F bits). Each user U_k has a cache Z_k of size MF bits. For a demand $\boldsymbol{d}$, we have:

Lemma 1. *For the* $(3,4)$ *cache network, achievable memory rate pairs* (M, R_3) *must satisfy the constraint*

$$8M + 4R_3 \geq 11$$

Proof. We have,

$$\begin{aligned}
8M + 4R_3 &\overset{(a)}{\geq} 2H(Z_1) + 2H(Z_2) + H(Z_3) + 3H(Z_4) + 2H(X_{(A,B,C,A)}) + H(X_{(B,C,A,A)}) + H(X_{(C,A,A,B)}) \\
&\overset{(b)}{\geq} H(Z_1, Z_2, X_{(A,B,C,A)}) + H(Z_2, Z_4, X_{(A,B,C,A)}) + H(Z_1, Z_4, X_{(B,C,A,A)}) + H(Z_3, Z_4, X_{(C,A,A,B)}) \\
&\overset{(c)}{=} H(A, B, Z_1, Z_2, X_{(A,B,C,A)}) + H(A, B, Z_2, Z_4, X_{(A,B,C,A)}) + H(A, B, Z_1, Z_4, X_{(B,C,A,A)}) \\
&\quad + H(A, B, Z_3, Z_4, X_{(C,A,A,B)}) \\
&\overset{(b)}{\geq} H(A, B, Z_2, X_{(A,B,C,A)}) + H(A, B, Z_1, Z_2, Z_4, X_{(A,B,C,A)}) + H(A, B, Z_1, Z_4, X_{(B,C,A,A)}) \\
&\quad + H(A, B, Z_3, Z_4, X_{(C,A,A,B)}) \\
&\geq H(A, B, Z_1, Z_2, Z_4) + H(A, B, Z_1, Z_4, X_{(B,C,A,A)}) + H(A, B, Z_2, X_{(A,B,C,A)}) \\
&\quad + H(A, B, Z_3, Z_4, X_{(C,A,A,B)}) \\
&\overset{(b)}{\geq} H(A, B, Z_1, Z_2, Z_4, X_{(B,C,A,A)}) + H(A, B, Z_1, Z_4) + H(A, B, Z_2, X_{(A,B,C,A)}) \\
&\quad + H(A, B, Z_3, Z_4, X_{(C,A,A,B)}) \\
&\overset{(c)}{=} H(A, B, C, Z_1, Z_2, Z_4, X_{(B,C,A,A)}) + H(A, B, Z_2, X_{(A,B,C,A)}) + H(A, B, Z_1, Z_4) \\
&\quad + H(A, B, Z_3, Z_4, X_{(C,A,A,B)})
\end{aligned}$$

$$\overset{(d)}{=}H(A,B,C)+H(A,B,Z_2,X_{(A,B,C,A)})+H(A,B,Z_1,Z_4)+H(A,B,Z_3,Z_4,X_{(C,A,A,B)})$$
$$\overset{(b)}{\geq}H(A,B,C)+H(A,B,Z_2,X_{(A,B,C,A)})+H(A,B,Z_4)+H(A,B,Z_1,Z_3,Z_4,X_{(C,A,A,B)})$$
$$\overset{(c)}{=}H(A,B,C)+H(A,B,Z_2,X_{(A,B,C,A)})+H(A,B,Z_4)+H(A,B,C,Z_1,Z_3,Z_4,X_{(C,A,A,B)})$$
$$\overset{(d)}{=}2H(A,B,C)+H(A,B,Z_2,X_{(A,B,C,A)})+H(A,B,Z_4)$$
$$\geq 2H(A,B,C)+H(A,B,X_{(A,B,C,A)})+H(A,B,Z_4)$$
$$\overset{(e)}{=}2H(A,B,C)+H(A,B,X_{(A,A,B,C)})+H(A,B,Z_4)$$
$$\overset{(b)}{\geq}2H(A,B,C)+H(A,B)+H(A,B,Z_4,X_{(A,A,B,C)})$$
$$\overset{(c)}{=}2H(A,B,C)+H(A,B)+H(A,B,C,Z_4,X_{(A,A,B,C)})$$
$$\overset{(c)}{=}3H(A,B,C)+H(A,B)\geq 11,$$

where

- (a) follows from (1.7) and (1.6),
- (b) follows from the submodularity property of entropy,
- (c) follows from (1.8),
- (d) follows from (1.9),
- (e) follows from (1.13).

□

The above result improves upon the previous results from [7, 18, 21] and is summarised in TABLE 2.2 and Fig. 2.1.

Memory	Rate [18]	Lower Bound [7, 21]	New Lower Bound
$\frac{1}{4} \leq M \leq \frac{3}{8}$	$\frac{11}{4} - 2M$	$R_3 \geq \max\left\{(3-3M), \left(\frac{8}{3}-2M\right)\right\}$	$R_3 \geq \frac{11}{4} - 2M$

Table 2.2: Rate memory tradeoff for the $(3,4)$ cache network

2.2.2 Case II: The $(2,4)$ Cache Network

Here, users $\{U_1,U_2,U_3,U_4\}$ are connected to a server with files $\{A,B\}$ (each of size F bits). Each user U_k has cache Z_k of size MF bits. For a demand $\boldsymbol{d}$, we have:

Lemma 2. *For the $(2,4)$ cache network, achievable memory rate pairs (M,R_2) must satisfy the*

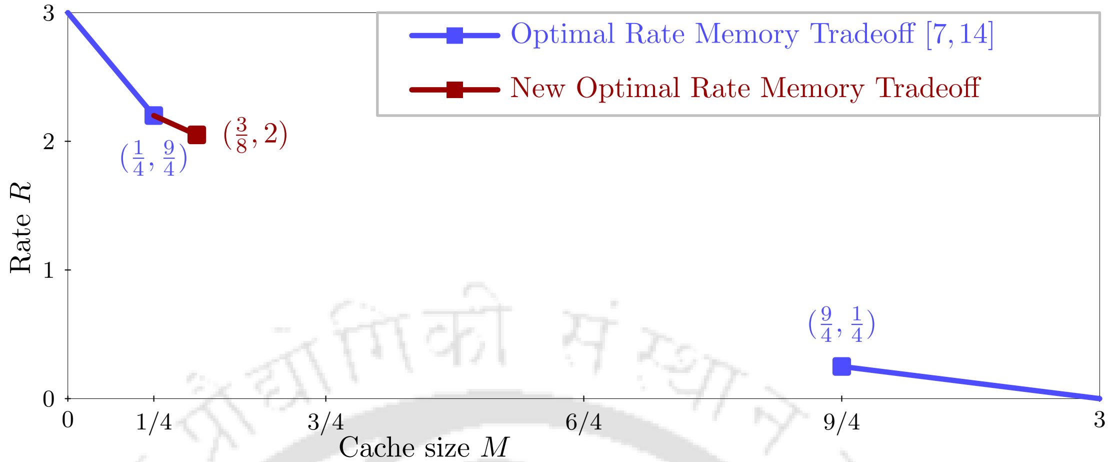

Figure 2.1: Rate memory tradeoff for the $(3, 4)$ cache network

constraint

$$8M + 6R_2 \geq 11. \tag{2.2}$$

Proof. We have,

$$\begin{aligned}
8M + 6R_2 &\geq 2H(Z_1) + H(Z_2) + 2H(Z_3) + 3H(Z_4) + 3H(X_{(A,B,A,A)}) + 2H(X_{(B,A,A,A)}) \\
&\quad + H(X_{(A,A,B,A)}) \\
&\overset{(a)}{\geq} H(Z_1, X_{(A,B,A,A)}) + H(Z_3, X_{(A,B,A,A)}) + H(Z_4, X_{(A,B,A,A)}) + H(Z_3, X_{(B,A,A,A)}) \\
&\quad + H(Z_4, X_{(B,A,A,A)}) + H(Z_4, X_{(A,A,B,A)}) + H(Z_2) + H(Z_1) \\
&\overset{(b)}{=} H(A, Z_1, X_{(A,B,A,A)}) + H(A, Z_3, X_{(A,B,A,A)}) + H(A, Z_4, X_{(A,B,A,A)}) + H(A, Z_3, X_{(B,A,A,A)}) \\
&\quad + H(A, Z_4, X_{(B,A,A,A)}) + H(A, Z_4, X_{(A,A,B,A)}) + H(Z_2) + H(Z_1) \\
&\overset{(a)}{\geq} H(A, Z_1, Z_3, Z_4, X_{(A,B,A,A)}) + 2H(A, X_{(A,B,A,A)}) + H(A, Z_3, Z_4, X_{(B,A,A,A)}) + H(A, X_{(B,A,A,A)}) \\
&\quad + H(A, Z_4, X_{(A,A,B,A)}) + H(Z_2) + H(Z_1) \\
&\overset{(c)}{=} H(A, Z_1, Z_3, Z_4, X_{(A,B,A,A)}) + H(A, Z_3, Z_4, X_{(B,A,A,A)}) + H(A, Z_4, X_{(A,A,B,A)}) + H(A, X_{(A,A,A,B)}) \\
&\quad + H(A, X_{(A,B,A,A)}) + H(A, X_{(B,A,A,A)}) + H(Z_2) + H(Z_1) \\
&\overset{(a)}{\geq} H(A, Z_1, Z_3, Z_4, X_{(A,B,A,A)}) + H(A, Z_3, Z_4, X_{(B,A,A,A)}) + H(A, Z_4, X_{(A,A,B,A)}) + H(A, X_{(A,A,A,B)}) \\
&\quad + H(A, Z_2, X_{(A,B,A,A)}) + H(A, Z_1, X_{(B,A,A,A)}) \\
&\overset{(b)}{=} H(A, Z_1, Z_3, Z_4, X_{(A,B,A,A)}) + H(A, Z_3, Z_4, X_{(B,A,A,A)}) + H(A, Z_4, X_{(A,A,B,A)}) + H(A, X_{(A,A,A,B)})
\end{aligned}$$

$$+ H(A, B, Z_2, X_{(A,B,A,A)}) + H(A, B, Z_1, X_{(B,A,A,A)})$$

$$\overset{(d)}{=} H(A, Z_1, Z_3, Z_4, X_{(A,B,A,A)}) + H(A, Z_3, Z_4, X_{(B,A,A,A)}) + H(A, Z_4, X_{(A,A,B,A)}) + H(A, X_{(A,A,A,B)})$$

$$+ 2H(A, B)$$

$$\overset{(a)}{\geq} H(A, Z_1, Z_3, Z_4, X_{(A,B,A,A)}, X_{(B,A,A,A)}) + H(A, Z_3, Z_4) + H(A, Z_4, X_{(A,A,B,A)}) + H(A, X_{(A,A,A,B)})$$

$$+ 2H(A, B)$$

$$\overset{(b)}{=} H(A, B, Z_1, Z_3, Z_4, X_{(A,B,A,A)}, X_{(B,A,A,A)}) + H(A, Z_3, Z_4) + H(A, Z_4, X_{(A,A,B,A)})$$

$$+ H(A, X_{(A,A,A,B)}) + 2H(A, B)$$

$$\overset{(d)}{=} 3H(A, B) + H(A, Z_3, Z_4) + H(A, Z_4, X_{(A,A,B,A)}) + H(A, X_{(A,A,A,B)})$$

$$\overset{(a)}{\geq} 3H(A, B) + H(A, Z_3, Z_4, X_{(A,A,B,A)}) + H(A, Z_4) + H(A, X_{(A,A,A,B)})$$

$$\overset{(b)}{=} 3H(A, B) + H(A, B, Z_3, Z_4, X_{(A,A,B,A)}) + H(A, Z_4) + H(A, X_{(A,A,A,B)})$$

$$\overset{(d)}{=} 4H(A, B) + H(A, Z_4) + H(A, X_{(A,A,A,B)})$$

$$\overset{(a)}{\geq} 4H(A, B) + H(A, Z_4, X_{(A,A,A,B)}) + H(A)$$

$$\overset{(b)}{=} 4H(A, B) + H(A, B, Z_4, X_{(A,A,A,B)}) + H(A)$$

$$\overset{(d)}{=} 5H(A, B) + H(A) \geq 11,$$

where

(a) follows from the submodularity property of entropy,

(b) follows from (1.8),

(c) follows from (1.13),

(d) follows from (1.9).

□

The above result improves upon the previous results from [7, 18, 21] and is summarised in TABLE 2.3 and Fig. 2.2.

Memory	Rate [18]	Lower Bound [7, 21]	New Lower Bound
$\frac{1}{4} \leq M \leq \frac{1}{2}$	$\frac{32}{18} - \frac{10}{9}M$	$R_2 \geq 2 - 2M$	$R_2 \geq \frac{11}{6} - \frac{4}{3}M$

Table 2.3: Rate memory tradeoff for the (2, 4) cache network

Remark 1. *It should be noted that, for the* $(2, 4)$ *cache network, the bound* $8M + 6R_2 \geq 11$ *is already*

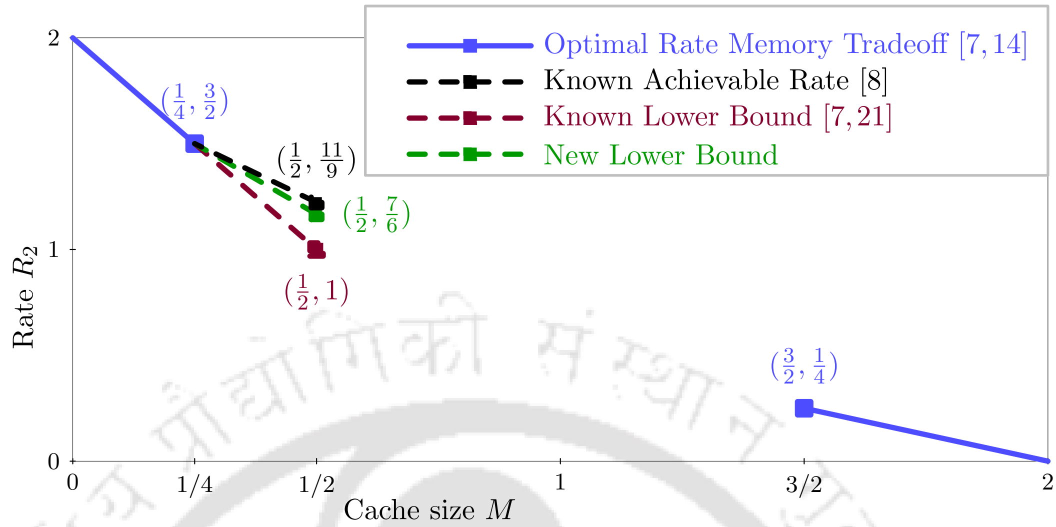

Figure 2.2: Rate memory tradeoff for the $(2, 4)$ cache network

mentioned in [8]. We present the proof above, which can be extended to the (N, K) *cache network.*

2.3 New Lower Bounds

In this section, we derive new lower bounds on the rate memory tradeoff for the (N, K) cache network where $N \leq K$ and cache size $M \in \left[\frac{1}{K}, \frac{N}{K}\right]$. The key ideas we employ are identities (1.8), (1.9) and the properties of symmetric caching schemes stated in (1.13). As in Section II, we consider two cases, namely $\lceil\frac{K+1}{2}\rceil \leq N \leq K$ and $1 \leq N \leq \lceil\frac{K+1}{2}\rceil$.

2.3.1 Case I: $\lceil\frac{K+1}{2}\rceil \leq N \leq K$

Consider the demand

$$\boldsymbol{d}_1 = (W_1, W_2, \ldots, W_N, W_1, W_2, \ldots, W_{K-N}) \tag{2.3}$$

Demands $\{\boldsymbol{d}_l : 2 \leq l \leq K\}$, are obtained from the demand $\boldsymbol{d}_1$ by cyclic left shifts as shown in TABLE 2.4. For the demand $\boldsymbol{d}_l$, let $X_{\boldsymbol{d}_l}$ denote the set of packets broadcast by the server. Consider the user index $\bar{l}$ defined as

$$\bar{l} = \begin{cases} N + 1 - l, \text{ for } 1 \leq l \leq N \\ K + N + 1 - l, \text{ for } N + 1 \leq l \leq K \end{cases} \tag{2.4}$$

It can be noted that in demand $\boldsymbol{d}_l$, the user $U_{\bar{l}}$ requires the file W_N. For $\boldsymbol{S} \subseteq \{U_1, \ldots, U_K\}$, let $Z_{\boldsymbol{S}}$ denote the cache contents of all the users in set $\boldsymbol{S}$. The following lemma are easy to obtain:

	Users	$\boldsymbol{d}_1$	…	$\boldsymbol{d}_i$	…	$\boldsymbol{d}_N$	$\boldsymbol{d}_{N+1}$	…	$\boldsymbol{d}_{N+i}$	…	$\boldsymbol{d}_K$
	U_1	W_1	…	W_i	…	W_N	W_1	…	W_i	…	W_{K-N}
	U_2	W_2	…	W_{i+1}	…	W_1	W_2	…	W_{i+1}	…	W_1
	…	…	…	…	…	…	…	…	…	…	…
	$U_{K-N-i+1}$	$W_{K-N-i+1}$	…	W_{K-N}	…	W_{K-N-i}	$W_{K-N-i+1}$	…	W_{K-N}	…	W_{K-N-i}
$\boldsymbol{A}_i$	$U_{K-N-i+2}$	$W_{K-N-i+2}$	…	W_{K-N+1}	…	$W_{K-N-i+1}$	$W_{K-N-i+2}$	…	W_1	…	$W_{K-N-i-1}$
	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{C}_i$	U_{K-N}	W_{K-N}	…	$W_{K-N+i-1}$	…	W_{K-N-1}	W_{K-N}	…	W_{i-1}	…	W_{K-N-1}
	…	…	…	…	…	…	…	…	…	…	…
	U_{N-i}	W_{N-i}	…	W_{N-1}	…	$W_{2N-K-i-1}$	W_{2N-K-i}	…	W_{2N-K-1}	…	W_{N-i-1}
	U_{N-i+1}	W_{N-i+1}	…	W_N	…	W_{2N-K-i}	$W_{2N-K-i+1}$	…	W_{2N-K}	…	W_{N-i}
	…	…	…	…	…	…	…	…	…	…	…
	U_N	W_N	…	W_{i-1}	…	W_{2N-K}	W_{2N-K+1}	…	W_{2N-K+i}	…	W_{N-1}
	U_{N+1}	W_1	…	W_i	…	W_{2N-K+1}	W_{2N-K+2}	…	$W_{2N-K+i+1}$	…	W_N
	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{E}$	U_{K-i+1}	$W_{K-N-i+1}$	…	W_{K-N}	…	W_{N-i}	W_{N-i+1}	…	W_N	…	W_{K-N-i}
	U_{K-i+2}	$W_{K-N-i+2}$	…	W_1	…	W_{N-i+1}	W_{N-i+2}	…	W_1	…	$W_{K-N-i+1}$
$\boldsymbol{B}_i$	…	…	…	…	…	…	…	…	…	…	…
	U_K	W_{K-N}	…	W_{i-1}	…	W_{N-1}	W_N	…	W_{i-1}	…	W_{K-N-1}

Table 2.4: The set of demands $\{\boldsymbol{d}_l : 1 \leq l \leq K\}$

Lemma 3. *For $\boldsymbol{S}, \boldsymbol{T} \subset \{U_1, \ldots, U_K\} \setminus \{U_{\bar{l}}\}$, we have the identity*

$$H(W_{[N-1]}, Z_{\boldsymbol{S}}, Z_{\bar{l}}) + H(W_{[N-1]}, Z_{\boldsymbol{T}}, X_{\boldsymbol{d}_l}) \geq H(W_{[N-1]}, Z_{\boldsymbol{S}\cap\boldsymbol{T}}) + N,$$

Proof. We have,

$$\begin{aligned} H(W_{[N-1]}, Z_{\boldsymbol{S}}, Z_{\bar{l}}) + H(W_{[N-1]}, Z_{\boldsymbol{T}}, X_{\boldsymbol{d}_l}) &\overset{(a)}{\geq} H(W_{[N-1]}, Z_{\boldsymbol{S}\cap\boldsymbol{T}}) + H(W_{[N-1]}, Z_{\boldsymbol{S}\cup\boldsymbol{T}}, Z_{\bar{l}}, X_{\boldsymbol{d}_l}) \\ &\overset{(b)}{=} H(W_{[N-1]}, Z_{\boldsymbol{S}\cap\boldsymbol{T}}) + H(W_{[N-1]}, W_N, Z_{\boldsymbol{S}\cup\boldsymbol{T}}, Z_{\bar{l}}, X_{\boldsymbol{d}_l}) \\ &\overset{(c)}{=} H(W_{[N-1]}, Z_{\boldsymbol{S}\cap\boldsymbol{T}}) + H(W_{[N]}) \\ &= H(W_{[N-1]}, Z_{\boldsymbol{S}\cap\boldsymbol{T}}) + N \end{aligned}$$

where

- (a) follows from the submodularity property of entropy,
- (b) follows from (1.8),
- (c) follows from (1.9).

□

Lemma 4. *For a sequence of sets* $\boldsymbol{S}_i \subset \{U_1, \ldots, U_K\} \setminus \{U_{\bar{i}}\}$*, such that* $\boldsymbol{S}_i = \boldsymbol{S}_{i+1} \cup \{U_{\overline{i+1}}\}$*, we have*

$$H(W_{[N-1]}, Z_{\boldsymbol{S}_l}) + \sum_{i=l+1}^{j} H(W_{[N-1]}, Z_{\boldsymbol{S}_i}, X_{\boldsymbol{d}_i}) \geq (j-l)N + H(W_{[N-1]}, Z_{\boldsymbol{S}_j})$$

Proof. We have,

$$\begin{aligned}
H(W_{[N-1]}, Z_{\boldsymbol{S}_l}) + &\sum_{i=l+1}^{j} H(W_{[N-1]}, Z_{\boldsymbol{S}_i}, X_{\boldsymbol{d}_i}) \\
&= H(W_{[N-1]}, Z_{\boldsymbol{S}_l}) + H(W_{[N-1]}, Z_{\boldsymbol{S}_{l+1}}, X_{\boldsymbol{d}_{l+1}}) + \sum_{i=l+2}^{j} H(W_{[N-1]}, Z_{\boldsymbol{S}_i}, X_{\boldsymbol{d}_i}) \\
&\overset{(a)}{=} \Big[H(W_{[N-1]}, Z_{\boldsymbol{S}_{l+1}}, Z_{\overline{l+1}}) + H(W_{[N-1]}, Z_{\boldsymbol{S}_{l+1}}, X_{\boldsymbol{d}_{l+1}})\Big] + \sum_{i=l+2}^{j} H(W_{[N-1]}, Z_{\boldsymbol{S}_i}, X_{\boldsymbol{d}_i}) \\
&\overset{(b)}{\geq} N + \Big[H(W_{[N-1]}, Z_{\boldsymbol{S}_{l+1}}) + H(W_{[N-1]}, Z_{\boldsymbol{S}_{l+2}}, X_{\boldsymbol{d}_{l+2}})\Big] + \sum_{i=l+3}^{j} H(W_{[N-1]}, Z_{\boldsymbol{S}_i}, X_{\boldsymbol{d}_i}) \\
&\overset{(c)}{\geq} 2N + H(W_{[N-1]}, Z_{\boldsymbol{S}_{l+2}}) + H(W_{[N-1]}, Z_{\boldsymbol{S}_{l+3}}, X_{\boldsymbol{d}_{l+3}}) + \sum_{i=l+4}^{j} H(W_{[N-1]}, Z_{\boldsymbol{S}_i}, X_{\boldsymbol{d}_i}) \\
&\overset{(d)}{\geq} (j-l)N + H(W_{[N-1]}, Z_{\boldsymbol{S}_j})
\end{aligned}$$

where

(a) follows from definition of set $\boldsymbol{S}_i$,

(b) follows from Lemma 3 with $\boldsymbol{S} \cup \{U_{\bar{l}}\} = \boldsymbol{S}_{l+1}$ and $\boldsymbol{T} = \boldsymbol{S}_{l+2}$,

(c) follows from Lemma 3 with $\boldsymbol{S} \cup \{U_{\bar{l}}\} = \boldsymbol{S}_{l+2}$ and $\boldsymbol{T} = \boldsymbol{S}_{l+3}$,

(d) follows from repeated use of Lemma 3 with $\boldsymbol{S} \cup \{U_{\bar{l}}\} = \boldsymbol{S}_i$ and $\boldsymbol{T} = \boldsymbol{S}_{i+1}$ for $l+3 \leq i \leq j$.

□

In a similar fashion, for a sequence of sets $\boldsymbol{T}_i \subset \{U_1, \ldots, U_K\} \setminus \{U_{\bar{i}}\}$, such that $\boldsymbol{T}_{i+1} = \boldsymbol{T}_i \cup \{U_{\bar{i}}\}$, we can obtain

$$H(W_{[N-1]}, Z_{\boldsymbol{T}_j}, Z_{\bar{j}}) + \sum_{i=l}^{j} H(W_{[N-1]}, Z_{\boldsymbol{T}_i}, X_{\boldsymbol{d}_i}) \geq (j-l+1)N + H(W_{[N-1]}, Z_{\boldsymbol{T}_l}) \tag{2.5}$$

For $1 \leq i \leq N$, consider the sets of users as shown below:

Set	Users	Number	Files Requested in Demand $\boldsymbol{d}_i$
$\boldsymbol{A}_i$	$U_1, \ldots, U_{N-i}$	$N-i$	$W_i, \ldots, W_{N-1}$
$\boldsymbol{B}_i$	$U_{K+2-i}, \ldots, U_K$	$i-1$	$W_1, \ldots, W_{i-1}$
$\boldsymbol{C}_i$	$U_{K+2-N-i}, \ldots, U_{N-i}$	$2N-K-1$	$W_{K-N+1}, \ldots, W_{N-1}$
$\boldsymbol{E}$	$U_{N+1}, \ldots, U_K$	$K-N$	$W_1, \ldots, W_{K-N}$

These sets are also indicated in TABLE 2.4. Note that

$$\boldsymbol{A}_N = \boldsymbol{B}_1 = \boldsymbol{C}_N = \phi \tag{2.6}$$

$$\boldsymbol{A}_{i+1} \cup \{U_{\overline{i+1}}\} = \boldsymbol{A}_i \tag{2.7}$$

$$\boldsymbol{B}_i \cup \{U_{\overline{N+i}}\} = \boldsymbol{B}_{i+1} \tag{2.8}$$

$$\boldsymbol{B}_{K-N} \cup \{U_{\overline{K}}\} = \boldsymbol{B}_{K-N+1} = \boldsymbol{E} \tag{2.9}$$

$$\boldsymbol{A}_i \cap \boldsymbol{C}_i = \boldsymbol{C}_i \tag{2.10}$$

$$\boldsymbol{B}_i \cap \boldsymbol{E} = \begin{cases} \boldsymbol{B}_i \text{ when } 1 \leq i \leq K-N \\ \boldsymbol{E} \text{ when } K-N+1 \leq i \leq N \end{cases} \tag{2.11}$$

It can be noted that in the demands $\boldsymbol{d}_i$ and $\boldsymbol{d}_{N+i}$, the users in set $\boldsymbol{B}_i$ are requesting for the same set of files $\{W_1, \ldots, W_{i-1}\}$ (for $1 \leq i \leq N$). Thus, from (1.13) we have

$$H(W_{[i-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) = H(W_{[i-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_{N+i}}) \tag{2.12}$$

Note that $\mid \boldsymbol{A}_i \cup \boldsymbol{B}_i \mid = \mid \boldsymbol{C}_i \cup \boldsymbol{E} \mid = N-1$. Thus, we have

$$(N-1)M + R_N \geq H(Z_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}) + H(X_{\boldsymbol{d}_i}) \geq H(Z_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) \tag{2.13}$$

Similarly,

$$(N-1)M + R_N \geq H(Z_{\boldsymbol{C}_i \cup \boldsymbol{E}}) + H(X_{\boldsymbol{d}_i}) \geq H(Z_{\boldsymbol{C}_i \cup \boldsymbol{E}}, X_{\boldsymbol{d}_i}) \tag{2.14}$$

Now, we have the following result:

Theorem 1. *For the (N, K) cache network, when $\lceil \frac{K+1}{2} \rceil \leq N \leq K$, achievable memory rate pairs*

(M, R_N) *must satisfy the constraint*

$$K(N-1)M + KR_N \geq KN - 1.$$

Proof. We have,

$$K(N-1)M + KR_N = \sum_{i=1}^{K-N} \Big[(N-1)M + R_N + (N-1)M + R_N\Big] + \sum_{i=K-N+1}^{N} \Big[(N-1)M + R_N\Big]$$

$$\overset{(a)}{\geq} \sum_{i=1}^{K-N} \Big[H(Z_{\boldsymbol{A}_i\cup\boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) + H(Z_{\boldsymbol{C}_i\cup\boldsymbol{E}}, X_{\boldsymbol{d}_i})\Big] + \sum_{i=K-N+1}^{N} H(Z_{\boldsymbol{A}_i\cup\boldsymbol{B}_i}, X_{\boldsymbol{d}_i})$$

$$\overset{(b)}{=} \sum_{i=1}^{K-N} \Big[H(W_{[N-1]}, Z_{\boldsymbol{A}_i\cup\boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) + H(W_{[N-1]}, Z_{\boldsymbol{C}_i\cup\boldsymbol{E}}, X_{\boldsymbol{d}_i})\Big]$$

$$+ \sum_{i=K-N+1}^{N} H(W_{[N-1]}, Z_{\boldsymbol{A}_i\cup\boldsymbol{B}_i}, X_{\boldsymbol{d}_i})$$

$$\overset{(c)}{\geq} \sum_{i=1}^{K-N} \Big[H(W_{[N-1]}, Z_{\boldsymbol{A}_i\cup\boldsymbol{E}}, X_{\boldsymbol{d}_i}) + H(W_{[N-1]}, Z_{\boldsymbol{B}_i\cup\boldsymbol{C}_i}, X_{\boldsymbol{d}_i})\Big]$$

$$+ \sum_{i=K-N+1}^{N} H(W_{[N-1]}, Z_{\boldsymbol{A}_i\cup\boldsymbol{B}_i}, X_{\boldsymbol{d}_i})$$

$$\geq \left[H(W_{[N-1]}, Z_{\boldsymbol{A}_1\cup\boldsymbol{E}}) + \sum_{i=2}^{K-N} H(W_{[N-1]}, Z_{\boldsymbol{A}_i\cup\boldsymbol{E}}, X_{\boldsymbol{d}_i})\right] + \sum_{i=K-N+1}^{N} H(W_{[N-1]}, Z_{\boldsymbol{A}_i\cup\boldsymbol{B}_i}, X_{\boldsymbol{d}_i})$$

$$+ \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_{\boldsymbol{B}_i\cup\boldsymbol{C}_i}, X_{\boldsymbol{d}_i})$$

$$\overset{(d)}{\geq} (K-N-1)N + \left[H(W_{[N-1]}, Z_{\boldsymbol{A}_{K-N}\cup\boldsymbol{E}}) + \sum_{i=K-N+1}^{N} H(W_{[N-1]}, Z_{\boldsymbol{A}_i\cup\boldsymbol{B}_i}, X_{\boldsymbol{d}_i})\right]$$

$$+ \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_{\boldsymbol{B}_i\cup\boldsymbol{C}_i}, X_{\boldsymbol{d}_i})$$

$$\overset{(e)}{=} (K-N-1)N + \left[H(W_{[N-1]}, Z_{\boldsymbol{A}_{K-N}\cup\boldsymbol{B}_{K-N+1}}) + \sum_{i=K-N+1}^{N} H(W_{[N-1]}, Z_{\boldsymbol{A}_i\cup\boldsymbol{B}_i}, X_{\boldsymbol{d}_i})\right]$$

$$+ \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_{\boldsymbol{B}_i\cup\boldsymbol{C}_i}, X_{\boldsymbol{d}_i})$$

$$\overset{(f)}{\geq} (N-1)N + H(W_{[N-1]}, Z_{\boldsymbol{A}_N\cup\boldsymbol{B}_{K-N+1}}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_{\boldsymbol{B}_i\cup\boldsymbol{C}_i}, X_{\boldsymbol{d}_i})$$

$$\overset{(g)}{=} (N-1)N + H(W_{[N-1]}, Z_{\boldsymbol{B}_{K-N+1}}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_{\boldsymbol{B}_i\cup\boldsymbol{C}_i}, X_{\boldsymbol{d}_i})$$

$$
\begin{aligned}
&\geq (N-1)N + H(W_{[N-1]}, Z_{\boldsymbol{B}_{K-N+1}}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) \\
&\overset{(h)}{=} (N-1)N + \left[H(W_{[N-1]}, Z_{\boldsymbol{B}_{K-N+1}}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_{N+i}}) \right] \\
&\overset{(i)}{\geq} (K-1)N + H(W_{[N-1]}, Z_{\boldsymbol{B}_1}) \\
&\overset{(g)}{=} (K-1)N + H(W_{[N-1]}) \geq KN - 1
\end{aligned}
$$

where

- (a) follows from (2.13) and (2.14),
- (b) follows from (1.8) and definition of sets $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, $\boldsymbol{C}_i$ and $\boldsymbol{E}$,
- (c) follows from the submodularity property of entropy, and the facts that $\boldsymbol{A}_i \cap \boldsymbol{C}_i = \boldsymbol{C}_i$ and $\boldsymbol{B}_i \cap \boldsymbol{E} = \boldsymbol{B}_i$ for $1 \leq i \leq K-N$ (refer (2.10) and (2.11)),
- (d) follows from Lemma 4 with $\boldsymbol{S}_i = \boldsymbol{A}_i \cup \boldsymbol{E}$, $l = 1$, $j = K-N$ and (2.7),
- (e) follows from (2.9)
- (f) follows from Lemma 4 with $\boldsymbol{S}_i = \boldsymbol{A}_i \cup \boldsymbol{B}_i$, $l = K-N$, $j = N$ and (2.7),
- (g) follows from (2.6),
- (h) follows from (2.12),
- (i) follows from (2.5) with $\boldsymbol{T}_i = \boldsymbol{B}_i$, $l = 1$, $j = K-N$ and (2.8).

□

2.3.2 Case II: $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$

Consider the demand

$$\boldsymbol{d}_1 = (W_1, W_2, \ldots, W_N, W_1, W_2, \ldots, W_{N-1}, W_1, W_1, \ldots, W_1) \tag{2.15}$$

Demands $\{\boldsymbol{d}_l : 2 \leq l \leq K\}$, are obtained from the demand $\boldsymbol{d}_1$ by cyclic left shifts as shown in TABLE 2.5.

Consider the user index $\bar{l}$ defined as

$$\bar{l} = \begin{cases} N+1-l, \text{ for } 1 \leq l \leq N \\ K+N+1-l, \text{ for } N+1 \leq l \leq K \end{cases} \tag{2.16}$$

It can be noted that in demand $\boldsymbol{d}_l$, the user $U_{\bar{l}}$ requires the file W_N. The following lemma is easy to

	Users	$\boldsymbol{d}_1$	$\dots$	$\boldsymbol{d}_i$	$\dots$	$\boldsymbol{d}_N$	$\boldsymbol{d}_{N+1}$	$\dots$	$\boldsymbol{d}_{N+i}$	$\dots$	$\boldsymbol{d}_{2N-1}$	$\boldsymbol{d}_{2N}$	$\dots$	$\boldsymbol{d}_j$	$\dots$	$\boldsymbol{d}_K$
	U_1	W_1	$\dots$	W_i	$\dots$	W_N	W_1	$\dots$	W_i	$\dots$	W_{N-1}	W_1	$\dots$	W_1	$\dots$	W_1
$\boldsymbol{A}_i$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{J}_i$	U_{N-i}	W_{N-i}	$\dots$	W_{N-1}	$\dots$	W_{N-i-1}	W_{N-i}	$\dots$	W_{N-1}	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_{i-1}
	U_{N-i+1}	W_{N-i+1}	$\dots$	W_N	$\dots$	W_{N-i}	W_{N-i+1}	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_{N-i}
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	U_{N+1}	W_1	$\dots$	W_i	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_N
$\boldsymbol{F}_i$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	U_{2N-i}	W_{N-i}	$\dots$	W_{N-1}	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_{N-i-1}
	U_{2N-i+1}	W_{N-i+1}	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_{N-i}
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	U_{K-j+2}	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1
	U_{K-j+3}	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_2	$\dots$	W_1
$\boldsymbol{S}_j$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{G}_i$	$U_{K+N-j+1}$	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_N	$\dots$	W_1
	$U_{K+N-j+2}$	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1
$\boldsymbol{P}_j$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	U_{K+2N-j}	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_{N-1}	$\dots$	W_1
	$U_{K+2N-j+1}$	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	$\dots$	W_1
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	U_{K-i+1}	W_1	$\dots$	W_1	$\dots$	W_{N-i}	W_{N-i+1}	$\dots$	W_N	$\dots$	W_{N-i}	W_{N-i+1}	$\dots$	W_1	$\dots$	W_1
$\boldsymbol{Q}_j$	U_{K-i+2}	W_1	$\dots$	W_1	$\dots$	W_{N-i+1}	W_{N-i+2}	$\dots$	W_1	$\dots$	W_{N-i+1}	W_{N-i+2}	$\dots$	W_1	$\dots$	W_1
$\boldsymbol{B}_i$	U_{K-i+3}	W_1	$\dots$	W_2	$\dots$	W_{N-i+2}	W_{N-i+3}	$\dots$	W_2	$\dots$	W_{N-i+2}	W_{N-i+3}	$\dots$	W_1	$\dots$	W_1
$\boldsymbol{K}_i$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	U_K	W_1	$\dots$	W_{i-1}	$\dots$	W_{N-1}	W_N	$\dots$	W_{i-1}	$\dots$	W_{N-2}	W_{N-1}	$\dots$	W_1	$\dots$	W_1

Table 2.5: Demand set $\{\boldsymbol{d}_l : 1 \leq l \leq K\}$

obtain.

Lemma 5. *Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \subset \{U_1, \dots, U_K\}$ be such that in demand $\boldsymbol{d}_l$, every user in $\boldsymbol{B}$ requests the file W_1 and users in $\boldsymbol{C}$ together request all the files in $\{W_2, \dots, W_N\}$. We have*

$$H(W_{[N-1]}, Z_{\boldsymbol{A}}, X_{\boldsymbol{d}_l}) + \sum_{i \in \boldsymbol{B}} H(Z_i) + \mid \boldsymbol{B} \mid H(X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid H(Z_{\boldsymbol{C}}) \geq H(W_{[N-1]}, Z_{\boldsymbol{A} \cup \boldsymbol{B}}, X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid N$$

Proof. We have,

$$\begin{aligned} H(W_{[N-1]}, Z_{\boldsymbol{A}}, X_{\boldsymbol{d}_l}) &+ \sum_{i \in \boldsymbol{B}} H(Z_i) + \mid \boldsymbol{B} \mid H(X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid H(Z_{\boldsymbol{C}}) \\ &= H(W_{[N-1]}, Z_{\boldsymbol{A}}, X_{\boldsymbol{d}_l}) + \sum_{i \in \boldsymbol{B}} \Big[H(Z_i) + H(X_{\boldsymbol{d}_l}) \Big] + \mid \boldsymbol{B} \mid H(Z_{\boldsymbol{C}}) \end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\geq} H(W_{[N-1]}, Z_{\boldsymbol{A}}, X_{\boldsymbol{d}_l}) + \sum_{i\in \boldsymbol{B}} H(Z_i, X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid H(Z_{\boldsymbol{C}})\\
&\overset{(b)}{=} H(W_{[N-1]}, Z_{\boldsymbol{A}}, X_{\boldsymbol{d}_l}) + \sum_{i\in \boldsymbol{B}} H(W_1, Z_i, X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid H(Z_{\boldsymbol{C}})\\
&\overset{(a)}{\geq} H(W_{[N-1]}, Z_{\boldsymbol{A}}, X_{\boldsymbol{d}_l}) + \Big[H(W_1, Z_{\boldsymbol{B}}, X_{\boldsymbol{d}_l}) + (\mid \boldsymbol{B} \mid -1) H(W_1, X_{\boldsymbol{d}_l})\Big] + \mid \boldsymbol{B} \mid H(Z_{\boldsymbol{C}})\\
&\overset{(a)}{\geq} H(W_{[N-1]}, Z_{\boldsymbol{A}\cup\boldsymbol{B}}, X_{\boldsymbol{d}_l}) + H(W_1, Z_{\boldsymbol{A}\cap\boldsymbol{B}}, X_{\boldsymbol{d}_l}) + (\mid \boldsymbol{B} \mid -1) H(W_1, X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid H(Z_{\boldsymbol{C}})\\
&\geq H(W_{[N-1]}, Z_{\boldsymbol{A}\cup\boldsymbol{B}}, X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid \Big[H(W_1, X_{\boldsymbol{d}_l}) + H(Z_{\boldsymbol{C}})\Big]\\
&\overset{(a)}{\geq} H(W_{[N-1]}, Z_{\boldsymbol{A}\cup\boldsymbol{B}}, X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid H(W_1, Z_{\boldsymbol{C}}, X_{\boldsymbol{d}_l})\\
&\overset{(c)}{=} H(W_{[N-1]}, Z_{\boldsymbol{A}\cup\boldsymbol{B}}, X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid H(W_{[N]}, Z_{\boldsymbol{C}}, X_{\boldsymbol{d}_l})\\
&\overset{(d)}{=} H(W_{[N-1]}, Z_{\boldsymbol{A}\cup\boldsymbol{B}}, X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid H(W_{[N]})\\
&\geq H(W_{[N-1]}, Z_{\boldsymbol{A}\cup\boldsymbol{B}}, X_{\boldsymbol{d}_l}) + \mid \boldsymbol{B} \mid N
\end{aligned}
$$

where

(a) follows form the submodularity property of entropy,

(b) follows from (1.8) and the definition of $\boldsymbol{B}$,

(c) follows from (1.8) and the definition of $\boldsymbol{C}$,

(d) follows from (1.9).

□

For $1 \leq i \leq N$, consider the sets of users as shown below:

Set	Users	Number	Files Requested in Demand $\boldsymbol{d}_i$
$\boldsymbol{A}_i$	$U_1, \ldots, U_{N-i}$	$N-i$	$W_i, \ldots, W_{N-1}$
$\boldsymbol{B}_i$	$U_{K-i+2}, \ldots, U_K$	$i-1$	$W_1 \ldots, W_{i-1}$
$\boldsymbol{F}_i$	$U_{N+1}, \ldots, U_{2N-i}$	$N-i$	$W_i, \ldots, W_{N-1}$
$\boldsymbol{G}_i$	$U_{2N-i+1}, \ldots, U_{K-i+1}$	$K-2N+1$	W_1
$\boldsymbol{J}_i$	$U_1, \ldots, U_{N-i+1}$	$N-i+1$	$W_i \ldots, W_N$
$\boldsymbol{K}_i$	$U_{K-i+3} \ldots U_K$	$i-2$	$W_2 \ldots, W_{i-1}$

These sets are also indicated in TABLE 2.5. Let

$$\boldsymbol{I}_i = \boldsymbol{J}_i \cup \boldsymbol{K}_i \tag{2.17}$$

$$\boldsymbol{L}_i = \boldsymbol{A}_i \cup \boldsymbol{B}_i \cup \boldsymbol{F}_i \cup \boldsymbol{G}_i \tag{2.18}$$

Note that

$$\boldsymbol{A}_N = \boldsymbol{B}_1 = \boldsymbol{F}_N = \boldsymbol{K}_1 = \boldsymbol{K}_2 = \phi \tag{2.19}$$

$$\boldsymbol{L}_{i+1} \cup \{U_{\overline{i+1}}\} = \boldsymbol{L}_i \tag{2.20}$$

$$\boldsymbol{B}_i \cup \{U_{\overline{N+i}}\} = \boldsymbol{B}_{i+1} \tag{2.21}$$

It can be noted that in the demands $\boldsymbol{d}_i$ and $\boldsymbol{d}_{N+i}$, users in the set $\boldsymbol{B}_i$ are requesting for the same set of files $\{W_1, \ldots, W_{i-1}\}$ (for $1 \leq i \leq N$). Thus, from (1.13) we have

$$H(W_{[i-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) = H(W_{[i-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_{N+i}}) \tag{2.22}$$

Note that $\mid \boldsymbol{A}_i \cup \boldsymbol{B}_i \mid = \mid \boldsymbol{B}_i \cup \boldsymbol{F}_i \mid = N-1$. Thus, we have

$$(N-1)M + R_N \geq H(Z_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}) + H(X_{\boldsymbol{d}_i}) \geq H(Z_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) \tag{2.23}$$

Similarly,

$$(N-1)M + R_N \geq H(Z_{\boldsymbol{B}_i \cup \boldsymbol{F}_i}) + H(X_{\boldsymbol{d}_i}) \geq H(Z_{\boldsymbol{B}_i \cup \boldsymbol{F}_i}, X_{\boldsymbol{d}_i}) \tag{2.24}$$

We can now obtain the following lemma:

Lemma 6. *The sets* $\mathbf{B}_i$ *and* $\mathbf{L}_i$*, defined as above, satisfy*

$$\begin{aligned}(N^2(K-2N+3) - 3N+1)M &+ (N(K-2N+3)-1)R_N \\ &\geq H(W_{[N-1]}, Z_{\boldsymbol{L}_N}) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_{N+i}}) + N((K-2N+2)N-1)\end{aligned}$$

Proof. We have,

$$\begin{aligned}&(N^2(K-2N+3) - 3N+1)M + (N(K-2N+3)-1)R_N \\ &= \left[\sum_{i=1}^{N}\Big[(N-1) + (K-2N+1) + (N-1)(K-2N+1)\Big] + \sum_{i=1}^{N-1}(N-1)\right] M \\ &\quad + \left[\sum_{i=1}^{N}(K-2N+2) + \sum_{i=1}^{N-1} 1\right] R_N\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} \Big[(N-1)M + R_N + (K-2N+1)M + (K-2N+1)(R_N + (N-1)M)\Big] \\
&\quad + \sum_{i=1}^{N-1} [(N-1)M + R_N] \\
&\overset{(a)}{\geq} \sum_{i=1}^{N} \Big[H(Z_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) + \sum_{j \in \boldsymbol{G}_i} H(Z_j) + \mid \boldsymbol{G}_i \mid H(X_{\boldsymbol{d}_i}) + \mid \boldsymbol{G}_i \mid H(Z_{\boldsymbol{I}_i})\Big] + \sum_{i=1}^{N-1} H(Z_{\boldsymbol{B}_i \cup \boldsymbol{F}_i}, X_{\boldsymbol{d}_i}) \\
&\overset{(b)}{=} \sum_{i=1}^{N} \Big[H(W_{[N-1]}, Z_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) + \sum_{j \in \boldsymbol{G}_i} H(Z_j) + \mid \boldsymbol{G}_i \mid H(X_{\boldsymbol{d}_i}) + \mid \boldsymbol{G}_i \mid H(Z_{\boldsymbol{I}_i})\Big] \\
&\quad + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{\boldsymbol{B}_i \cup \boldsymbol{F}_i}, X_{\boldsymbol{d}_i}) \\
&\overset{(c)}{\geq} \sum_{i=1}^{N} \Big[H(W_{[N-1]}, Z_{\boldsymbol{A}_i \cup \boldsymbol{B}_i \cup \boldsymbol{G}_i}, X_{\boldsymbol{d}_i}) + \mid \boldsymbol{G}_i \mid N\Big] + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{\boldsymbol{B}_i \cup \boldsymbol{F}_i}, X_{\boldsymbol{d}_i}) \\
&= H(W_{[N-1]}, Z_{\boldsymbol{A}_N \cup \boldsymbol{B}_N \cup \boldsymbol{G}_N}, X_{\boldsymbol{d}_N}) + \sum_{i=1}^{N-1} \Big[H(W_{[N-1]}, Z_{\boldsymbol{A}_i \cup \boldsymbol{B}_i \cup \boldsymbol{G}_i}, X_{\boldsymbol{d}_i}) + H(W_{[N-1]}, Z_{\boldsymbol{B}_i \cup \boldsymbol{F}_i}, X_{\boldsymbol{d}_i})\Big] \\
&\quad + \sum_{i=1}^{N} \mid \boldsymbol{G}_i \mid N \\
&\overset{(d)}{\geq} H(W_{[N-1]}, Z_{\boldsymbol{A}_N \cup \boldsymbol{B}_N \cup \boldsymbol{G}_N \cup \boldsymbol{F}_N}, X_{\boldsymbol{d}_N}) + \sum_{i=1}^{N-1} \Big[H(W_{[N-1]}, Z_{\boldsymbol{A}_i \cup \boldsymbol{B}_i \cup \boldsymbol{G}_i \cup \boldsymbol{F}_i}, X_{\boldsymbol{d}_i}) + H(W_{[N-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_i})\Big] \\
&\quad + (K-2N+1)N^2 \\
&\overset{(e)}{=} \sum_{i=1}^{N} H(W_{[N-1]}, Z_{\boldsymbol{L}_i}, X_{\boldsymbol{d}_i}) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) + (K-2N+1)N^2 \\
&\geq \Big[H(W_{[N-1]}, Z_{\boldsymbol{L}_1}) + \sum_{i=2}^{N} H(W_{[N-1]}, Z_{\boldsymbol{L}_i}, X_{\boldsymbol{d}_i})\Big] + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) + (K-2N+1)N^2 \\
&\overset{(f)}{\geq} \Big[(N-1)N + H(W_{[N-1]}, Z_{\boldsymbol{L}_N})\Big] + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_i}) + (K-2N+1)N^2 \\
&\overset{(g)}{=} H(W_{[N-1]}, Z_{\boldsymbol{L}_N}) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{\boldsymbol{B}_i}, X_{\boldsymbol{d}_{N+i}}) + N((K-2N+2)N-1)
\end{aligned}
$$

where

- (a) follows from (2.23) and (2.24),
- (b) follows from (1.8) and definition of set $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, and $\boldsymbol{F}_i$,
- (c) follows from Lemma 5 with $\boldsymbol{A} = \boldsymbol{A}_i \cup \boldsymbol{B}_i$, $\boldsymbol{B} = \boldsymbol{G}_i$ and $\boldsymbol{C} = \boldsymbol{I}_i$,
- (d) follows from (2.19) the submodularity property of entropy,
- (e) follows from the definition of $\boldsymbol{L}_i$,

(f) follows from Lemma 4 with $\boldsymbol{S}_i = \boldsymbol{L}_i$, $l = 1$, $j = N$ and (2.20),

(g) follows from (2.22).

□

Now, for $2N \leq j \leq K$, consider another sets of users as shown below:

Set	Users	Number	Files Requested in Demand $\boldsymbol{d}_j$
$\boldsymbol{P}_j$	$U_{K+N+2-j}, \ldots, U_{K+2N-j}$	$N-1$	$W_1, \ldots, W_{N-1}$
$\boldsymbol{Q}_j$	$U_{K+2N+1-j}, \ldots, U_K$	$j-2N$	W_1
$\boldsymbol{S}_j$	$U_{K-j+3}, \ldots, U_{K+N-j+2}$	$N-1$	$W_2, \ldots, W_N$

These sets are also indicated in TABLE 2.5. Let

$$\boldsymbol{T}_j = \boldsymbol{P}_j \cup \boldsymbol{Q}_j \tag{2.25}$$

Note that

$$\boldsymbol{Q}_{2N} = \boldsymbol{S}_{2N} = \phi \tag{2.26}$$

$$\boldsymbol{T}_{j+1} \cup \{U_{\overline{j+1}}\} = \boldsymbol{T}_j \tag{2.27}$$

$$\boldsymbol{T}_K \cup \{U_{\overline{K}}\} = \boldsymbol{L}_N \tag{2.28}$$

$$\boldsymbol{B}_{N-1} \cup \{U_{\overline{2N-1}}\} = \boldsymbol{B}_N = \boldsymbol{T}_{2N} \tag{2.29}$$

Note that $\mid \boldsymbol{P}_j \mid = N-1$. Thus, we have

$$(N-1)M + R_N \geq H(Z_{\boldsymbol{P}_j}) + H(X_{\boldsymbol{d}_j}) \geq H(Z_{\boldsymbol{P}_j}, X_{\boldsymbol{d}_j}) \tag{2.30}$$

The following lemma is easy to obtain:

Lemma 7. *The set $\boldsymbol{T}_j$, as defined above, satisfy*

$$\frac{(K-2N+1)}{2}\Big[\big(N(K-2N+2)-2\big)M + (K-2N+2)R_N\Big] \geq \sum_{j=2N}^{K} H(W_{[N-1]}, Z_{\boldsymbol{T}_j}, X_{\boldsymbol{d}_j}) + \frac{(K-2N+1)(K-2N)}{2}N$$

Proof. We have,

$$
\begin{aligned}
&\frac{(K-2N+1)}{2}\Big[\big(N(K-2N+2)-2\big)M+(K-2N+2)R_N\Big] \\
&= \frac{(K-2N+1)(N(K-2N+2)-2)}{2}M + \frac{(K-2N+1)(K-2N+2)}{2}R_N \\
&= \sum_{j=2N}^{K}\Big(N-1+N(j-2N)\Big)M + \sum_{j=2N}^{K}(j-2N)R_N \\
&= \sum_{j=2N}^{K}\Big[((N-1)M+R_N)+(j-2N)M+(j-2N)(R_N+(N-1)M)\Big] \\
&\overset{(a)}{\geq} \sum_{j=2N}^{K}\Bigg[H(Z_{\boldsymbol{P}_j},X_{\boldsymbol{d}_j})+\sum_{l\in \boldsymbol{Q}_j}H(Z_l)+\mid \boldsymbol{Q}_j\mid H(X_{\boldsymbol{d}_j})+\mid \boldsymbol{Q}_j\mid H(Z_{\boldsymbol{S}_j})\Bigg] \\
&\overset{(b)}{=} \sum_{j=2N}^{K}\Bigg[H(W_{[N-1]},Z_{\boldsymbol{P}_j},X_{\boldsymbol{d}_j})+\sum_{l\in \boldsymbol{Q}_j}H(Z_l)+\mid \boldsymbol{Q}_j\mid H(X_{\boldsymbol{d}_j})+\mid \boldsymbol{Q}_j\mid H(Z_{\boldsymbol{S}_j})\Bigg] \\
&\overset{(c)}{\geq} \sum_{j=2N}^{K}\Big[H(W_{[N-1]},Z_{\boldsymbol{P}_j\cup \boldsymbol{Q}_j},X_{\boldsymbol{d}_j})+\mid \boldsymbol{Q}_j\mid N\Big] \\
&\overset{(d)}{=} \sum_{j=2N}^{K}H(W_{[N-1]},Z_{\boldsymbol{T}_j},X_{\boldsymbol{d}_j})+\sum_{j=2N}^{K}(j-2N)N \\
&= \sum_{j=2N}^{K}H(W_{[N-1]},Z_{\boldsymbol{T}_j},X_{\boldsymbol{d}_j})+\frac{(K-2N+1)(K-2N)}{2}N
\end{aligned}
$$

where

- (a) follows from (2.30) and the fact that $\boldsymbol{Q}_{2N}=\phi$,
- (b) follows from definition of sets $\boldsymbol{P}_j$, $\boldsymbol{Q}_j$ and (1.8),
- (c) follows from Lemma 5 with $\boldsymbol{A}=\boldsymbol{P}_j$, $\boldsymbol{B}=\boldsymbol{Q}_j$ and $\boldsymbol{C}=\boldsymbol{S}_j$,
- (d) follows from the definition of $\boldsymbol{T}_j$.

□

Using the above lemma, we can obtain the following result:

Theorem 2. *For the (N,K) cache network, when $1\leq N<\lceil\frac{K+1}{2}\rceil$, achievable memory rate pairs (M,R_N) must satisfy the constraint*

$$\frac{K(N(K+3)-2(N^2+1))}{2}M+\frac{K(K+3-2N)}{2}R_N\geq\frac{NK(K-2N+3)-2}{2}$$

Proof. We have,

$$
\begin{aligned}
&\frac{K(N(K+3)-2(N^2+1))}{2}M+\frac{K(K+3-2N)}{2}R_N\\
&=\Big[(N^2(K-2N+3)-3N+1)M+(N(K-2N+3)-1)R_N\Big]\\
&+\left[\frac{(K-2N+1)}{2}\Big(\big(N(K-2N+2)-2\big)M+(K-2N+2)R_N\Big)\right]\\
&\overset{(a)}{\geq}\left[H(W_{[N-1]},Z_{\boldsymbol{L}_N})+\sum_{i=1}^{N-1}H(W_{[N-1]},Z_{\boldsymbol{B}_i},X_{\boldsymbol{d}_{N+i}})+N((K-2N+2)N-1)\right]\\
&+\left[\sum_{j=2N}^{K}H(W_{[N-1]},Z_{\boldsymbol{T}_j},X_{\boldsymbol{d}_j})+\frac{(K-2N+1)(K-2N)}{2}N\right]\\
&\overset{(b)}{=}\left[H(W_{[N-1]},Z_{\boldsymbol{T}_K},Z_{\overline{K}})+\sum_{j=2N}^{K}H(W_{[N-1]},Z_{\boldsymbol{T}_j},X_{\boldsymbol{d}_j})\right]+\sum_{i=1}^{N-1}H(W_{[N-1]},Z_{\boldsymbol{B}_i},X_{\boldsymbol{d}_{(N+i)}})\\
&+\frac{(K(K-2N+1)+2N-2)}{2}N\\
&\overset{(c)}{\geq}\Big[(K-2N+1)N+H(W_{[N-1]},Z_{\boldsymbol{T}_{2N}})\Big]+\sum_{i=1}^{N-1}H(W_{[N-1]},Z_{\boldsymbol{B}_i},X_{\boldsymbol{d}_{(N+i)}})\\
&+\frac{(K(K-2N+1)+2N-2)}{2}N\\
&\overset{(d)}{=}\left[H(W_{[N-1]},Z_{\boldsymbol{B}_{N-1}},Z_{\overline{2N-1}})+\sum_{i=1}^{N-1}H(W_{[N-1]},Z_{\boldsymbol{B}_i},X_{\boldsymbol{d}_{(N+i)}})\right]+\frac{(K(K-2N+3)-2N)}{2}N\\
&\overset{(e)}{\geq}\Big[(N-1)N+H(W_{[N-1]},Z_{\boldsymbol{B}_1})\Big]+\frac{(K(K-2N+3)-2N)}{2}N\\
&\overset{(f)}{=}H(W_{[N-1]})+\frac{(K(K-2N+3)-2)}{2}N\\
&\geq\frac{NK(K-2N+3)-2}{2}
\end{aligned}
$$

where

- (a) follows from Lemma 6 and Lemma 7,
- (b) follows from (2.28),
- (c) follows from (2.5) with $\boldsymbol{T}_i=\boldsymbol{T}_j$, $l=2N$, $j=K$ and (2.27),
- (d) follows from (2.29),
- (e) follows from (2.5) with $\boldsymbol{T}_i=\boldsymbol{B}_i$, $l=1$, $j=N-1$ and (2.21),
- (f) follows from (2.19).

□

2.4 Comparison with Previous Bounds

In [7], Maddah-Ali and Niesen derived a lower bound on achievable rates using cut set arguments, which was further improved in [21, 24–26, 64]. A comparison between these lower bounds and the new lower bounds in Section III, at cache size $M = \left(\frac{1}{K} + \frac{1}{K(N-1)}\right)$, is given in TABLE 2.6. It can be noted that the new bounds improve upon the previous ones. For the (N, K) cache

Lower Bound	Case I: $\lceil\frac{K+1}{2}\rceil \leq N \leq K$	Case II: $1 \leq N \leq \lceil\frac{K+1}{2}\rceil$
Cut Set bound [7]	$N - \frac{N^2}{(N-1)K}$	$N - \frac{N^2}{(N-1)K}$
Ghasemi & Ramamoorthy [25]	$N - \frac{N^2}{(N-1)K}$	$N - \frac{N^2}{(N-1)K}$
Ajaykrishnan et al. [64]	$N - \frac{N^2}{(N-1)K}$	$N - \frac{N^2}{(N-1)K}$
Wang et al. [24]	$N - \frac{N^2}{(N-1)K}$	$N - \frac{N^2}{(N-1)K}$
Yu et al. [26]	$N - \frac{N^2}{(N-1)K} + \frac{1}{K(N-1)}\left(N - K + \frac{K}{N}\right)$	$N - \frac{N^2}{(N-1)K}$
Sengupta et al. [21]	$N - \frac{N^2}{(N-1)K} + \frac{1}{K(N-1)}\left(N - K + \frac{K}{N}\right)$	$N - \frac{N^2}{(N-1)K}$
New lower bound	$N - \frac{N^2}{(N-1)K} + \frac{1}{K(N-1)}$	$N - \frac{N^2}{K(N-1)} + \frac{2}{K(N-1)(K+3-2N)}$

Table 2.6: Comparison with previous lower bounds for $M = \left(\frac{1}{K} + \frac{1}{K(N-1)}\right)$

network, the scheme proposed by Gómez-Vilardebó in [18] achieves memory rate pairs $(M, R_N^G) = \left(M, \frac{KN-1}{K} - (N-1)M\right)$, for $M \in \left[\frac{1}{K}, \left(\frac{1}{K} + \frac{1}{K(N-1)}\right)\right]$. From Theorem 1, when $\lceil\frac{K+1}{2}\rceil \leq N \leq K$, we have that all achievable memory rate pairs satisfy the constraint

$$R_N \geq \frac{KN-1}{K} - (N-1)M = R_N^G(M)$$

Thus we have:

Theorem 3. *For the (N, K) cache network, when $\lceil\frac{K+1}{2}\rceil \leq N \leq K$, the optimal rate memory tradeoff is given by*

$$R_N^*(M) = \frac{KN-1}{K} - (N-1)M \tag{2.31}$$

where $M \in \left[\frac{1}{K}, \left(\frac{1}{K} + \frac{1}{K(N-1)}\right)\right]$.

Remark 2. *In [18], with the help of the lower bounds derived in [21] and [26], Gómez-Vilardebó showed that when $K = N$ and $M = \frac{1}{N-1}$, his scheme is optimal. We extend his result to the case where $\lceil\frac{K+1}{2}\rceil \leq N \leq K$ in Theorem 3.*

2.5 Conclusions

In this chapter, we considered the canonical (N, K) cache network where $N \leq K$ and $M \in \left[0, \frac{N}{K}\right]$. We derived a new set of lower bounds on the achievable rate when each file in the server is requested by at least one user. Using these lower bounds, we showed that when $\lceil \frac{K+1}{2} \rceil \leq N \leq K$ the scheme proposed in [18] is optimal for $M \in \left[\frac{1}{K}, \left(\frac{1}{K} + \frac{1}{K(N-1)}\right)\right]$. For the case $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$, the new lower bound was shown to improve upon the previous lower bounds, but a matching scheme is still not known. The work presented forms another step in the attempt to find a characterization of the optimal rate memory tradeoff for coded caching and is illustrated in Fig. 2.3.

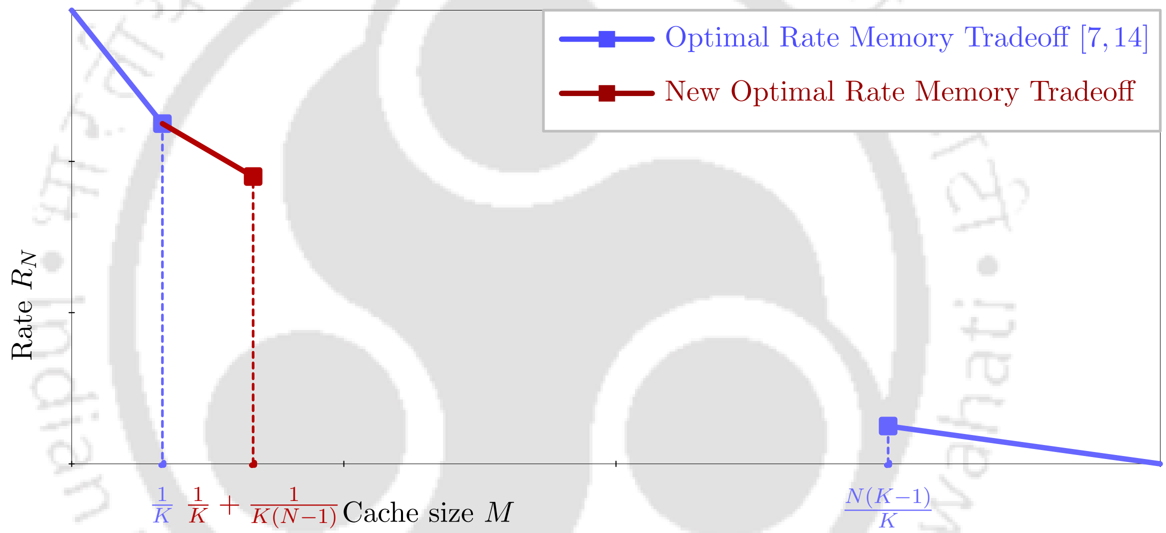

Figure 2.3: Optimal rate memory tradeoff for the (N, K) cache network when $\lceil \frac{K+1}{2} \rceil < N \leq K$

3

Large Caches

Contents

3.1 Introduction

We continue our research into enhancing the characterisation of the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_N$ in this chapter. The performance obtained by previous schemes is summarised in TABLE 3.1. In this chapter, we focus on large caches where $M \geq \frac{N}{K}(K-2)$. For the (N, K) cache network where $N \leq K$, we propose a new coding scheme and derive a set of new lower bounds for the demands in $\boldsymbol{D}_N$, which extend characterization of the optimal rate memory tradeoff.

Caching Scheme	Cache Size, (M)	Rate Memory Tradeoff	Condition
Chen et al. [14]	$\left[0, \frac{1}{K}\right]$	$R_N^*(M) = N - NM$	$N \leq K$
Gómez-Vilardebó [18]	$\left[\frac{1}{N}, \frac{1}{(N-1)}\right]$	$R_N^*(M) = \dfrac{N^2-1}{N} - (N-1)M$	$K = N$
Maddah-Ali and Niesen [7]	$\left[\frac{N}{K}(K-1), N\right]$,	$R_N^*(M) = 1 - \frac{1}{N}M$	-
Yu et al. [13]	$[0, N]$	$R_N(M) = R(r) + (R(r) - R(r+1))\left(r - \frac{N}{K}M\right)$ where $R(r) = \frac{{}^{K}C_{r+1} - {}^{K-N}C_{r+1}}{{}^{K}C_r}$ and $r = \frac{KM}{N}$	Optimal among uncoded prefetching schemes
Vijith et al. [61], [62]	$\left[N-1-\frac{1}{N}, N-1\right]$	$R_N^*(M) = \frac{N+1}{N} - \frac{1}{N-1}M$	$N = K$
Chapter 3	$\left[\frac{1}{K}, \left(\frac{1}{K} + \frac{1}{K(N-1)}\right)\right]$	$R_N^*(M) = \dfrac{KN-1}{K} - (N-1)M$	$\left\lceil\frac{K+1}{2}\right\rceil \leq N \leq K$
This chapter	$\left[\left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)}\right), \frac{N}{K}(K-1)\right]$	$R_N^*(M) = \frac{(KN-1)}{K(N-1)} - \frac{1}{(N-1)}M$	$\left\lceil\frac{K+1}{2}\right\rceil \leq N \leq K$
	$\left[\frac{N}{K}(K-2), \frac{N}{K}(K-1)\right]$	$R_N^*(M) = \frac{K^2+K-2}{K(K-1)} - \frac{(K+1)}{N(K-1)}M$	$1 \leq N \leq \left\lceil\frac{K+1}{2}\right\rceil$

Table 3.1: Summary of previous work in coded caching

3.2 Example Networks

As a prelude to the results presented in Section 3.3 and 3.4, we consider two example networks.

3.2.1 Case I: The $(3, 4)$ Cache Network

Here, users $\{U_1, U_2, U_3, U_4\}$ are connected to a server with files $\{A, B, C\}$ (each of size F bits). Each user U_k has a cache Z_k of size MF bits. We now describe a symmetric caching scheme for the case $M = \frac{25}{12}$. During the placement phase, every file is split into 12 disjoint subfiles, each of size $\frac{1}{12}F$ bits. The subfiles are:

File	Subfiles
A	A^{12}, A^{21}, A^{13}, A^{31}, A^{14}, A^{41}, A^{23}, A^{32}, A^{24}, A^{42}, A^{34}, A^{43}
B	B^{12}, B^{21}, B^{13}, B^{31}, B^{14}, B^{41}, B^{23}, B^{32}, B^{24}, B^{42}, B^{34}, B^{43}
C	C^{12}, C^{21}, C^{13}, C^{31}, C^{14}, C^{41}, C^{23}, C^{32}, C^{24}, C^{42}, C^{34}, C^{43}

The server places 18 uncoded packets (stage 1) and 7 coded packets (stage 2) in each user's cache as shown in TABLE 3.2. Each of these packets are of size $\frac{1}{12}F$ bits and they together occupy $\frac{25}{12}F$ bits.

Cache	Stage 1						Stage 2		
Z_1	A^{23}	A^{24}	A^{32}	A^{34}	A^{42}	A^{43}	$A^{12}-A^{13}$	$A^{12}-A^{14}$	
	B^{23}	B^{24}	B^{32}	B^{34}	B^{42}	B^{43}	$B^{12}-B^{13}$	$B^{12}-B^{14}$	
	C^{23}	C^{24}	C^{32}	C^{34}	C^{42}	C^{43}	$C^{12}-C^{13}$	$C^{12}-C^{14}$	$A^{12}+B^{12}+C^{12}$
Z_2	A^{13}	A^{14}	A^{31}	A^{34}	A^{41}	A^{43}	$A^{23}-A^{24}$	$A^{23}-A^{21}$	
	B^{13}	B^{14}	B^{31}	B^{34}	B^{41}	B^{43}	$B^{23}-B^{24}$	$B^{23}-B^{21}$	
	C^{13}	C^{14}	C^{31}	C^{34}	C^{41}	C^{43}	$C^{23}-C^{24}$	$C^{23}-C^{21}$	$A^{23}+B^{23}+C^{23}$
Z_3	A^{21}	A^{24}	A^{12}	A^{14}	A^{42}	A^{41}	$A^{34}-A^{31}$	$A^{34}-A^{32}$	
	B^{21}	B^{24}	B^{12}	B^{14}	B^{42}	B^{41}	$B^{34}-B^{31}$	$B^{34}-B^{32}$	
	C^{21}	C^{24}	C^{12}	C^{14}	C^{42}	C^{41}	$C^{34}-C^{31}$	$C^{34}-C^{32}$	$A^{34}+B^{34}+C^{34}$
Z_4	A^{23}	A^{21}	A^{32}	A^{31}	A^{12}	A^{13}	$A^{41}-A^{42}$	$A^{41}-A^{43}$	
	B^{23}	B^{21}	B^{32}	B^{31}	B^{12}	B^{13}	$B^{41}-B^{42}$	$B^{41}-B^{43}$	
	C^{23}	C^{21}	C^{32}	C^{31}	C^{12}	C^{13}	$C^{41}-C^{42}$	$C^{41}-C^{43}$	$A^{41}+B^{41}+C^{41}$

Table 3.2: Cache contents placed in stage 1 and stage 2

To understand how the delivery phase works, consider a demand $\boldsymbol{d} = (P, P, Q, R)$ where P, Q and R are distinct files in $\{A, B, C\}$. In response to this demand, the server broadcasts a set of packets

$$X_{\boldsymbol{d}} = \left\{ \begin{array}{c} Q^{13} + R^{14} - P^{12} \\ Q^{23} + R^{24} - P^{21} \\ R^{34} + \frac{1}{2}P^{31} + \frac{1}{2}P^{32} \\ Q^{43} + \frac{1}{2}P^{41} + \frac{1}{2}P^{42} \end{array} \right\}.$$

As $X_{\boldsymbol{d}}$ has four packets, of size $\frac{1}{12}F$ bits each, the load experienced by the shared link is $R_3F = \frac{1}{3}F$ bits and thus the rate is $R_3 = \frac{1}{3}$.

Let us consider U_1 to understand how the requested file is obtained from $X_{\boldsymbol{d}}$ and Z_1. Note that the user has subfiles $P^{23}, P^{32}, P^{24}, P^{42}, P^{34}$ and P^{43} in Z_1 and require subfiles P^{12}, P^{21}, P^{13}, P^{31}, P^{14} and P^{41} to compute the requested file P. The user can compute subfiles P^{21}, P^{31}, P^{41} and P^{12}

by combining received packets and cached packets as shown below:

Received Packet	Cached Packets	Computed Subfile
$Q^{23} + R^{24} - P^{21}$	Q^{23}, R^{24}	P^{21}
$R^{34} - \frac{1}{2}P^{31} - \frac{1}{2}P^{32}$	R^{34}, P^{32}	P^{31}
$Q^{43} - \frac{1}{2}P^{41} - \frac{1}{2}P^{42}$	Q^{43}, P^{42}	P^{41}
$Q^{13} + R^{14} - P^{12}$	$Q^{12} - Q^{13}$, $R^{12} - R^{14}$, $P^{12} + Q^{12} + R^{12}$	P^{12}

Combining the subfile P^{12} with cached packets $P^{12} - P^{13}$ and $P^{12} - P^{14}$, U_1 obtains subfiles P^{13} and P^{14}. The other users can proceed in similar fashion. We summarise as:

Lemma 8. *The memory rate pair* $\left(\frac{25}{12}, \frac{1}{3}\right)$ *is achievable by symmetric caching schemes for the* $(3,4)$ *cache network.*

The caching scheme proposed in [7] achieves the memory rate pair $\left(\frac{9}{4}, \frac{1}{4}\right)$, and by memory sharing between that scheme and the proposed scheme, we can achieve all memory rate pairs $\left(M, \frac{11}{8} - \frac{1}{2}M\right)$, where $M \in \left[\frac{25}{12}, \frac{9}{4}\right]$. We obtain a matching lower bound in the following lemma:

Lemma 9. *For the* $(3,4)$ *cache network, achievable memory rate pairs* (M, R_3) *must satisfy the constraint*

$$4M + 8R_3 \geq 11.$$

Proof. We have,

$$\begin{aligned}
4M + 8R_3 &\overset{(a)}{\geq} 2H(Z_1) + H(Z_2) + H(Z_3) + 2H(X_{(A,B,C,A)}) + 2H(X_{(B,C,A,A)}) + 3H(X_{(A,A,B,C)}) \\
&\quad + H(X_{(C,A,A,B)}) \\
&\overset{(b)}{\geq} H(Z_1, X_{(A,B,C,A)}, X_{(B,C,A,A)}) + H(Z_1, X_{(B,C,A,A)}, X_{(A,A,B,C)}) + H(Z_2, X_{(A,B,C,A)}, X_{(A,A,B,C)}) \\
&\quad + H(Z_3, X_{(C,A,A,B)}, X_{(A,A,B,C)}) \\
&\overset{(c)}{=} H(A, B, Z_1, X_{(A,B,C,A)}, X_{(B,C,A,A)}) + H(A, B, Z_1, X_{(B,C,A,A)}, X_{(A,A,B,C)}) \\
&\quad + H(A, B, Z_2, X_{(A,B,C,A)}, X_{(A,A,B,C)}) + H(A, B, Z_3, X_{(C,A,A,B)}, X_{(A,A,B,C)}) \\
&\overset{(b)}{\geq} H(A, B, Z_1, X_{(B,C,A,A)}) + H(A, B, Z_1, X_{(A,B,C,A)}, X_{(B,C,A,A)}, X_{(A,A,B,C)}) \\
&\quad + H(A, B, Z_2, X_{(A,B,C,A)}, X_{(A,A,B,C)}) + H(A, B, Z_3, X_{(C,A,A,B)}, X_{(A,A,B,C)}) \\
&\geq H(A, B, X_{(A,B,C,A)}, X_{(B,C,A,A)}, X_{(A,A,B,C)}) + H(A, B, Z_2, X_{(A,B,C,A)}, X_{(A,A,B,C)}) \\
&\quad + H(A, B, Z_3, X_{(C,A,A,B)}, X_{(A,A,B,C)}) + H(A, B, Z_1, X_{(B,C,A,A)})
\end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{\geq} H(A,B,Z_2,X_{(A,B,C,A)},X_{(B,C,A,A)},X_{(A,A,B,C)}) + H(A,B,X_{(A,B,C,A)},X_{(A,A,B,C)}) \\
&\quad + H(A,B,Z_3,X_{(C,A,A,B)},X_{(A,A,B,C)}) + H(A,B,Z_1,X_{(B,C,A,A)}) \\
&\overset{(c)}{=} H(A,B,C,Z_2,X_{(A,B,C,A)},X_{(B,C,A,A)},X_{(A,A,B,C)}) + H(A,B,X_{(A,B,C,A)},X_{(A,A,B,C)}) \\
&\quad + H(A,B,Z_3,X_{(C,A,A,B)},X_{(A,A,B,C)}) + H(A,B,Z_1,X_{(B,C,A,A)}) \\
&\overset{(d)}{=} H(A,B,C) + H(A,B,X_{(A,B,C,A)},X_{(A,A,B,C)}) + H(A,B,Z_3,X_{(C,A,A,B)},X_{(A,A,B,C)}) \\
&\quad + H(A,B,Z_1,X_{(B,C,A,A)}) \\
&\overset{(b)}{\geq} H(A,B,C) + H(A,B,Z_3,X_{(C,A,A,B)},X_{(A,B,C,A)},X_{(A,A,B,C)}) + H(A,B,X_{(A,A,B,C)}) \\
&\quad + H(A,B,Z_1,X_{(B,C,A,A)}) \\
&\overset{(c)}{=} H(A,B,C) + H(A,B,C,Z_3,X_{(C,A,A,B)},X_{(A,B,C,A)},X_{(A,A,B,C)}) + H(A,B,X_{(A,A,B,C)}) \\
&\quad + H(A,B,Z_1,X_{(B,C,A,A)}) \\
&\overset{(d)}{=} 2H(A,B,C) + H(A,B,X_{(A,A,B,C)}) + H(A,B,Z_1,X_{(B,C,A,A)}) \\
&\geq 2H(A,B,C) + H(A,B,X_{(A,A,B,C)}) + H(A,B,Z_1) \\
&\overset{(e)}{=} 2H(A,B,C) + H(A,B,X_{(A,A,B,C)}) + H(A,B,Z_4) \\
&\overset{(b)}{\geq} H(A,B,Z_4,X_{(A,A,B,C)}) + H(A,B) + 2H(A,B,C) \\
&\overset{(c)}{=} H(A,B,C,Z_4,X_{(A,A,B,C)}) + H(A,B) + 2H(A,B,C) \\
&\overset{(d)}{=} H(A,B,C) + H(A,B) + 2H(A,B,C) \geq 11
\end{aligned}
$$

where

- (a) follows from (1.6) and (1.7),
- (b) follows from the submodularity property of entropy,
- (c) follows from (1.8),
- (d) follows from (1.9),
- (e) follows from (1.13).

□

The above observations improve upon the previous results from [7,13] and is summarised in TABLE 3.3 and Fig. 3.1.

Memory	Rate [7,13]	Lower Bound [7,13]	New Rate	New Lower Bound
$\frac{25}{12} \leq M \leq \frac{9}{4}$	$\frac{3}{2} - \frac{5}{9}M$	$R_3 \geq 1 - \frac{1}{3}M$	$\frac{11}{8} - \frac{1}{2}M$	$R_3 \geq \frac{11}{8} - \frac{1}{2}M$

Table 3.3: Rate memory tradeoff for the $(3,4)$ cache network

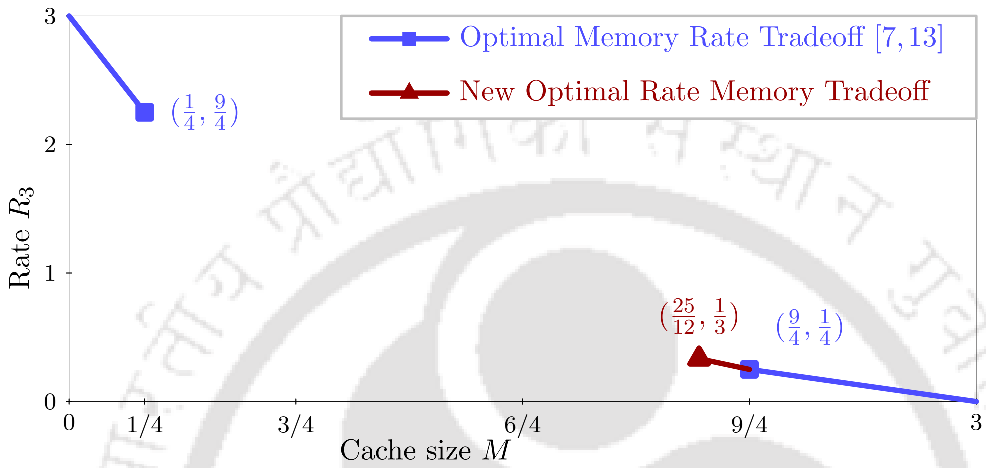

Figure 3.1: Optimal rate memory tradeoff for the $(3,4)$ cache network

3.2.2 Case II: The $(2,4)$ Cache Network

Here, users $\{U_1, U_2, U_3, U_4\}$ are connected to a server with files $\{A, B\}$ (each of size F bits). Each user U_k has cache Z_k of size MF bits. The caching scheme proposed in [13] can achieve all memory rate pairs $\left(M, \frac{3}{2} - \frac{5}{6}M\right)$, where $M \in \left[1, \frac{3}{2}\right]$. We obtain a matching lower bound in the following lemma:

Lemma 10. *For the $(2,4)$ cache network, achievable memory rate pairs (M, R_2) must satisfy the constraint*

$$5M + 6R_2 \geq 9.$$

Proof. We have,

$$\begin{aligned}
5M+6R_2 &\overset{(a)}{\geq} 3H(Z_1) + 2H(Z_2) + H(Z_4) + 3H(X_{(A,A,B,A)}) + 2H(X_{(A,A,A,B)}) + H(X_{(A,B,A,A)}) - H(Z_1) \\
&\overset{(b)}{\geq} H(Z_1, X_{(A,A,B,A)}) + H(Z_1, X_{(A,A,A,B)}) + H(Z_1, X_{(A,B,A,A)}) + H(Z_2, X_{(A,A,B,A)}) + H(Z_2, X_{(A,A,A,B)}) \\
&\quad + H(Z_4, X_{(A,A,B,A)}) - H(Z_1) \\
&\overset{(c)}{=} H(A, Z_1, X_{(A,A,B,A)}) + H(A, Z_1, X_{(A,A,A,B)}) + H(A, Z_1, X_{(A,B,A,A)}) + H(A, Z_2, X_{(A,A,B,A)}) \\
&\quad + H(A, Z_2, X_{(A,A,A,B)}) + H(A, Z_4, X_{(A,A,B,A)}) - H(Z_1)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{\geq} H(A, Z_1, X_{(A,A,B,A)}, X_{(A,A,A,B)}, X_{(A,B,A,A)}) + 2H(A, Z_1) + H(A, Z_2, X_{(A,A,B,A)}, X_{(A,A,A,B)}) \\
&\quad + H(A, Z_2) + H(A, Z_4, X_{(A,A,B,A)}) - H(Z_1) \\
&\overset{(d)}{=} H(A, Z_1, X_{(A,A,B,A)}, X_{(A,A,A,B)}, X_{(A,B,A,A)}) + H(A, Z_2, X_{(A,A,B,A)}, X_{(A,A,A,B)}) \\
&\quad + H(A, Z_4, X_{(A,A,B,A)}) + H(A, Z_2) + H(A, Z_1) + H(B, Z_1) - H(Z_1) \\
&\overset{(e)}{=} H(A, Z_1, X_{(A,A,B,A)}, X_{(A,A,A,B)}, X_{(A,B,A,A)}) + H(A, Z_2, X_{(A,A,B,A)}, X_{(A,A,A,B)}) \\
&\quad + H(A, Z_4, X_{(A,A,B,A)}) + H(A, Z_3) + H(A, Z_1) + H(B, Z_1) - H(Z_1) \\
&\overset{(b)}{\geq} H(A, Z_1, X_{(A,A,B,A)}, X_{(A,A,A,B)}, X_{(A,B,A,A)}) + H(A, Z_2, X_{(A,A,B,A)}, X_{(A,A,A,B)}) \\
&\quad + H(A, Z_4, X_{(A,A,B,A)}) + H(A, Z_3) + H(A, B, Z_1) + H(Z_1) - H(Z_1) \\
&\overset{(f)}{=} H(A, Z_1, X_{(A,A,B,A)}, X_{(A,A,A,B)}, X_{(A,B,A,A)}) + H(A, Z_2, X_{(A,A,B,A)}, X_{(A,A,A,B)}) \\
&\quad + H(A, Z_4, X_{(A,A,B,A)}) + H(A, Z_3) + H(A, B) \\
&\overset{(b)}{\geq} H(A, Z_1, Z_2, X_{(A,A,B,A)}, X_{(A,A,A,B)}, X_{(A,B,A,A)}) + H(A, X_{(A,A,B,A)}, X_{(A,A,A,B)}) \\
&\quad + H(A, Z_4, X_{(A,A,B,A)}) + H(A, Z_3) + H(A, B) \\
&\overset{(c)}{=} H(A, B, Z_1, Z_2, X_{(A,A,B,A)}, X_{(A,A,A,B)}, X_{(A,B,A,A)}) + H(A, X_{(A,A,B,A)}, X_{(A,A,A,B)}) \\
&\quad + H(A, Z_4, X_{(A,A,B,A)}) + H(A, Z_3) + H(A, B) \\
&\overset{(f)}{=} 2H(A, B) + H(A, X_{(A,A,B,A)}, X_{(A,A,A,B)}) + H(A, Z_4, X_{(A,A,B,A)}) + H(A, Z_3) \\
&\overset{(b)}{\geq} 2H(A, B) + H(A, Z_4, X_{(A,A,B,A)}, X_{(A,A,A,B)}) + H(A, X_{(A,A,B,A)}) + H(A, Z_3) \\
&\overset{(c)}{=} 2H(A, B) + H(A, B, Z_4, X_{(A,A,B,A)}, X_{(A,A,A,B)}) + H(A, X_{(A,A,B,A)}) + H(A, Z_3) \\
&\overset{(f)}{=} 3H(A, B) + H(A, X_{(A,A,B,A)}) + H(A, Z_3) \\
&\overset{(b)}{\geq} 3H(A, B) + H(A, Z_3, X_{(A,A,B,A)}) + H(A) \\
&\overset{(c)}{=} 3H(A, B) + H(A, B, Z_3, X_{(A,A,B,A)}) + H(A) \\
&\overset{(f)}{=} 4H(A, B) + H(A) \geq 9,
\end{aligned}
$$

where

(*a*) follows from (1.6) and (1.7),

(*b*) follows from the submodularity property of entropy,

(*c*) follows from (1.8),

(*d*) follows from (1.15),

(*e*) follows from (1.13),

(*f*) follows from (1.9)

□

The above observations improve upon the previous results from [7,13] and is summarised in TABLE 3.4 and Fig. 3.2.

Memory	Rate [7,13]	Lower Bound [7,13]	New Lower Bound
$1 \leq M \leq \frac{3}{2}$	$\frac{3}{2} - \frac{5}{6}M$	$R_2 \geq 1 - \frac{1}{2}M$	$R_2 \geq \frac{3}{2} - \frac{5}{6}M$

Table 3.4: Rate memory tradeoff for the (2, 4) cache network

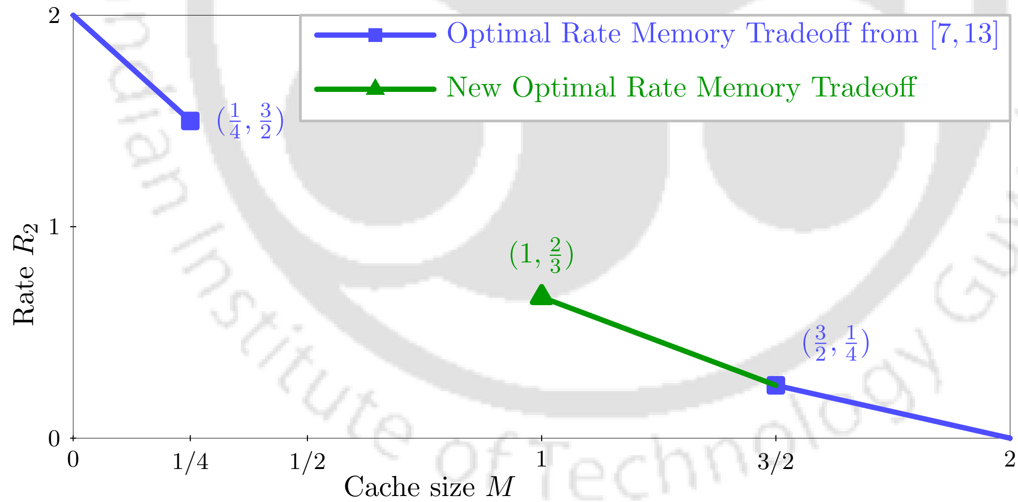

Figure 3.2: Optimal rate memory tradeoff for the (2, 4) cache network

3.3 Case I: $\lceil \frac{K+1}{2} \rceil \leq N \leq K$

In this section we propose a new symmetric caching scheme that achieves the memory rate pair

$$(M, R_N) = \left(\left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)} \right), \frac{1}{K-1} \right) \tag{3.1}$$

for the (N, K) cache network. This scheme can be seen as a generalization of the scheme presented for $(3, 4)$ cache network in Section 3.2.1 and is an extension of the scheme we proposed in [61,62]. For $\lceil \frac{K+1}{2} \rceil \leq N \leq K$, we prove a matching lower bound to establish the optimal rate memory tradeoff when $M \geq \left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)}\right)$. Let $\mathbf{I}$ denote the indicator function and let $\mathbf{S}_k$ denote the set $[K] \setminus \{k\}$.

3.3.1 Placement and Delivery Phase

During the placement phase, the server splits every file into $2\ {}^{K}C_2$ disjoint subfiles of size $\frac{1}{K(K-1)}F$ bits. Subfiles of the file W_n are:

$$\{W_n^{ij} : i, j \in [K] \text{ and } i \neq j\}.$$

The placement phase proceeds in two stages. In the first stage, the server copy subfiles W_n^{ij} in user U_k's cache, Z_k, if $k \notin \{i, j\}$. In the second stage, functions of subfiles are computed and placed into each user's cache resulting in the cache of the k^{th} user, Z_k, having the contents:

Stages	Packets	Constraints	Number
Stage 1	W_n^{ij}	$n \in [N]$, $i, j \in \mathbf{S}_k$ and $i \neq j$	$2N\ {}^{K-1}C_2$
Stage 2	$W_n^{k(k+1)} - W_n^{kj}$	$n \in [N]$ and $j \in \mathbf{S}_k$	$N(K-2)$
	$\sum_{n=1}^{N} W_n^{k(k+1)}$		1

It can be noted that subfiles W_n^{kj}, for $n \in [N]$ and $j \in \mathbf{S}_k$, are contained in user U_k's cache in coded form. The total number of packets, each of size $\frac{1}{K(K-1)}F$ bits, placed in each user's cache is,

$$2N\ {}^{K-1}C_2 + N(K-2) + 1 = NK(K-2) + 1$$

utilising the entire cache of size $\left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)}\right) F$ bits.

In the delivery phase, let the server receive a demand $\boldsymbol{d}$. Let N_k^i represent the number of users in the set $\mathbf{S}_k$ requesting the file W_{d_i}. For each $k \in [K]$, the server constructs a packet,

$$X_{\boldsymbol{d}}^k = \sum_{s \in \mathbf{S}_k} \left(\frac{\alpha_k^s}{N_k^s}\right) W_{d_s}^{ks} \tag{3.2}$$

where

$$\alpha_k^s = 1 - 2\mathbf{I}\{W_{d_k} = W_{d_s}\}. \tag{3.3}$$

The set of packets broadcast by the server in response to the demand $\boldsymbol{d}$ is,

$$X_{\boldsymbol{d}} = \{X_{\boldsymbol{d}}^1, \ldots, X_{\boldsymbol{d}}^K\}. \tag{3.4}$$

Thus, K packets, each of size $\frac{1}{K(K-1)}F$ bits, are transmitted and the rate corresponding to the demand $\boldsymbol{d}$ is,

$$R_N = \frac{1}{K-1}. \tag{3.5}$$

3.3.2 File Recovery by Users

To understand how the requested files are recovered by the users, let us consider user U_k who needs to recover the file W_{d_k} from its cache contents Z_k and the received packets $X_{\boldsymbol{d}}$. Subfiles $W_{d_k}^{ij}$, for $i, j \in \mathbf{S}_k$, are available in Z_k. To reconstruct the file W_{d_k}, the user needs to compute subfiles $W_{d_k}^{jk}$, and $W_{d_k}^{kj}$, for $j \in \mathbf{S}_k$. The user obtains these subfiles in two stages. In the first stage, the user obtains subfiles $W_{d_k}^{jk}$, for $j \in \mathbf{S}_k$. One of the packet available in $X_{\boldsymbol{d}}$ is,

$$X_{\boldsymbol{d}}^j = \left(\frac{\alpha_j^k}{N_j^k}\right) W_{d_k}^{jk} + \sum_{s \in \mathbf{S}_j \setminus \{k\}} \left(\frac{\alpha_j^s}{N_j^s}\right) W_{d_s}^{js}. \tag{3.6}$$

Since subfiles W_n^{js}, for $s \in \mathbf{S}_j \setminus \{k\} = \mathbf{S}_k \setminus \{j\}$, are available in Z_k, the user can evaluate

$$\sum_{s \in \mathbf{S}_j \setminus \{k\}} \left(\frac{\alpha_j^s}{N_j^s}\right) W_{d_s}^{js} \tag{3.7}$$

The subfile $W_{d_k}^{jk}$ can computed from (3.6) and (3.7). In the second stage, the user recovers subfiles $W_{d_k}^{kj}$, for $j \in \mathbf{S}_k$. Another packet available in $X_{\boldsymbol{d}}$ is,

$$X_{\boldsymbol{d}}^k = \sum_{j \in \mathbf{S}_k} \left(\frac{\alpha_k^j}{N_k^j}\right) W_{d_j}^{kj} \tag{3.8}$$

Since $W_n^{k(k+1)} - W_n^{kj}$, for $j \in \mathbf{S}_k$, are available in Z_k, the user can evaluate

$$\sum_{j \in \mathbf{S}_k} \left(\frac{\alpha_k^j}{N_k^j}\right) (W_{d_j}^{k(k+1)} - W_{d_j}^{kj}) \tag{3.9}$$

Combining (3.8) and (3.9) the user can compute

$$\sum_{j \in \mathbf{S}_k} \left(\frac{\alpha_k^j}{N_k^j} \right) W_{d_j}^{k(k+1)} \tag{3.10}$$

This can be rewritten as

$$\left(\sum_{n \in [N] \setminus d_k} \sum_{j \in \mathbf{S}_k : d_j = n} \left(\frac{\alpha_k^j}{N_k^j} \right) W_{d_j}^{k(k+1)} \right) + \left(\sum_{j \in \mathbf{S}_k : d_j = d_k} \left(\frac{\alpha_k^j}{N_k^j} \right) W_{d_j}^{k(k+1)} \right) \tag{3.11}$$

Note that when $W_{d_j} \neq W_{d_k}$, $\alpha_k^j = 1$ and when $W_{d_j} = W_{d_k}$, $\alpha_k^j = -1$. Recall that N_k^k denotes the number of users request for file W_{d_k} in set $\mathbf{S}_k$. Now the above expression simplifies to

$$\left(\sum_{n \in [N] \setminus d_k} W_n^{k(k+1)} \right) - I\{N_k^k \neq 0\} W_{d_k}^{k(k+1)} \tag{3.12}$$

With the help of the cached function

$$\sum_{n \in [N]} W_n^{k(k+1)} = \left(\sum_{n \in [N] \setminus \{d_k\}} W_n^{k(k+1)} \right) + W_{d_k}^{k(k+1)} \tag{3.13}$$

and (3.12) the user can compute the subfile $W_{d_k}^{k(k+1)}$. Combining this with $W_{d_k}^{k(k+1)} - W_{d_k}^{kj}$ available in Z_k user can obtain subfiles $W_{d_k}^{kj}$. Using all the recovered subfiles the user can reconstruct the requested file W_{d_k}.

The above observations can be summarised as:

Theorem 4. *The memory rate pair* $\left(\left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)} \right), \frac{1}{K-1} \right)$ *is achievable by symmetric caching schemes for the* (N, K) *cache network.*

The caching scheme proposed in [7, 13] achieves the memory rate pair $\left(\frac{N}{K}(K-1), \frac{1}{K} \right)$, and by memory sharing between that scheme and the proposed scheme, we can achieve all memory rate pairs $\left(M, \frac{(KN-1)}{K(N-1)} - \frac{1}{(N-1)} M \right)$, where $M \in \left[\left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)} \right), \frac{N}{K}(K-1) \right]$. By deriving a matching lower bound, we show that this is the optimal rate memory tradeoff when $\lceil \frac{K+1}{2} \rceil \leq N \leq K$.

3.3.3 Matching Lower Bound

Consider the demand

$$\boldsymbol{d}_1 = (W_1, W_2, \ldots, W_N, W_1, W_2, \ldots, W_{K-N}). \tag{3.14}$$

	Demands	U_1	$\dots$	U_i	$\dots$	U_N	U_{N+1}	$\dots$	U_{N+i}	$\dots$	U_K
	$\boldsymbol{d}_1$	W_1	$\dots$	W_i	$\dots$	W_N	W_1	$\dots$	W_i	$\dots$	W_{K-N}
	$\boldsymbol{d}_2$	W_2	$\dots$	W_{i+1}	$\dots$	W_1	W_2	$\dots$	W_{i+1}	$\dots$	W_1
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{A}_i$	$\boldsymbol{d}_{K-N-i+1}$	$W_{K-N-i+1}$	$\dots$	W_{K-N}	$\dots$	W_{K-N-i}	$W_{K-N-i+1}$	$\dots$	W_{K-N}	$\dots$	W_{K-N-i}
	$\boldsymbol{d}_{K-N-i+2}$	$W_{K-N-i+2}$	$\dots$	W_{K-N+1}	$\dots$	$W_{K-N-i+1}$	$W_{K-N-i+2}$	$\dots$	W_1	$\dots$	$W_{K-N-i-1}$
$\boldsymbol{C}_i$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	$\boldsymbol{d}_{N-i}$	W_{N-i}	$\dots$	W_{N-1}	$\dots$	$W_{2N-K-i-1}$	W_{2N-K-i}	$\dots$	W_{2N-K-1}	$\dots$	W_{N-i-1}
	$\boldsymbol{d}_{N-i+1}$	W_{N-i+1}	$\dots$	W_N	$\dots$	W_{2N-K-i}	$W_{2N-K-i+1}$	$\dots$	W_{2N-K}	$\dots$	W_{N-i}
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	$\boldsymbol{d}_N$	W_N	$\dots$	W_{i-1}	$\dots$	W_{2N-K}	W_{2N-K+1}	$\dots$	W_{2N-K+i}	$\dots$	W_{N-1}
	$\boldsymbol{d}_{N+1}$	W_1	$\dots$	W_i	$\dots$	W_{2N-K+1}	W_{2N-K+2}	$\dots$	$W_{2N-K+i+1}$	$\dots$	W_N
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{J}$	$\boldsymbol{d}_{K-i+1}$	$W_{K-N-i+1}$	$\dots$	W_{K-N}	$\dots$	W_{N-i}	W_{N-i+1}	$\dots$	W_N	$\dots$	W_{K-N-i}
	$\boldsymbol{d}_{K-i+2}$	$W_{K-N-i+2}$	$\dots$	W_1	$\dots$	W_{N-i+1}	W_{N-i+2}	$\dots$	W_1	$\dots$	$W_{K-N-i+1}$
$\boldsymbol{B}_i$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	$\boldsymbol{d}_K$	W_{K-N}	$\dots$	W_{i-1}	$\dots$	W_{N-1}	W_N	$\dots$	W_{i-1}	$\dots$	W_{K-N-1}

Table 3.5: The set of demands $\{\boldsymbol{d}_l : 1 \leq l \leq K\}$

Demands $\{\boldsymbol{d}_l : 2 \leq l \leq K\}$, are obtained from the demand $\boldsymbol{d}_1$ by cyclic left shifts as shown in TABLE 3.5. Consider the demand $\boldsymbol{b}_l$ defined as

$$\boldsymbol{b}_l = \begin{cases} \boldsymbol{d}_{N-l+1}, \text{ for } 1 \leq l \leq N \\ \boldsymbol{d}_{K+N-l+1}, \text{ for } N+1 \leq l \leq K \end{cases}$$

It can be noted that in demand $\boldsymbol{b}_l$, the user U_l requests for the file W_N. Let $X_{\boldsymbol{d}_l}$ denote the set of packets broadcast by the server in response to the demand $\boldsymbol{d}_l$. For $\boldsymbol{S} \subseteq \{\boldsymbol{d}_1, \dots, \boldsymbol{d}_K\}$, let $X_{\boldsymbol{S}}$ denote the set of all packets broadcast in response to the demands in the set $\boldsymbol{S}$. The following lemma are easy to obtain:

Lemma 11. *For $\boldsymbol{S}$, $\boldsymbol{T} \subset \{\boldsymbol{d}_1, \dots, \boldsymbol{d}_K\} \setminus \{\boldsymbol{b}_l\}$, we have the identity,*

$$H(W_{[N-1]}, X_{\boldsymbol{S} \cup \{\boldsymbol{b}_l\}}) + H(W_{[N-1]}, Z_l, X_{\boldsymbol{T}}) \geq H(W_{[N-1]}, X_{\boldsymbol{S} \cap \boldsymbol{T}}) + N$$

Proof. We have,

$$\begin{aligned} H(W_{[N-1]}, X_{\boldsymbol{S}\cup\{\boldsymbol{b}_l\}}) + H(W_{[N-1]}, Z_l, X_{\boldsymbol{T}}) &\overset{(a)}{\geq} H(W_{[N-1]}, X_{\boldsymbol{S}\cap\boldsymbol{T}}) + H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}\cup\boldsymbol{T}\cup\{\boldsymbol{b}_l\}}) \\ &\overset{(b)}{=} H(W_{[N-1]}, X_{\boldsymbol{S}\cap\boldsymbol{T}}) + H(W_{[N-1]}, W_N, Z_l, X_{\boldsymbol{S}\cup\boldsymbol{T}\cup\{\boldsymbol{b}_l\}}) \\ &\overset{(c)}{=} H(W_{[N-1]}, X_{\boldsymbol{S}\cap\boldsymbol{T}}) + H(W_{[N]}) \\ &= H(W_{[N-1]}, X_{\boldsymbol{S}\cap\boldsymbol{T}}) + N \end{aligned}$$

where

(a) follows from the submodularity property of entropy,

(b) follows from (1.8),

(c) follow from (1.9).

□

Lemma 12. *For a sequence of sets $\boldsymbol{S}_i \subset \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_K\} \setminus \{\boldsymbol{b}_i\}$, such that $\boldsymbol{S}_i = \boldsymbol{S}_{i+1} \cup \{\boldsymbol{b}_{i+1}\}$, we have the identity*

$$H(W_{[N-1]}, X_{\boldsymbol{S}_l}) + \sum_{i=l+1}^{j} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \geq H(W_{[N-1]}, X_{\boldsymbol{S}_j}) + (j-l)N$$

Proof. We have,

$$\begin{aligned} H(W_{[N-1]}, X_{\boldsymbol{S}_l}) + \sum_{i=l+1}^{j} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) &= H(W_{[N-1]}, X_{\boldsymbol{S}_l}) + H(W_{[N-1]}, Z_{l+1}, X_{\boldsymbol{S}_{l+1}}) \\ &\quad + \sum_{i=l+2}^{j} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \\ &= \Big(H(W_{[N-1]}, X_{\boldsymbol{S}_{l+1}}, X_{\boldsymbol{b}_{l+1}}) + H(W_{[N-1]}, Z_{l+1}, X_{\boldsymbol{S}_{l+1}})\Big) + \sum_{i=l+2}^{j} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \\ &\overset{(a)}{\geq} H(W_{[N]}) + \Big(H(W_{[N-1]}, X_{\boldsymbol{S}_{l+1}}) + H(W_{[N-1]}, Z_{l+2}, X_{\boldsymbol{S}_{l+2}})\Big) + \sum_{i=l+3}^{j} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \\ &\overset{(a)}{\geq} 2H(W_{[N]}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{l+2}}) + H(W_{[N-1]}, Z_{l+3}, X_{\boldsymbol{S}_{l+3}}) + \sum_{i=l+4}^{j} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \\ &\overset{(b)}{\geq} (j-l)H(W_{[N]}) + H(W_{[N-1]}, X_{\boldsymbol{S}_j}) \\ &= H(W_{[N-1]}, X_{\boldsymbol{S}_j}) + (j-l)N \end{aligned}$$

where

(a) follows from Lemma 11 with $\boldsymbol{S} = \boldsymbol{T} = \boldsymbol{S}_{l+1}$,

(b) follows from repeated use of Lemma 11 with $\boldsymbol{S} \cup \{\boldsymbol{b}_l\} = \boldsymbol{S}_i$ and $\boldsymbol{T} = \boldsymbol{S}_{i+1}$ for $l+3 \leq i \leq j$.

□

In a similar fashion, for a sequence of sets $\boldsymbol{T}_i \subset \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_K\} \setminus \{\boldsymbol{b}_i\}$, such that $\boldsymbol{T}_i = \boldsymbol{T}_{i-1} \cup \{\boldsymbol{b}_{i-1}\}$, we can obtain

$$H(W_{[N-1]}, X_{\boldsymbol{T}_{j+1}}) + \sum_{i=l}^{j} H(W_{[N-1]}, Z_i, X_{\boldsymbol{T}_i}) \geq H(W_{[N-1]}, X_{\boldsymbol{T}_l}) + (j-l+1)N \tag{3.15}$$

For $1 \leq i \leq N$, let us consider the sets of demands as shown below:

Set	Demands	Number	Files Requested by U_i
$\boldsymbol{A}_i$	$\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{N-i}$	$N-i$	$W_i, \ldots, W_{(N-1)}$
$\boldsymbol{B}_i$	$\boldsymbol{d}_{K-i+2}, \ldots, \boldsymbol{d}_K$	$i-1$	$W_1, \ldots, W_{i-1}$
$\boldsymbol{C}_i$	$\boldsymbol{d}_{K-N-i+2}, \ldots, \boldsymbol{d}_{N-i}$	$2N-K-1$	$W_{(K-N+1)}, \ldots, W_{N-1}$
$\boldsymbol{J}$	$\boldsymbol{d}_{N+1}, \ldots, \boldsymbol{d}_K$	$K-N$	$W_1, \ldots, W_{K-N}$

These set are also indicated in TABLE 3.5. Note that

$$\boldsymbol{A}_N = \boldsymbol{B}_1 = \boldsymbol{C}_N = \phi \tag{3.16}$$

$$\boldsymbol{A}_{i+1} \cup \{\boldsymbol{b}_{i+1}\} = \boldsymbol{A}_i \tag{3.17}$$

$$\boldsymbol{B}_i \cup \{\boldsymbol{b}_{N+i}\} = \boldsymbol{B}_{i+1} \tag{3.18}$$

$$\boldsymbol{A}_i \cap \boldsymbol{C}_i = \boldsymbol{C}_i \tag{3.19}$$

$$\boldsymbol{B}_{K-N} \cup \{\boldsymbol{b}_K\} = \boldsymbol{B}_{K-N+1} = \boldsymbol{J} \tag{3.20}$$

$$\boldsymbol{B}_i \cap \boldsymbol{J} = \begin{cases} \boldsymbol{B}_i \text{ when } 1 \leq i \leq K-N \\ \boldsymbol{J} \text{ when } K-N+1 \leq i \leq N \end{cases} \tag{3.21}$$

It can be noted that in the demand set $\boldsymbol{B}_i$, both users U_i and U_{N+i} are requesting for the same set of files (for $1 \leq i \leq K-N$). Note that $\mid \boldsymbol{A}_i \cup \boldsymbol{B}_i \mid = \mid \boldsymbol{J} \cup \boldsymbol{C}_i \mid = N-1$. Thus, we have

$$M + (N-1)R_N \geq H(Z_i) + H(X_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}) \geq H(Z_i, X_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}) \tag{3.22}$$

Similarly,

$$M + (N-1)R_N \geq H(Z_i) + H(X_{\boldsymbol{J}\cup \boldsymbol{C}_i}) \geq H(Z_i, X_{\boldsymbol{J}\cup \boldsymbol{C}_i}) \tag{3.23}$$

Now we can obtain the following result:

Theorem 5. *For the* (N, K) *cache network, when* $\lceil \frac{K+1}{2} \rceil \leq N \leq K$*, achievable memory rate pairs* (M, R_N) *must satisfy the constraint*

$$KM + K(N-1)R_N \geq KN - 1.$$

Proof. We have,

$$\begin{aligned}
KM+K(N-1)R_N &= N(M+(N-1)R_N) + (K-N)(M+(N-1)R_N) \\
&\overset{(a)}{\geq} \sum_{i=1}^{N} H(Z_i, X_{\boldsymbol{A}_i\cup \boldsymbol{B}_i}) + \sum_{i=1}^{K-N} H(Z_i, X_{\boldsymbol{J}\cup \boldsymbol{C}_i}) \\
&\overset{(b)}{=} \sum_{i=1}^{N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{A}_i\cup \boldsymbol{B}_i}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{J}\cup \boldsymbol{C}_i}) \\
&= \sum_{i=1}^{K-N} \Big(H(W_{[N-1]}, Z_i, X_{\boldsymbol{A}_i\cup \boldsymbol{B}_i}) + H(W_{[N-1]}, Z_i, X_{\boldsymbol{J}\cup \boldsymbol{C}_i}) \Big) + \sum_{j=K-N+1}^{N} H(W_{[N-1]}, Z_j, X_{\boldsymbol{A}_j\cup \boldsymbol{B}_j}) \\
&\overset{(c)}{\geq} \sum_{i=1}^{K-N} \Big(H(W_{[N-1]}, Z_i, X_{\boldsymbol{A}_i\cup \boldsymbol{J}}) + H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i\cup \boldsymbol{C}_i}) \Big) + \sum_{j=K-N+1}^{N} H(W_{[N-1]}, Z_j, X_{\boldsymbol{B}_j\cup \boldsymbol{A}_j}) \\
&\geq \Big(H(W_{[N-1]}, X_{\boldsymbol{A}_1\cup \boldsymbol{J}}) + \sum_{i=2}^{K-N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{A}_i\cup \boldsymbol{J}}) \Big) + \sum_{j=K-N+1}^{N} H(W_{[N-1]}, Z_j, X_{\boldsymbol{B}_j\cup \boldsymbol{A}_j}) \\
&\quad + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i\cup \boldsymbol{C}_i}) \\
&\overset{(d)}{\geq} (K-N-1)N + \Big(H(W_{[N-1]}, X_{\boldsymbol{A}_{K-N}\cup \boldsymbol{J}}) + \sum_{j=K-N+1}^{N} H(W_{[N-1]}, Z_j, X_{\boldsymbol{B}_j\cup \boldsymbol{A}_j}) \Big) \\
&\quad + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i\cup \boldsymbol{C}_i}) \\
&\overset{(e)}{\geq} (N-1)N + H(W_{[N-1]}, X_{\boldsymbol{A}_N\cup \boldsymbol{J}}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i\cup \boldsymbol{C}_i}) \\
&\geq (N-1)N + H(W_{[N-1]}, X_{\boldsymbol{A}_N\cup \boldsymbol{J}}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i})
\end{aligned}$$

$$
\begin{aligned}
&\overset{(f)}{=}(N-1)N + H(W_{[N-1]}, X_{\boldsymbol{J}}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i}) \\
&\overset{(g)}{=}(N-1)N + H(W_{[N-1]}, X_{\boldsymbol{J}}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_{i+N}, X_{\boldsymbol{B}_i}) \\
&\overset{(h)}{=}(N-1)N + \left(H(W_{[N-1]}, X_{\boldsymbol{B}_{K-N+1}}) + \sum_{i=1}^{K-N} H(W_{[N-1]}, Z_{i+N}, X_{\boldsymbol{B}_i}) \right) \\
&\overset{(i)}{\geq}(N-1)N + (K-N)N + H(W_{[N-1]}, X_{\boldsymbol{B}_1}) \\
&\overset{(f)}{=}(K-1)N + H(W_{[N-1]}) \geq KN - 1
\end{aligned}
$$

where

- (a) follows from (3.22) and (3.23),
- (b) follows from (1.8) and the definition of sets $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, $\boldsymbol{C}_i$ and $\boldsymbol{J}$,
- (c) follows from the facts that $\boldsymbol{A}_i \cap \boldsymbol{C}_i = \boldsymbol{C}_i$, $\boldsymbol{J} \cap \boldsymbol{B}_i = \boldsymbol{B}_i$ for $1 \leq i \leq K-N$ (refer (3.19) and (3.21)) and the submodularity property of entropy,
- (d) follows from Lemma 12 with $\boldsymbol{S}_i = \boldsymbol{A}_i \cup \boldsymbol{J}$, $l=1$, $j = K-N$ and (3.17),
- (e) follows from Lemma 12, with $\boldsymbol{S}_i = \boldsymbol{A}_i \cup \boldsymbol{J}$, $l = K-N$, $j = N$, the fact that $\boldsymbol{J} \cap \boldsymbol{B}_i = \boldsymbol{J}$ for $K-N+1 \leq i \leq K$(refer (3.21)) and (3.17),
- (f) follows from (3.16),
- (g) follows from (1.13),
- (h) follows from (3.20),
- (i) follows from (3.15) with $\boldsymbol{T}_i = \boldsymbol{B}_i$, $l=1$, $j = K-N$ and (3.18).

□

The above observations improve upon the previous results from [7, 13] as shown in TABLE 3.6.

Memory	Rate [7, 13]	Lower Bound [7, 13]	New Rate	New Lower Bound
$\left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)}\right) \leq M \leq \frac{N(K-1)}{K}$	$\frac{(K^2+K-2)}{K(K-1)} - \frac{(K+1)M}{N(K-1)}$	$R_N \geq 1 - \frac{M}{N}$	$\frac{(KN-1)}{K(N-1)} - \frac{M}{(N-1)}$	$R_N \geq \frac{(KN-1)}{K(N-1)} - \frac{M}{(N-1)}$

Table 3.6: Rate memory tradeoff when $\lceil \frac{K+1}{2} \rceil \leq N \leq K$

We summarise as:

Theorem 6. *For the (N, K) cache network, when $\lceil \frac{K+1}{2} \rceil \leq N \leq K$, the optimal rate memory tradeoff is given by*

$$R_N^*(M) = \frac{(KN-1)}{K(N-1)} - \frac{1}{(N-1)}M \tag{3.24}$$

where $M \geq \left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)}\right)$.

3.4 Case II: $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$

The caching scheme proposed by Yu et al. in [13] can achieve all memory rate pairs

$$\left(M, \frac{K^2+K-2}{K(K-1)} - \frac{(K+1)}{N(K-1)}M\right), \tag{3.25}$$

where $M \in \left[\frac{N(K-2)}{K}, \frac{N(K-1)}{K}\right]$, for the (N, K) cache network. By deriving a matching lower bound, we show that this is the optimal rate memory tradeoff when $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$. Consider the demand

$$\boldsymbol{d}_1 = (W_1, W_2, \ldots, W_N, W_1, W_2, \ldots, W_{N-1}, W_1, W_1, \ldots, W_1) \tag{3.26}$$

Demands $\{\boldsymbol{d}_l : 2 \leq l \leq K\}$, are obtained from the demand $\boldsymbol{d}_1$ by cyclic left shifts as shown in TABLE 3.7. Consider the demand $\boldsymbol{b}_l$ defined as,

$$\boldsymbol{b}_l = \begin{cases} \boldsymbol{d}_{N-l+1}, \text{ for } 1 \leq l \leq N \\ \boldsymbol{d}_{K+N-l+1}, \text{ for } N+1 \leq l \leq K \end{cases}$$

It can be noted that in demand $\boldsymbol{b}_l$, the user U_l requests for the file W_N. The following lemma is easy to obtain:

Lemma 13. *Let $\boldsymbol{S}, \boldsymbol{T} \subset \{\boldsymbol{d}_1 \ldots, \boldsymbol{d}_K\} \setminus \{\boldsymbol{b}_l\}$ be such that for every demand in $\boldsymbol{T}$, user U_l requests the file W_1. We have*

$$H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}}) + \sum_{j \in \boldsymbol{T}} H(X_j) + \frac{|\boldsymbol{T}|}{N} H(Z_l) \geq H(W_{[N-1]}, Z_l, X_{\boldsymbol{S} \cup \boldsymbol{T}}) + |\boldsymbol{T}|$$

Proof.

$$\begin{aligned} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}}) &+ \sum_{j \in \boldsymbol{T}} H(X_j) + \frac{|\boldsymbol{T}|}{N} H(Z_l) \\ &= H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}}) + \sum_{j \in \boldsymbol{T}} H(X_j) + |\boldsymbol{T}| H(Z_l) - \frac{|\boldsymbol{T}|(N-1)}{N} H(Z_l) \\ &\overset{(a)}{\geq} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}}) + \sum_{j \in \boldsymbol{T}} H(Z_l, X_j) - \frac{|\boldsymbol{T}|(N-1)}{N} H(Z_l) \end{aligned}$$

Set	Demand	U_1	…	U_i	…	U_N	U_{N+1}	…	U_{N+i}	…	U_{2N-1}	U_{2N}	…	U_j	…	U_K
$\boldsymbol{A}_i$	$\boldsymbol{d}_1$	W_1	…	W_i	…	W_N	W_1	…	W_i	…	W_{N-1}	W_1	…	W_1	…	W_1
$\boldsymbol{A}_i$	$\boldsymbol{d}_2$	W_2	…	W_{i+1}	…	W_1	W_2	…	W_{i+1}	…	W_1	W_1	…	W_1	…	W_1
$\boldsymbol{A}_i$	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{A}_i$	$\boldsymbol{d}_{N-i}$	W_{N-i}	…	W_{N-1}	…	W_{N-i-1}	W_{N-i}	…	W_{N-1}	…	W_1	W_1	…	W_1	…	W_{i-1}
	$\boldsymbol{d}_{N-i+1}$	W_{N-i+1}	…	W_N	…	W_{N-i}	W_{N-i+1}	…	W_1	…	W_1	W_1	…	W_1	…	W_i
	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{E}_i$	$\boldsymbol{d}_{N+1}$	W_1	…	W_i	…	W_1	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_N
$\boldsymbol{E}_i$	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{E}_i$	$\boldsymbol{d}_{2N-i}$	W_{N-i}	…	W_{N-1}	…	W_1	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_{N-i-1}
$\boldsymbol{G}_i$	$\boldsymbol{d}_{2N-i+1}$	W_{N-i+1}	…	W_1	…	W_1	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_{N-i}
$\boldsymbol{G}_i$	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{G}_i$	$\boldsymbol{d}_{K-j+2}$	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_1
$\boldsymbol{G}_i$	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{G}_i$, $\boldsymbol{P}_j$	$\boldsymbol{d}_{K+N-j+1}$	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_1	W_1	…	W_N	…	W_1
$\boldsymbol{G}_i$, $\boldsymbol{P}_j$	$\boldsymbol{d}_{K+N-j+2}$	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_1
$\boldsymbol{G}_i$, $\boldsymbol{P}_j$	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{G}_i$, $\boldsymbol{P}_j$	$\boldsymbol{d}_{K+2N-j}$	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_1	W_1	…	W_{N-1}	…	W_1
$\boldsymbol{G}_i$, $\boldsymbol{Q}_j$	$\boldsymbol{d}_{K+2N-j+1}$	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_1	W_1	…	W_1	…	W_1
$\boldsymbol{G}_i$, $\boldsymbol{Q}_j$	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{G}_i$, $\boldsymbol{Q}_j$	$\boldsymbol{d}_{K-i+1}$	W_1	…	W_1	…	W_{N-i}	W_{N-i+1}	…	W_N	…	W_{N-i}	W_{N-i+1}	…	W_1	…	W_1
$\boldsymbol{B}_i$, $\boldsymbol{Q}_j$	$\boldsymbol{d}_{K-i+2}$	W_1	…	W_1	…	W_{N-i+1}	W_{N-i+2}	…	W_1	…	W_{N-i+1}	W_{N-i+2}	…	W_1	…	W_1
$\boldsymbol{B}_i$, $\boldsymbol{Q}_j$	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
$\boldsymbol{B}_i$, $\boldsymbol{Q}_j$	$\boldsymbol{d}_K$	W_1	…	W_{i-1}	…	W_{N-1}	W_N	…	W_{i-1}	…	W_{N-2}	W_{N-1}	…	W_1	…	W_1

Table 3.7: Demand set $\{\boldsymbol{d}_l : 1 \leq l \leq K\}$

$$
\begin{aligned}
&\overset{(b)}{=} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}}) + \sum_{j \in \boldsymbol{T}} H(W_1, Z_l, X_j) - \frac{|\,\boldsymbol{T}\,|\,(N-1)}{N} H(Z_l) \\
&\overset{(a)}{\geq} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}}) + H(W_1, Z_l, X_{\boldsymbol{T}}) + (|\,\boldsymbol{T}\,| - 1) H(W_1, Z_l) - \frac{|\,\boldsymbol{T}\,|\,(N-1)}{N} H(Z_l) \\
&\overset{(a)}{\geq} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S} \cup \boldsymbol{T}}) + |\,\boldsymbol{T}\,|\, H(W_1, Z_l) - \frac{|\,\boldsymbol{T}\,|\,(N-1)}{N} H(Z_l) \\
&\overset{(c)}{=} H(W_{[N-1]}, Z_j, X_{\boldsymbol{S} \cup \boldsymbol{T}}) + \frac{|\,\boldsymbol{T}\,|}{N} \left(\sum_{i=1}^{N} H(W_i, Z_l) \right) - \frac{|\,\boldsymbol{T}\,|\,(N-1)}{N} H(Z_l) \\
&\overset{(a)}{\geq} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S} \cup \boldsymbol{T}}) + \frac{|\,\boldsymbol{T}\,|}{N} \Big(H(W_{[N]}, Z_l) \Big) + \left(\frac{|\,\boldsymbol{T}\,|\,(N-1)}{N} H(Z_l) \right) - \frac{|\,\boldsymbol{T}\,|\,(N-1)}{N} H(Z_l) \\
&\overset{(d)}{=} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S} \cup \boldsymbol{T}}) + \frac{|\,\boldsymbol{T}\,|}{N} H(W_{[N]}) = H(W_{[N-1]}, Z_l, X_{\boldsymbol{S} \cup \boldsymbol{T}}) + |\,\boldsymbol{T}\,|
\end{aligned}
$$

where

(a) follows from submodularity property of entropy,

(b) follows from (1.8),

(c) follows from (1.15),

(d) follows from (1.9).

□

For $1 \leq i \leq N$, let us consider the set of demands as shown below:

Set	Demands	Number	Files Requested by U_i
$\boldsymbol{A}_i$	$\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{N-i}$	$N-i$	$W_i, \ldots, W_{N-1}$
$\boldsymbol{B}_i$	$\boldsymbol{d}_{K-i+2}, \ldots, \boldsymbol{d}_K$	$i-1$	$W_1, \ldots, W_{i-1}$
$\boldsymbol{E}_i$	$\boldsymbol{d}_{N+1}, \ldots, \boldsymbol{d}_{2N-i}$	$N-i$	$W_i, \ldots, W_{N-1}$
$\boldsymbol{G}_i$	$\boldsymbol{d}_{2N-i+1}, \ldots, \boldsymbol{d}_{K-i+1}$	$K-2N+1$	W_1

These set are also indicated in TABLE 3.7. We also have a set of demands $\boldsymbol{L}_i$ defined as

$$\boldsymbol{L}_i = \boldsymbol{A}_i \cup \boldsymbol{B}_i \cup \boldsymbol{E}_i \cup \boldsymbol{G}_i. \tag{3.27}$$

Note that

$$\boldsymbol{A}_N = \boldsymbol{B}_1 = \boldsymbol{E}_N = \phi \tag{3.28}$$

$$\boldsymbol{L}_{i+1} \cup \{\boldsymbol{b}_{i+1}\} = \boldsymbol{L}_i \tag{3.29}$$

$$\boldsymbol{B}_i \cup \{\boldsymbol{b}_{N+i}\} = \boldsymbol{B}_{i+1} \tag{3.30}$$

It can be noted that in the demand set $\boldsymbol{B}_i$, both users U_i and U_{N+i} are requesting for the same set of files (for $1 \leq i \leq N-1$). Note that $|\ \boldsymbol{A}_i \cup \boldsymbol{B}_i\ |=|\ \boldsymbol{B}_i \cup \boldsymbol{E}_i\ |= N-1$. Thus, we have

$$M + (N-1)R_N \geq H(Z_i) + H(X_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}) \geq H(Z_i, X_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}) \tag{3.31}$$

Similarly,

$$M + (N-1)R_N \geq H(Z_i) + H(X_{\boldsymbol{B}_i \cup \boldsymbol{E}_i}) \geq H(Z_i, X_{\boldsymbol{B}_i \cup \boldsymbol{E}_i}) \tag{3.32}$$

The following lemma is easy to obtain:

Lemma 14. *The demand sets* $\boldsymbol{B}_i$ *and* $\boldsymbol{L}_i$*, defined as above, satisfy*

$$KM + (KN - 2N + 1)R_N \geq H(W_{[N-1]}, X_{\boldsymbol{L}_N}) + \sum_{i=1}^{N-1} \left(H(W_{[N-1]}, Z_{N+i}, X_{\boldsymbol{B}_i}) \right) + N(K - N)$$

Proof. We have,

$$\begin{aligned}
KM + (KN - 2N + 1)R_N &= N \left(M + (N-1)R_N + \frac{(K-2N+1)}{N} M + (K - 2N + 1)R_N \right) \\
&\quad + (N-1)(M + (N-1)R_N) \\
&\overset{(a)}{\geq} \sum_{i=1}^{N} \left(H(Z_i, X_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}) + \frac{(K-2N+1)}{N} H(Z_i) + \sum_{l \in \boldsymbol{G}_i} H(X_l) \right) + \sum_{i=1}^{N-1} (H(Z_i, X_{\boldsymbol{B}_i \cup \boldsymbol{E}_i})) \\
&\overset{(b)}{=} \sum_{i=1}^{N} \left(H(W_{[N-1]}, Z_i, X_{\boldsymbol{A}_i \cup \boldsymbol{B}_i}) + \frac{(K-2N+1)}{N} H(Z_i) + \sum_{l \in \boldsymbol{G}_i} H(X_l) \right) \\
&\quad + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i \cup \boldsymbol{E}_i}) \\
&\overset{(c)}{\geq} \sum_{i=1}^{N} \left(H(W_{[N-1]}, Z_i, X_{\boldsymbol{A}_i \cup \boldsymbol{B}_i \cup \boldsymbol{G}_i}) + (K - 2N + 1) \right) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i \cup \boldsymbol{E}_i}) \\
&= H(W_{[N-1]}, Z_N, X_{\boldsymbol{A}_N \cup \boldsymbol{B}_N \cup \boldsymbol{G}_N}) + \sum_{i=1}^{N-1} \left(H(W_{[N-1]}, Z_i, X_{\boldsymbol{A}_i \cup \boldsymbol{B}_i \cup \boldsymbol{G}_i}) + H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i \cup \boldsymbol{E}_i}) \right) \\
&\quad + N(K - 2N + 1) \\
&\overset{(d)}{\geq} H(W_{[N-1]}, Z_N, X_{\boldsymbol{A}_N \cup \boldsymbol{B}_N \cup \boldsymbol{G}_N \cup \boldsymbol{E}_N}) + \sum_{i=1}^{N-1} \left(H(W_{[N-1]}, Z_i, X_{\boldsymbol{A}_i \cup \boldsymbol{B}_i \cup \boldsymbol{E}_i \cup \boldsymbol{G}_i}) + H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i}) \right) \\
&\quad + N(K - 2N + 1) \\
&\overset{(e)}{=} \sum_{i=1}^{N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{L}_i}) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i}) + N(K - 2N + 1) \\
&\geq \left(H(W_{[N-1]}, X_{\boldsymbol{L}_1}) + \sum_{i=2}^{N} H(W_{[N-1]}, Z_i, X_{\boldsymbol{L}_i}) \right) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i}) + N(K - 2N + 1) \\
&\overset{(f)}{\geq} (N-1)N + H(W_{[N-1]}, X_{\boldsymbol{L}_N}) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_i, X_{\boldsymbol{B}_i}) + N(K - 2N + 1) \\
&\overset{(g)}{=} H(W_{[N-1]}, X_{\boldsymbol{L}_N}) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{N+i}, X_{\boldsymbol{B}_i}) + N(K - N)
\end{aligned}$$

where

(a) follows from (3.31) and (3.32),

(b) follows from (1.8) and definition of sets $\boldsymbol{A}_i$, $\boldsymbol{B}_i$ and $\boldsymbol{E}_i$,

(c) follows from Lemma 13 with $\boldsymbol{S} = \boldsymbol{A}_i \cup \boldsymbol{B}_i$ and $\boldsymbol{T} = \boldsymbol{G}_i$,

(d) follows from the submodularity property of entropy and the fact that $\boldsymbol{E}_N = \phi$,

(e) follows from (3.27),

(f) follows from Lemma 12 with $\boldsymbol{S}_i = \boldsymbol{L}_i$, $l = 1$, $j = N$ and (3.29),

(g) follows from (1.13).

□

Now, for $2N \leq j \leq K$, consider another set of demands as shown below:

Set	Demands	Number	Files Requested by U_j
$\boldsymbol{P}_j$	$\boldsymbol{d}_{K+N-i+2}, \ldots, \boldsymbol{d}_{K+2N-j}$	$N-1$	$W_1, \ldots, W_{N-1}$
$\boldsymbol{Q}_j$	$\boldsymbol{d}_{K+2N-j+1}, \ldots, \boldsymbol{d}_K$	$j-2N$	W_1

These set are also indicated in TABLE 3.7. We also have a set of demands $\boldsymbol{T}_j$ defined as

$$\boldsymbol{T}_j = \boldsymbol{P}_j \cup \boldsymbol{Q}_j \tag{3.33}$$

Note that

$$\boldsymbol{Q}_{2N} = \phi \tag{3.34}$$

$$\boldsymbol{T}_{j+1} \cup \{\boldsymbol{b}_{j+1}\} = \boldsymbol{T}_j \tag{3.35}$$

$$\boldsymbol{T}_K \cup \{\boldsymbol{b}_K\} = \boldsymbol{L}_N \tag{3.36}$$

$$\boldsymbol{B}_{N-1} \cup \{\boldsymbol{b}_{2N-1}\} = \boldsymbol{T}_{2N} \tag{3.37}$$

Note that $\mid \boldsymbol{P}_j \mid= N-1$. Thus, we have

$$M + (N-1)R_N \geq H(Z_j) + H(X_{\boldsymbol{P}_j}) \geq H(Z_j, X_{\boldsymbol{P}_j}) \tag{3.38}$$

The following lemma is easy to obtain:

Lemma 15. *The demand set* $\boldsymbol{T}_j$*, defined as above, satisfy*

$$\frac{K(K-2N+1)}{2N}M + \frac{(K-2)(K-2N+1)}{2}R_N \geq \sum_{j=2N}^{K} H(W_{[N-1]}, Z_j, X_{\boldsymbol{T}_j}) + \frac{(K-2N+1)(K-2N)}{2}$$

Proof.

$$\begin{aligned}
&\frac{K(K-2N+1)}{2N}M + \frac{(K-2)(K-2N+1)}{2}R_N \\
&=(K-2N+1)\left(M+(N-1)R_N\right) + \sum_{j=2N}^{K}\left(\frac{(j-2N)}{N}M + (j-2N)R_N\right) \\
&\overset{(a)}{\geq} \sum_{j=2N}^{K}\left(H(Z_j, X_{\boldsymbol{P}_j}) + \frac{(j-2N)}{N}H(Z_j) + \sum_{l\in\boldsymbol{Q}_j} H(X_l)\right) \\
&\overset{(b)}{=} \sum_{j=2N}^{K}\left(H(W_{[N-1]}, Z_j, X_{\boldsymbol{P}_j}) + \frac{(j-2N)}{N}H(Z_j) + \sum_{l\in\boldsymbol{Q}_j} H(X_l)\right) \\
&\overset{(c)}{\geq} \sum_{j=2N}^{K}\left(H(W_{[N-1]}, Z_j, X_{\boldsymbol{P}_j\cup\boldsymbol{Q}_j}) + \frac{(j-2N)}{N}H(W_{[N]})\right) \\
&\overset{(d)}{=} \sum_{j=2N}^{K} H(W_{[N-1]}, Z_j, X_{\boldsymbol{T}_j}) + \frac{(K-2N+1)(K-2N)}{2}
\end{aligned}$$

where

- (a) follows from (3.38) and the fact that $\boldsymbol{Q}_{2N} = \phi$,
- (b) follows from (1.8) and definition of set $\boldsymbol{P}_j$,
- (c) follows from Lemma 13 with $\boldsymbol{S} = \boldsymbol{P}_j$ and $\boldsymbol{T} = \boldsymbol{Q}_j$,
- (d) follows from (3.33).

□

Using the above lemma, we can obtain the following result:

Theorem 7. *For the (N, K) cache network, when $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$, achievable memory rate pairs (M, R_N) must satisfy the constraint*

$$\frac{K(K+1)}{N}M + K(K-1)R_N \geq K^2 + K - 2$$

Proof. We have,

$$\begin{aligned}
\frac{K(K+1)}{2N}M + \frac{K(K-1)}{2}R_N &= KM + (KN-2N+1)R_N + \frac{K(K-2N+1)}{2N}M + \frac{(K-2)(K-2N+1)}{2}R_N \\
&\overset{(a)}{\geq} H(W_{[N-1]}, X_{\boldsymbol{L}_N}) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{N+i}, X_{\boldsymbol{B}_i}) + N(K-N) + \sum_{j=2N}^{K} H(W_{[N-1]}, Z_j, X_{\boldsymbol{T}_j}) \\
&\quad + \frac{(K-2N+1)(K-2N)}{2}
\end{aligned}$$

$$\overset{(b)}{=} \left(H(W_{[N-1]}, X_{\boldsymbol{T}_K}, X_{\boldsymbol{b}_K}) + \sum_{j=2N}^{K} H(W_{[N-1]}, Z_j, X_{\boldsymbol{T}_j}) \right) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{N+i}, X_{\boldsymbol{B}_i})$$
$$+ \frac{(K+1)(K-2N)+2N^2}{2}$$
$$\overset{(c)}{\geq} (K-2N+1)N + \left(H(W_{[N-1]}, X_{\boldsymbol{T}_{2N}}) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{N+i}, X_{\boldsymbol{B}_i}) \right)$$
$$+ \frac{(K+1)(K-2N)+2N^2}{2}$$
$$\overset{(d)}{=} \left(H(W_{[N-1]}, X_{\boldsymbol{B}_{N-1}}, X_{\boldsymbol{b}_{2N-1}}) + \sum_{i=1}^{N-1} H(W_{[N-1]}, Z_{N+i}, X_{\boldsymbol{B}_i}) \right)$$
$$+ \frac{(K+1+2N)(K-2N)+2N(N+1)}{2}$$
$$\overset{(e)}{\geq} (N-1)N + H(W_{[N-1]}, X_{\boldsymbol{B}_1}) + \frac{(K+1+2N)(K-2N)+2N(N+1)}{2}$$
$$\overset{(f)}{=} H(W_{[N-1]}) + \frac{K^2+K-2N}{2} \geq \frac{K^2+K-2}{2}$$

where

- (a) follows from Lemma 14 and Lemma 15,
- (b) follows from (3.36),
- (c) follows from (3.15) with $\boldsymbol{T}_i = \boldsymbol{T}_j$, $l = 2N$, $j = K$ and (3.35),
- (d) follows from (3.37),
- (e) follows from (3.15) with $\boldsymbol{T}_i = \boldsymbol{B}_i$, $l = 1$, $j = N-1$ and (3.30),
- (f) follows from (3.28).

□

The above observations improve upon the previous results from [7, 13] as shown in TABLE 3.8.

Memory	Rate [7, 13]	Lower Bound [7, 13]	New Lower Bound
$\frac{N(K-2)}{K} \leq M \leq \frac{N(K-1)}{K}$	$\frac{K^2+K-2}{K(K-1)} - \frac{(K+1)}{N(K-1)}M$	$R_N \geq 1 - \frac{1}{N}M$	$R_N \geq \frac{K^2+K-2}{K(K-1)} - \frac{(K+1)}{N(K-1)}M$

Table 3.8: Rate memory tradeoff when $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$

We summarise as:

Theorem 8. *For the (N, K) cache network, when $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$, the optimal rate memory tradeoff is given by*

$$R_N^*(M) = \frac{K^2+K-2}{K(K-1)} - \frac{(K+1)}{N(K-1)}M \tag{3.39}$$

where $M \geq \frac{N}{K}(K-2)$.

3.5 Conclusions

In this chapter, we considered the problem of characterizing the optimal rate memory tradeoff for the canonical (N, K) cache network, where we focused on the case of large caches. For $\lceil \frac{K+1}{2} \rceil \leq N \leq K$, in Section 3.3, we proposed a new coded caching scheme and derived a matching lower bound leading to the characterization of the optimal rate memory tradeoff when $M \geq \left(\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)} \right)$. For $1 \leq N \leq \lceil \frac{K+1}{2} \rceil$, in Section 3.4, we derived a new lower bound matching the scheme proposed in [13], thereby providing a characterization of the optimal rate memory tradeoff when $M \geq \frac{N}{K}(K-2)$.

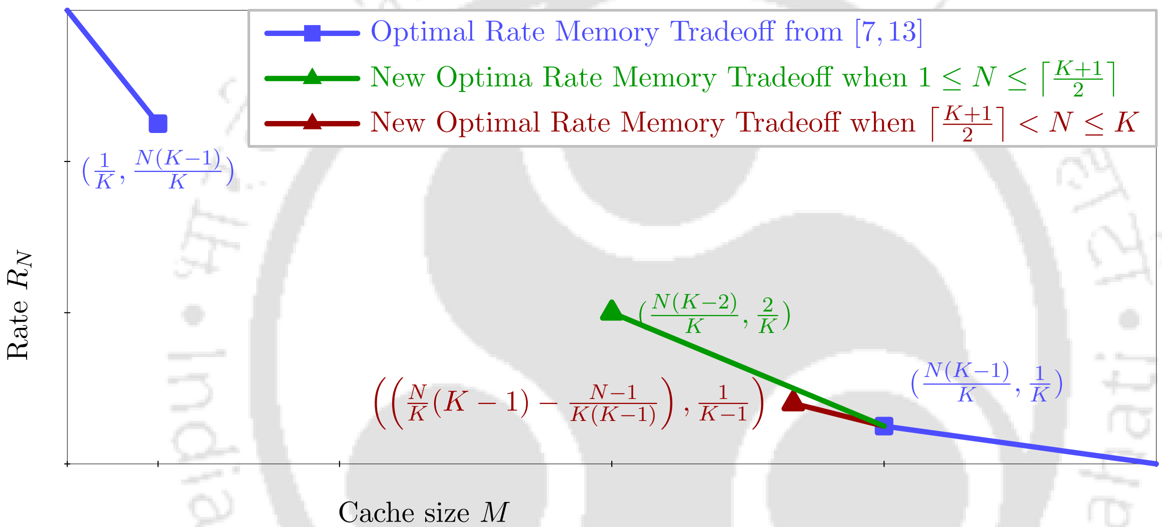

Figure 3.3: Optimal rate memory tradeoff for the (N, K) cache network

4

Pareto Optial Schemes in Coded Caching

Contents

4.1 Introduction

In the previous chapters, we studied the demand set D_N and investigated schemes that minimized the rate R_N required in the delivery phase, for small and large caches. In this chapter, we focus on the multi-objective nature of the problem of coded caching where each rate R_p corresponding to demand set $\boldsymbol{D}_p$, $1 \leq p \leq N$, is of interest. A natural question is whether a placement strategy can be devised which simultaneously minimizes all the rates of interest. In [13], Yu, Maddah-Ali and Avestimehr considered the collection of coded caching schemes where the placement phase is restricted to be uncoded and obtained the optimal rate memory tradeoff for each of the rates R_p. In a surprising result, they demonstrated the existence of a universal scheme among them, referred to as the YMA scheme in this chapter, which was shown attain the optimal rate memory tradeoff for each of the rates R_p, simultaneously. Thus, for coded caching schemes where the placement phase is restricted to be uncoded, the Pareto optimal frontier of the rates R_p, $1 \leq p \leq N$, consist of a single point and the YMA scheme operates at this point.

We first investigate the existence of such universal schemes in the more general setting where coding is permitted in both the placement phase and the delivery phase and prove that such schemes do not exist. To this end, we introduce new lower bounds which jointly constrain the rates R_p, $1 \leq p \leq N$, and provide better insight into how the performance for one demand set affects the performance for other demand sets. Using such bounds we identify two coded caching schemes, the CFL scheme [14] and the YMA scheme, that operate at the pareto optimal frontier of the problem of coded caching. We begin the chapter by considering the example of the $(2, 2)$ cache network for which we show that there are no universal schemes and that the CFL scheme and the YMA scheme are Pareto optimal. We then formally define Pareto optimal schemes for the (N, K) cache network and extend results obtained for the $(2, 2)$ cache network to the (N, K) cache network. Results for the $(2, 2)$ network is presented in section 4.2 and the (N, K) network is considered in section 4.3, and section 4.4. The Pareto optimal property of the YMA scheme is investigated in sections 4.5 and 4.6. We conclude the chapter in section 4.7.

4.2 The $(2, 2)$ Cache Network

Here, users $\{U_1, U_2\}$ are connected to a server with two files $\{A, B\}$, each of size F bits, and each user is equipped with an isolated cache of size MF bits, where $M \in [0, 2]$. The demand set $\boldsymbol{D}_1$

represents the set of all demands where both users request for the same file and the demand set $\boldsymbol{D}_2$ represents the set of all demands where both users request for different files. Let R_1 and R_2 denote the rates corresponding to $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, respectively. For the demand set $\boldsymbol{D}_1$, achievable memory rate pairs (M, R_1) must satisfy the following constraint [8],

$$M + 2R_1 \geq 2.$$

As shown in Figure 4.1, this constraint is tight for the YMA scheme and thus characterize the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_1$. For the demand set $\boldsymbol{D}_2$, achievable memory rate pairs

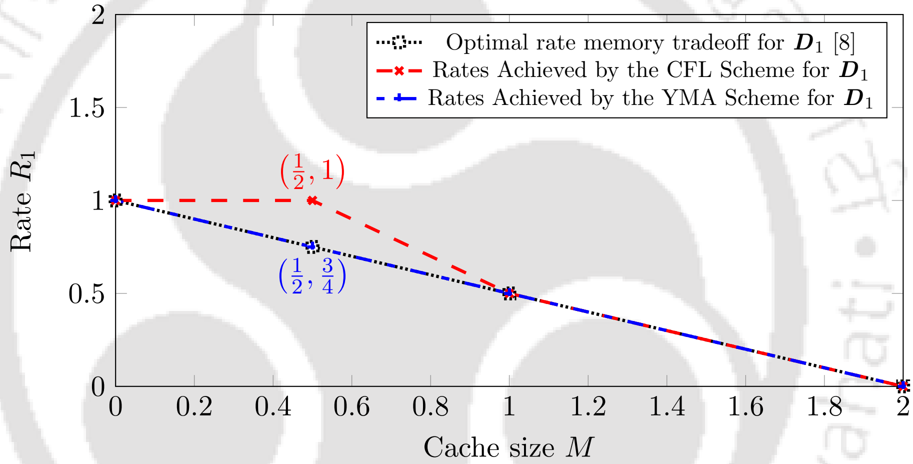

Figure 4.1: Rate memory tradeoff for demand set $\boldsymbol{D}_1$.

(M, R_2) must satisfy the following constraints [7],

$$M + 2R_2 \geq 2, \;\; 2M + R_2 \geq 2, \;\; 2M + 2R_2 \geq 3.$$

As shown in Figure 4.2, these constraints are tight for the CFL scheme and thus characterize the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_2$. Note that the CFL scheme meets the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_2$, but does not meet the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_1$. Similarly, the YMA scheme meets the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_1$, but does not meet the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_2$. This leads us to the question as to whether there exist a universal scheme that is simultaneously optimal for both the demands sets. To answer this we obtain a new lower bound that jointly constrains the rates achievable for these demand sets.

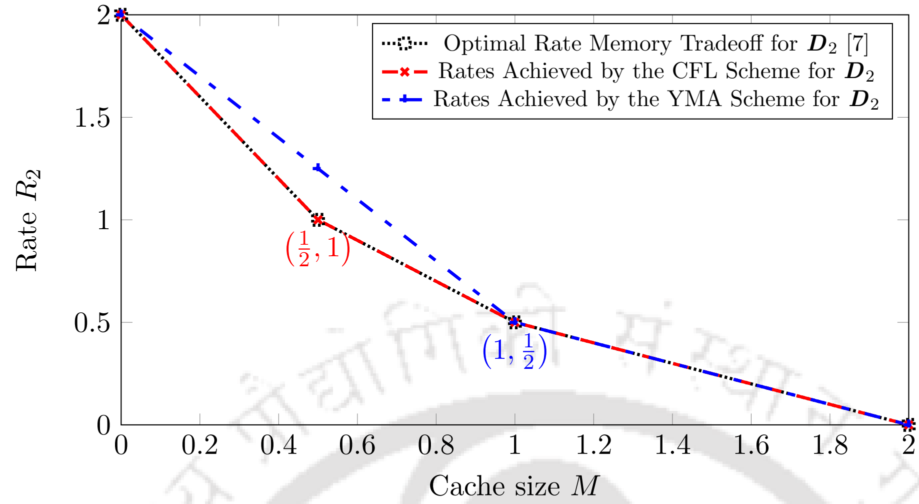

Figure 4.2: Rate memory tradeoff for demand set $\boldsymbol{D}_2$.

Lemma 16. *For the* $(2,2)$ *cache network with cache size* M*, achievable rates* R_1 *and* R_2 *must satisfy the following constraints:*

$$2M + R_1 + R_2 \geq 3,$$
$$M + R_1 + R_2 \geq 2.$$

Proof. We have,

$$\begin{aligned}
2M + R_1 + R_2 &\geq H(Z_1) + H(Z_2) + H(X_{(A,A)}) + H(X_{(B,A)}) \\
&\overset{(a)}{\geq} H(Z_1, X_{(A,A)}) + H(Z_2, X_{(B,A)}) \\
&\overset{(b)}{=} H(A, Z_1, X_{(A,A)}) + H(A, Z_2, X_{(B,A)}) \\
&\overset{(a)}{\geq} H(A) + H(A, Z_2, Z_1, X_{(B,A)}, X_{(A,A)}) \\
&\overset{(b)}{=} H(A) + H(A, B, Z_2, Z_1, X_{(B,A)}, X_{(A,A)}) \\
&\overset{(c)}{=} H(A) + H(A, B) = 3.
\end{aligned}$$

We also have,

$$\begin{aligned}
M + R_2 + R_1 &\geq H(Z_1) + H(X_{(A,A)}) + H(X_{(B,A)}) \\
&\overset{(a)}{\geq} H(Z_1, X_{(A,A)}, X_{(B,A)})
\end{aligned}$$

$$\overset{(b)}{=} H(A, B, Z_1, X_{(A,A)}, X_{(B,A)})$$
$$\overset{(c)}{=} H(A, B) = 2,$$

where

(a) follows from the submodularity of entropy,

(b) follows from (1.8),

(c) follows from (1.9). □

The optimal rate memory tradeoffs $R_1^*(M)$ and $R_2^*(M)$ and a joint rate constraint are shown in Table 4.1. Any scheme that is optimal for $\boldsymbol{D}_1$ and achieves $R_1 = R_1^*(M) = 1 - \frac{1}{2}M$ must satisfy the

Cache Size	Optimal Rate $R_1^*(M)$	Optimal Rate $R_2^*(M)$	Rate Constraint
$0 \leq M \leq \frac{1}{2}$	$1 - \frac{1}{2}M$	$2 - 2M$	$2M + R_1 + R_2 \geq 3$

Table 4.1: Optimal rate memory tradeoff for the $(2,2)$ cache network

constraint $2M + R_1 + R_2 \geq 3$ and will necessarily have $R_2(M) \geq 2 - \frac{3}{2}M > 2 - 2M = R_2^*(M)$. Thus no scheme can be simultaneously optimal for both the demand sets. We summarize as:

Theorem 9. *For the $(2,2)$ cache network, for $M \leq \frac{1}{2}$, there exist no universal caching scheme which achieves the optimal rates for all the demand sets simultaneously.*

Since there are no universal schemes in the general setting where coding is permitted in both the placement phase and the delivery phase, finding schemes that operate at various points on the Pareto optimal frontier is of interest. The performance of the CFL scheme is summarized in Table 4.2 where it can be noted that the scheme is tight with respect to the joint constraint on the rates. Thus no scheme

Cache Size	R_1	R_2	Constraint
$1 \leq M \leq \frac{1}{2}$	1	$2 - 2M$	$2M + R_1 + R_2 = 3$

Table 4.2: Performance of the CFL scheme

can have a smaller rate R_1 while being optimal for the demand set $\boldsymbol{D}_2$ and any attempt at improving the rate R_1 will necessarily involve a degradation of the performance for rate R_2. We conclude that the CFL scheme is Pareto optimal.

Theorem 10. *For the $(2,2)$ cache network, for $0 \leq M \leq \frac{1}{2}$, the CFL scheme is Pareto optimal.*

The performance of the YMA scheme is summarized in Table 4.3 where it can be noted that the scheme is tight with respect to the joint constraint on the rates. Thus no scheme can have a smaller

Cache Size	R_1	R_2	Constraints
$1 \leq M \leq 1$	$1 - \frac{1}{2}M$	$2 - \frac{3}{2}M$	$2M + R_1 + R_2 = 3$
$1 \leq M \leq 2$	$1 - \frac{1}{2}M$	$1 - \frac{1}{2}M$	$M + R_1 + R_2 = 2$

Table 4.3: Performance of the YMA scheme

rate R_2 while being optimal for the demand set $\boldsymbol{D}_1$ and any attempt at improving the rate R_2 will necessarily involve a degradation of the performance for rate R_1. We conclude that the CFL scheme is Pareto optimal.

Theorem 11. *For the* $(2, 2)$ *cache network, for* $0 \leq M \leq 2$*, the YMA scheme is Pareto optimal.*

4.3 Non-existence of Universal Schemes

In this section, we consider the (N, K) cache network, where the number of files is at most the number of users, i.e., $N \leq K$. Here users $\{U_1, \ldots, U_K\}$ are connected to a server with N files $\{W_1, \ldots, W_N\}$, each of size F bits, and each user is equipped with an isolated cache of size MF bits, where $M \in [0, N]$. Note that the demand set $\boldsymbol{D}_p$ denotes the set of demands where users request p distinct files and R_p denotes the corresponding rate. For the demand set $\boldsymbol{D}_1$, achievable memory rate pairs (M, R_1) must satisfy the following constraint [8],

$$M + NR_1 \geq N \tag{4.1}$$

This constraint is tight for the YMA scheme and thus characterize the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_1$, when $0 \leq M \leq N$. For the demand set $\boldsymbol{D}_N$, achievable memory rate pairs (M, R_N) must satisfy the following constraint [7],

$$NM + R_N \geq N \tag{4.2}$$

This constraint is tight for the CFL scheme and thus characterize the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_N$, when $0 \leq M \leq \frac{1}{K}$. From Table 4.4, we can note that the CFL scheme meets the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_N$, but does not meet the optimal rate memory

Rates	The CFL Scheme	The YMA Scheme	Optimal Rate
R_1	1	$1-\frac{1}{N}M$	$1-\frac{1}{N}M$
R_N	$N-NM$	$N-\frac{(N+1)}{2}M$	$N-NM$

Table 4.4: Rates achieved for $\boldsymbol{D}_1$ and $\boldsymbol{D}_N$ when $0 \leq M \leq \frac{1}{K}$

tradeoff for the demand set $\boldsymbol{D}_1$. Similarly, the YMA scheme meets the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_1$, but does not meet the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_N$. This leads us to the question as to whether there exist a universal scheme that is simultaneously optimal for both the demands sets. To answer this we obtain a new lower bound that jointly constrains the rates achievable for these demand sets.

Consider a demand $\boldsymbol{d}_p \in \boldsymbol{D}_p$, where $p \leq (N-1)$ distinct files are requested. Without loss of generality, we assume that in demand $\boldsymbol{d}_p$, the first p users are requesting for files $\{W_1, \ldots, W_p\}$. In response to this demand $\boldsymbol{d}_p$, the server broadcasts a packet $X_{\boldsymbol{d}_p}$ of size R_p. Now consider demands $\boldsymbol{b}_i \in \boldsymbol{D}_N$ as shown below, where the user i requests the file W_N. In response to the demand $\boldsymbol{b}_i$, the server

Demand	U_1	U_2	...	U_i	U_{i+1}	...	U_N	U_{N+1}	...	U_K
$\boldsymbol{b}_i$	W_{N-i+1}	W_{N-i+2}	...	W_N	W_1	...	W_{N-i}	W_1	...	W_1

broadcasts a packet $X_{\boldsymbol{b}_i}$ of size R_N. We use the notation $\overline{Z_k}$ to represent the set $\{Z_1 \ldots, Z_N\} \setminus \{Z_k\}$, where $k \in [N]$. The following lemma is easy to obtain:

Lemma 17. *For the (N, K) cache network, all coded caching schemes satisfy the constraint*

$$H(W_{[p]}, Z_{[p]\setminus[k]}) + H(W_{[N-1]}, \overline{Z_{k+1}}, X_{\boldsymbol{b}_{k+1}}) \geq H(W_{[N]}) + H(W_{[p]}, Z_{[p]\setminus[k+1]}),$$

where $k+1 \leq p \leq N$.

Proof. We have,

$$\begin{aligned}
&H(W_{[p]}, Z_{[p]\setminus[k]}) + H(W_{[N-1]}, \overline{Z_{k+1}}, X_{\boldsymbol{b}_{k+1}}) \\
&\qquad = H(W_{[p]}, Z_{[p]\setminus[k+1]}, Z_{k+1}) + H(W_{[N-1]}, \overline{Z_{k+1}}, X_{\boldsymbol{b}_{k+1}}) \\
&\qquad \overset{(a)}{\geq} H(W_{[N-1]}, W_{[p]}, \overline{Z_{k+1}}, Z_{[p]\setminus[k+1]}, Z_{k+1}, X_{\boldsymbol{b}_{k+1}}) + H(W_{[p]}, Z_{[p]\setminus[k+1]}) \\
&\qquad \overset{(b)}{=} H(W_{[N-1]}, W_N, Z_{[N]}, X_{\boldsymbol{b}_{k+1}}) + H(W_{[p]}, Z_{[p]\setminus[k+1]})
\end{aligned}$$

$$\overset{(c)}{=} H(W_{[N]}) + H(W_{[p]}, Z_{[p]\setminus[k+1]})$$

where

(a) follows from submodularity of entropy,

(b) follows (1.8) and the definition of demand $\boldsymbol{b}_i$,

(c) follows from (1.9).

□

Now we have the following result:

Theorem 12. *For the (N, K) cache network with cache size M, achievable rates R_N and R_p, where $1 \leq p \leq N-1$, satisfy the constraint*

$$pNM + pR_N + R_p \geq p(N+1)$$

Proof. We have:

$$\begin{aligned}
pNM + R_p + pR_N &= pM + R_p + p(N-1)M + pR_N \\
&\geq H(Z_{[p]}) + H(X_{\boldsymbol{d}_p}) + \sum_{i=1}^{p} H(\overline{Z_i}) + \sum_{i=1}^{p} H(X_{\boldsymbol{b}_i}) \\
&\overset{(a)}{\geq} H(Z_{[p]}, X_{\boldsymbol{d}_p}) + \sum_{i=1}^{p} H(\overline{Z_i}, X_{\boldsymbol{b}_i}) \\
&\overset{(b)}{=} H(W_{[p]}, Z_{[p]}, X_{\boldsymbol{d}_p}) + \sum_{i=1}^{p} H(W_{[N-1]}, \overline{Z_i}, X_{\boldsymbol{b}_i}) \\
&\geq H(W_{[p]}, Z_{[p]}) + H(W_{[N-1]}, \overline{Z_1}, X_{\boldsymbol{b}_1}) + \sum_{i=2}^{p} H(W_{[N-1]}, \overline{Z_i}, X_{\boldsymbol{b}_i}) \\
&\overset{(c)}{\geq} H(W_{[N]}) + H(W_{[p]}, Z_{[p]\setminus\{1\}}) + H(W_{[N-1]}, \overline{Z_2}, X_{\boldsymbol{b}_2}) + \sum_{i=3}^{p} H(W_{[N-1]}, \overline{Z_i}, X_{\boldsymbol{b}_i}) \\
&\overset{(c)}{\geq} 2H(W_{[N]}) + H(W_{[p]}, Z_{[p]\setminus[2]}) + H(W_{[N-1]}, \overline{Z_3}, X_{\boldsymbol{b}_3}) + \sum_{i=4}^{p} H(W_{[N-1]}, \overline{Z_i}, X_{\boldsymbol{b}_i}) \\
&\overset{(d)}{\geq} (p)H(W_{[N]}) + H(W_{[p]}, Z_{[p]\setminus[p]}) \\
&= pH(W_{[N]}) + H(W_{[p]}) = p(N+1)
\end{aligned}$$

where

(a) follows from submodularity of entropy,

(b) follows from (1.8),

(c) follows from Lemma 17,

(d) follows from repeated use of Lemma 17.

□

The optimal rate memory tradeoffs $R_1^*(M)$ and $R_N^*(M)$ and the joint rate constraint obtained above are shown in Table 4.5. Any scheme that is optimal for $\boldsymbol{D}_1$ and achieves $R_1 = R_1^*(M) = 1 - \frac{1}{N}M$ must

Cache Size	Optimal Rate $R_1^*(M)$	Optimal Rate $R_N^*(M)$	Constraint
$0 \leq M \leq \frac{1}{K}$	$1 - \frac{1}{N}M$	$N - NM$	$NM + R_N + R_1 \geq N + 1$

Table 4.5: Optimal rate memory tradeoff for the (N, K) cache network

satisfy the constraint $NM + R_N + R_1 \geq N + 1$ and will necessarily have $R_N(M) \geq N - \frac{N^2-1}{N}M > N - NM = R_N^*(M)$. Thus no scheme can be simultaneously optimal for both the demand sets. We summarize as:

Theorem 13. *For the (N, K) cache network, for $M \leq \frac{1}{K}$, there exist no universal caching scheme which achieves the optimal rates for all the demand sets simultaneously.*

4.4 Pareto Optimal schemes

In the previous section we saw that in the general setting where coding is permitted in the placement phase and the delivery phase, universal codes that achieve the optimal rate memory tradeoff for all the demand sets, simultaneously, do not exist. This leads us to investigate whether there exist coded caching schemes that operate at the Pareto optimal frontier of the problem where any improvement for one demand set has to necessarily come at the expense of another, as seen for the $(2, 2)$ cache network. The notion of such Pareto optimal schemes is formalized in the following:

Definition 1. *For the (N, K) cache network with cache size MF bits, a caching scheme $\mathcal{X}$ is said to dominate another caching scheme $\mathcal{X}^*$ if the scheme $\mathcal{X}$ achieves the same or better performance as scheme $\mathcal{X}^*$ for all demand sets and achieves strictly better performance for at least one demand set. If no scheme dominates a particular scheme $\mathcal{X}$, the scheme $\mathcal{X}$ is said to be Pareto optimal.*

For the $(2,2)$ cache network, it was seen that the CFL scheme is Pareto optimal for $M \leq \frac{1}{2}$. The performance of the CFL scheme for the (N, K) cache network is summarized in Table 4.6 where it can be noted that the scheme is tight with respect to the joint constraint on the rates obtained in Theorem 12. Thus no scheme can have a smaller rate R_p while being optimal for the demand set $\boldsymbol{D}_N$

Cache Size	R_p	R_N	Constraint
$0 \leq M \leq \frac{1}{K}$	p	$N - NM$	$pNM + pR_N + R_p = p(N+1)$

Table 4.6: Performance of the CFL scheme

and any attempt at improving the rate R_p will necessarily involve a degradation of the performance for rate R_N. We conclude that the CFL scheme is Pareto optimal.

Theorem 14. *For the (N, K) cache network, for $0 \leq M \leq \frac{1}{K}$, the CFL scheme is Pareto optimal.*

4.5 The YMA scheme: $(3,3)$ Cache Network

As noted in Theorem 11, in the general setting where coding is permitted in the placement phase and the delivery phase, the YMA scheme is Pareto optimal for the $(2,2)$ cache network, for all cache sizes. This result is particularly interesting as the placement phase for the YMA scheme is uncoded and it is universal among such schemes. As a prelude to investigating the Pareto optimal nature of the YMA scheme for (N, K) cache network, we consider the example of the $(3,3)$ cache network in this section. Here, users $\{U_1, U_2, U_3\}$ are connected to a server with files $\{A, B, C\}$, each of size F bits, and each user is equipped with a cache memory of size MF bits, where $M \in [0,3]$. The network has three sets of demands $\boldsymbol{D}_1$, $\boldsymbol{D}_2$ and $\boldsymbol{D}_3$ with rate R_1, R_2 and R_3, respectively. We derive a new set of lower bounds that jointly constrain these rates for all coded caching schemes for the $(3,3)$ network.

Lemma 18. *For the $(3,3)$ cache network with cache size M, achievable rates R_1, R_2 and R_3 must satisfy the constraints:*

$$R_2 + 3R_1 + 2M \geq 5,\ R_3 + 3R_1 + 3M \geq 6$$

$$3R_2 + 3R_1 + 3M \geq 8,\ 3R_3 + 3R_1 + 3M \geq 8$$

$$2R_2 + R_1 + M \geq 3,\ 2R_3 + R_1 + M \geq 3$$

Proof. We have,

$$\begin{aligned}
R_2 + 3R_1 + 2M &\geq H(Z_1) + H(X_{(A,A,A)}) + H(X_{(B,C,C)}) + H(Z_2) + H(X_{(B,B,B)}) + H(X_{(A,A,A)}) \\
&\overset{(a)}{\geq} H(Z_1, X_{(B,C,C)}, X_{(A,A,A)}) + H(Z_2, X_{(B,B,B)}, X_{(A,A,A)}) \\
&\overset{(b)}{=} H(A, B, Z_1, X_{(B,C,C)}, X_{(A,A,A)}) + H(A, B, Z_2, X_{(B,B,B)}, X_{(A,A,A)}) \\
&\geq H(A, B, Z_1, X_{(B,C,C)}) + H(A, B, Z_2) \\
&\overset{(a)}{\geq} H(A, B, Z_1, Z_2, X_{(B,C,C)}) + H(A, B) \\
&\overset{(b)}{=} H(A, B, C, Z_1, Z_2, X_{(B,C,C)}) + H(A, B) \\
&\overset{(c)}{=} H(A, B, C) + H(A, B) = 5
\end{aligned}$$

$$\begin{aligned}
R_3 + 3R_1 + 3M &\geq H(Z_1) + H(X_{(A,B,C)}) + H(Z_2) + H(X_{(A,A,A)}) + H(Z_3) + H(X_{(A,A,A)}) \\
&\quad + H(X_{(B,B,B)}) \\
&\overset{(a)}{\geq} H(Z_1, X_{(A,B,C)}) + H(Z_2, X_{(A,A,A)}) + H(Z_3, X_{(A,A,A)}, X_{(B,B,B)}) \\
&\overset{(b)}{=} H(A, Z_1, X_{(A,B,C)}) + H(A, Z_2, X_{(A,A,A)}) + H(A, B, Z_3, X_{(A,A,A)}, X_{(B,B,B)}) \\
&\geq H(A, Z_1, X_{(A,B,C)}) + H(A, Z_2) + H(A, B, Z_3) \\
&\overset{(a)}{\geq} H(A, Z_1, Z_2, X_{(A,B,C)}) + H(A) + H(A, B, Z_3) \\
&\overset{(b)}{=} H(A, B, Z_1, Z_2, X_{(A,B,C)}) + H(A) + H(A, B, Z_3) \\
&\overset{(a)}{\geq} H(A, B, Z_1, Z_2, Z_3, X_{(A,B,C)}) + H(A) + H(A, B) \\
&\overset{(b)}{=} H(A, B, C, Z_1, Z_2, Z_3, X_{(A,B,C)}) + H(A) + H(A, B) \\
&\overset{(b)}{=} H(A, B, C) + H(A) + H(A, B) = 6
\end{aligned}$$

$$\begin{aligned}
3R_2 + 3R_1 + 3M &\geq H(Z_1) + H(X_{(A,A,C)}) + H(X_{(B,C,C)}) + H(Z_2) + H(X_{(A,A,C)}) + H(X_{(B,B,B)}) \\
&\quad + H(Z_3) + H(X_{(A,A,A)}) + H(X_{(B,B,B)}) \\
&\overset{(a)}{\geq} H(Z_1, X_{(A,A,C)}, X_{(B,C,C)}) + H(Z_2, X_{(A,A,C)}, X_{(B,B,B)}) + H(Z_3, X_{(A,A,A)}, X_{(B,B,B)}) \\
&\overset{(b)}{=} H(A, B, Z_1, X_{(A,A,C)}, X_{(B,C,C)}) + H(A, B, Z_2, X_{(A,A,C)}, X_{(B,B,B)}) \\
&\quad + H(A, B, Z_3, X_{(A,A,A)}, X_{(B,B,B)})
\end{aligned}$$

$$
\begin{aligned}
&\geq H(A,B,Z_1,X_{(A,A,C)},X_{(B,C,C)}) + H(A,B,Z_2,X_{(A,A,C)}) + H(A,B,Z_3) \\
&\overset{(a)}{\geq} H(A,B,Z_1,Z_2,X_{(A,A,C)},X_{(B,C,C)}) + H(A,B,X_{(A,A,C)}) + H(A,B,Z_3) \\
&\overset{(a)}{\geq} H(A,B,Z_1,Z_2,X_{(A,A,C)},X_{(B,C,C)}) + H(A,B,Z_3,X_{(A,A,C)}) + H(A,B) \\
&\overset{(b)}{=} H(A,B,C,Z_1,Z_2,X_{(A,A,C)},X_{(B,C,C)}) + H(A,B,C,Z_3,X_{(A,A,C)}) + H(A,B) \\
&\overset{(c)}{=} 2H(A,B,C) + H(A,B) = 8
\end{aligned}
$$

$$
\begin{aligned}
3R_3 + 3R_1 + 3M &\geq H(Z_1) + H(X_{(A,B,C)}) + H(X_{(B,C,A)}) + H(Z_2) + H(X_{(A,B,C)}) + H(X_{(B,B,B)}) \\
&+ H(Z_3) + H(X_{(A,A,A)}) + H(X_{(B,B,B)}) \\
&\overset{(a)}{\geq} H(Z_1,X_{(A,B,C)},X_{(B,C,A)}) + H(Z_2,X_{(A,B,C)},X_{(B,B,B)}) + H(Z_3,X_{(A,A,A)},X_{(B,B,B)}) \\
&\overset{(b)}{=} H(A,B,Z_1,X_{(A,B,C)},X_{(B,C,A)}) + H(A,B,Z_2,X_{(A,B,C)},X_{(B,B,B)}) \\
&+ H(A,B,Z_3,X_{(A,A,A)},X_{(B,B,B)}) \\
&\geq H(A,B,Z_1,X_{(A,B,C)},X_{(B,C,A)}) + H(A,B,Z_2,X_{(A,B,C)}) + H(A,B,Z_3) \\
&\overset{(a)}{\geq} H(A,B,Z_1,Z_2,X_{(A,B,C)},X_{(B,C,A)}) + H(A,B,X_{(A,B,C)}) + H(A,B,Z_3) \\
&\overset{(a)}{\geq} H(A,B,Z_1,Z_2,X_{(A,B,C)},X_{(B,C,A)}) + H(A,B,Z_3,X_{(A,B,C)}) + H(A,B) \\
&\overset{(b)}{=} H(A,B,C,Z_1,Z_2,X_{(A,B,C)},X_{(B,C,A)}) + H(A,B,C,Z_3,X_{(A,B,C)}) + H(A,B) \\
&\overset{(c)}{=} 2H(A,B,C) + H(A,B) = 8
\end{aligned}
$$

$$
\begin{aligned}
2R_2 + R_1 + M &\geq H(Z_1) + H(X_{(A,A,C)}) + H(X_{(B,C,C)}) + H(X_{(C,C,C)}) \\
&\overset{(a)}{\geq} H(Z_1,X_{(A,A,C)},X_{(B,C,C)},X_{(C,C,C)}) \\
&\overset{(b)}{=} H(A,B,C,Z_1,X_{(A,A,C)},X_{(B,C,C)},X_{(C,C,C)}) \\
&\overset{(c)}{=} H(A,B,C) = 3
\end{aligned}
$$

$$
\begin{aligned}
2R_3 + R_1 + M &\geq H(Z_1) + H(X_{(A,B,C)}) + H(X_{(B,C,A)}) + H(X_{(C,C,C)}) \\
&\overset{(a)}{\geq} H(Z_1,X_{(A,B,C)},X_{(B,C,A)},X_{(C,C,C)}) \\
&\overset{(b)}{=} H(A,B,C,Z_1,X_{(A,B,C)},X_{(B,C,A)},X_{(C,C,C)})
\end{aligned}
$$

$$\overset{(c)}{=} H(A, B, C) = 3$$

where

(a) follows from the submodularity property of entropy,

(b) follows from (1.8),

(c) follows from (1.9)

□

The rate memory tradeoff of the YMA scheme is summarized in Table 4.7. It can be noted that the rates achieved by the YMA scheme is tight with respect to the lower bounds presented in Lemma 18. Thus even when coding is allowed in the placement phase no code can dominate the YMA scheme,

Cache Size	$R_1(M)$	$R_2(M)$	$R_3(M)$	Constraints
$0 \leq M \leq 1$	$1 - \frac{1}{3}M$	$2 - M$	$3 - 2M$	$R_2 + R_1 + 2M = 5$, $R_3 + 3R_1 + 3M = 6$
$1 \leq M \leq 2$	$1 - \frac{1}{3}M$	$\frac{5}{3} - \frac{2}{3}M$	$\frac{5}{3} - \frac{2}{3}M$	$3R_2 + 3R_1 + 3M = 8$, $3R_3 + 3R_1 + 3M = 8$
$2 \leq M \leq 3$	$1 - \frac{1}{3}M$	$1 - \frac{1}{3}M$	$1 - \frac{1}{3}M$	$2R_2 + R_1 + M = 3$, $2R_3 + R_1 + M = 3$

Table 4.7: Performance of the YMA scheme

i.e., it is operating at the Pareto optimal frontier of the $(3, 3)$ cache network. These observations leads to the following result:

Theorem 15. *For the* $(3, 3)$ *cache network, for* $0 \leq M \leq 3$*, the YMA scheme is Pareto optimal.*

4.6 The YMA scheme: (N, K) Cache Network

Let us consider the canonical (N, K) cache network where users $\{U_1, \ldots, U_K\}$ are connected to a server with files $\{W_1, \ldots, W_N\}$, each of size F bits and each user is equipped with an isolated cache of size MF bits. The network has N sets of demands $\boldsymbol{D}_p$ with rates $R_p(M)$ for $1 \leq p \leq N$. We derive a new set of lower bounds that jointly constrain these rates for all coded caching schemes. Three cache regions are considered, namely $0 \leq M \leq \frac{N}{K}$, $\frac{N}{K}(K-2) \leq M \leq \frac{N}{K}(K-1)$ and $\frac{N}{K}(K-1) \leq M \leq N$.

4.6.1 Case I: $0 \leq M \leq \frac{N}{K}$

For $1 \leq i \leq N$, consider the demands $\boldsymbol{a}_i \in \boldsymbol{D}_1$, defined as

$$\boldsymbol{a}_i = (W_i, W_i, \ldots, W_i, W_i) \tag{4.3}$$

Let $\boldsymbol{A}_i$, with $|\boldsymbol{A}_i| = i$, be defined as

$$\boldsymbol{A}_i = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_i\} \tag{4.4}$$

We have

$$M + iR_1 \geq H(Z_l) + H(X_{\boldsymbol{A}_i}) \geq H(Z_l, X_{\boldsymbol{A}_i}) \tag{4.5}$$

Let $\boldsymbol{c}_p \in \boldsymbol{D}_p$ be defined as

$$\boldsymbol{c}_p = (W_{N-p+1}, W_{N-p+2}, \ldots, W_{N-1}, W_N, W_N, \ldots, W_N) \tag{4.6}$$

where the first p users request distinct files $\{W_{N-p+1}, \ldots, W_N\}$.

Demand	U_1	U_2	$\ldots$	U_{p-1}	U_p	U_{p+1}	$\ldots$	U_K
$\boldsymbol{c}_p$	W_{N-p+1}	W_{N-p+2}	$\ldots$	W_{N-1}	W_N	W_N	$\ldots$	W_N

Theorem 16. *For the (N, K) cache network with cache size M, achievable rates R_1 and R_p must satisfy the constraint*

$$R_p + \frac{p}{2}(2N - p - 1)R_1 + pM \geq \frac{p}{2}(2N - p + 1)$$

where $2 \leq p \leq \min(N, K)$.

Proof. We have,

$$\begin{aligned}
R_p + \frac{p}{2}(2N - p - 1)R_1 + pM &= R_p + \sum_{i=1}^{p}(M + (N - p + i - 1)R_1) \\
&\overset{(a)}{\geq} H(X_{\boldsymbol{c}_p}) + \sum_{i=1}^{p} H(Z_i, X_{\boldsymbol{A}_{N-p+i-1}}) \\
&\overset{(b)}{=} H(X_{\boldsymbol{c}_p}) + \sum_{i=1}^{p} H(W_{[N-p+i-1]}, Z_i, X_{\boldsymbol{A}_{N-p+i-1}}) \\
&= H(X_{\boldsymbol{c}_p}) + H(W_{[N-p]}, Z_1, X_{\boldsymbol{A}_{N-p}}) + \sum_{i=2}^{p} H(W_{[N-p+i-1]}, Z_i, X_{\boldsymbol{A}_{N-p+i-1}})
\end{aligned}$$

$$
\begin{aligned}
&\overset{(c)}{\geq} H(W_{[N-p]}, Z_1, X_{\boldsymbol{c}_p}, X_{\boldsymbol{A}_{N-p}}) + \sum_{i=2}^{p} H(W_{[N-p+i-1]}, Z_i, X_{\boldsymbol{A}_{N-p+i-1}}) \\
&\overset{(b)}{=} H(W_{[N-p+1]}, Z_1, X_{\boldsymbol{c}_p}, X_{\boldsymbol{A}_{N-p}}) + \sum_{i=2}^{p} H(W_{[N-p+i-1]}, Z_i, X_{\boldsymbol{A}_{N-p+i-1}}) \\
&\geq H(W_{[N-p+1]}, Z_1, X_{\boldsymbol{c}_p}) + \sum_{i=2}^{p} H(W_{[N-p+i-1]}, Z_i) \\
&= H(W_{[N-p+1]}, Z_1, X_{\boldsymbol{c}_p}) + H(W_{[N-p+1]}, Z_2) + \sum_{i=3}^{p} H(W_{[N-p+i-1]}, Z_i) \\
&\overset{(c)}{\geq} H(W_{[N-p+1]}) + H(W_{[N-p+1]}, Z_1, Z_2, X_{\boldsymbol{c}_p}) + \sum_{i=3}^{p} H(W_{[N-p+i-1]}, Z_i) \\
&\overset{(b)}{=} H(W_{[N-p+1]}) + H(W_{[N-p+2]}, Z_1, Z_2, X_{\boldsymbol{c}_p}) + H(W_{[N-p+2]}, Z_3) + \sum_{i=4}^{p} H(W_{[N-p+i-1]}, Z_i) \\
&= H(W_{[N-p+1]}) + H(W_{[N-p+2]}, Z_{[2]}, X_{\boldsymbol{c}_p}) + H(W_{[N-p+2]}, Z_3) + \sum_{i=4}^{p} H(W_{[N-p+i-1]}, Z_i) \\
&\overset{(c)}{\geq} H(W_{[N-p+1]}) + H(W_{[N-p+2]}) + H(W_{[N-p+2]}, Z_{[2]}, Z_3, X_{\boldsymbol{c}_p}) + \sum_{i=4}^{p} H(W_{[N-p+i-1]}, Z_i) \\
&\overset{(b)}{=} \sum_{i=1}^{2} H(W_{[N-p+i]}) + H(W_{[N-p+3]}, Z_{[3]}, X_{\boldsymbol{c}_p}) + \sum_{i=4}^{p} H(W_{[N-p+i-1]}, Z_i) \\
&\overset{(d)}{\geq} \sum_{i=1}^{p-2} H(W_{[N-p+i]}) + H(W_{[N-1]}, Z_{[p-1]}, X_{\boldsymbol{c}_p}) + H(W_{[N-1]}, Z_p) \\
&\overset{(c)}{\geq} \sum_{i=1}^{p-2} H(W_{[N-p+i]}) + H(W_{[N-1]}) + H(W_{[N-1]}, Z_{[p]}, X_{\boldsymbol{c}_p}) \\
&= \sum_{i=1}^{p-1} H(W_{[N-p+i]}) + H(W_{[N-1]}, Z_{[p]}, X_{\boldsymbol{c}_p}) \\
&\overset{(b)}{=} \sum_{i=1}^{p-1} H(W_{[N-p+i]}) + H(W_{[N]}, Z_{[p]}, X_{\boldsymbol{c}_p}) \\
&\overset{(e)}{=} \sum_{i=1}^{p} H(W_{[N-p+i]}) = \sum_{i=1}^{p} (N-p+i) = \frac{p}{2}(2N-p+1)
\end{aligned}
$$

where

- (a) follows from (4.5),
- (b) follows from (1.8),
- (c) follows from the submodularity property of entropy,
- (d) follows from repeated use of the submodularity property of entropy and (1.8),
- (e) follows from (1.9)

□

When cache size $M \in \left[0, \frac{N}{K}\right]$, the rate achieved by YMA scheme for any demand $\boldsymbol{d} \in \boldsymbol{D}_p$ (where $1 \leq p \leq \min(N, K)$) is given in Table 4.8. The constraint of Theorem 16 is satisfied with an equality by the scheme and any attempt at modifying the scheme to obtain better performance for R_p would entail a performance loss for R_1. Thus for $0 \leq M \leq \frac{N}{K}$, no scheme can dominate the YMA scheme.

$R_1(M)$	$R_p(M)$	Constraint
$1 - \frac{1}{N}M$	$p - \frac{p(p+1)}{2N}M$	$pM + R_p + \frac{p}{2}(2N-p-1)R_1 = \frac{p}{2}(2N-p+1)$

Table 4.8: Rate achieved by YMA scheme when $M \in \left[0, \frac{N}{K}\right]$.

4.6.2 Case II: $\frac{N}{K}(K-2) \leq M \leq \frac{N}{K}(K-1)$

We begin by considering the case where $K \leq N$. Consider the demands $\boldsymbol{q}_i$, $1 \leq i \leq K-1$, defined as below (and illustrated in Table 4.9):

$$\boldsymbol{q}_i = \begin{cases} \underbrace{W_i, W_i, \ldots, W_i}_{\text{first } (K-p-i+2) \text{ users}}, W_{K-p+2}, W_{K-p+3}, \ldots, W_K, \underbrace{\star, \star, \ldots, \star}_{\text{last } (i-1) \text{ users}} & \text{for } 1 \leq i \leq K-p+1 \\ W_i, W_{i+1}, \ldots, W_K, \underbrace{\star, \star \ldots, \star, \ldots, \star \ldots, \star, \ldots, \star, \ldots, \star}_{\text{last } (i-1) \text{ users}} & \text{for } K-p+2 \leq i \leq K-1 \end{cases}$$

where $\star$ can be any file such that $\boldsymbol{q}_i \in \boldsymbol{D}_p$. For $1 \leq i \leq K$, let us consider the sets of demands as shown below:

Set	Demands	Number of Demands	Files Requested by User U_i	
			$1 \leq i \leq K-p+1$	$K-p+2 \leq i \leq K-1$
$\boldsymbol{Q}_i$	$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_{K-i}$	$K-i$	$W_1, \ldots, W_{K-p-i+2}, W_{K-p+2}, \ldots, W_{K-1}$	$W_i, \ldots, W_{K-1}$
$\boldsymbol{B}_i$	$\boldsymbol{a}_{K-p-i+3}, \boldsymbol{a}_{K-p-i+4}, \ldots, \boldsymbol{a}_{K-p+1}$	$i-1$	$W_{K-p-i+3}, W_{K-p-i+4}, \ldots, W_{K-p+1}$	$W_{K-p-i+3}, W_{K-p-i+4}, \ldots, W_{K-p+1}$
$\boldsymbol{A}_{i-1}$	$\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{i-1}$	$i-1$	$W_1, W_2, \ldots, W_{i-1}$	$W_1, W_2, \ldots, W_{i-1}$
$\boldsymbol{C}$	$\boldsymbol{a}_{K+1}, \boldsymbol{a}_{K+2}, \ldots, \boldsymbol{a}_N$	$N-K$	$W_{K+1}, W_{K+2}, \ldots, W_N$	$W_{K+1}, W_{K+2}, \ldots, W_N$

Note that

$$\boldsymbol{A}_0 = \boldsymbol{B}_1 = \boldsymbol{Q}_K = \phi \tag{4.7}$$

$$\boldsymbol{Q}_{i+1} \cap \boldsymbol{Q}_i = \boldsymbol{Q}_{i+1} \tag{4.8}$$

Demand	U_1	U_2	$\dots$	U_{i-1}	U_i	U_{i+1}	$\dots$	U_{K-p}	U_{K-p+1}	U_{K-p+2}	$\dots$	U_{j-1}	U_j	U_{j+1}	$\dots$	U_{K-1}	U_K
$\boldsymbol{q}_1$	W_1	W_1	$\dots$	W_1	W_1	W_1	$\dots$	W_1	W_1	W_{K-p+2}	$\dots$	W_{j-1}	W_j	W_{j+1}	$\dots$	W_{K-1}	W_K
$\boldsymbol{q}_2$	W_2	W_2	$\dots$	W_2	W_2	W_2	$\dots$	W_2	W_{K-p+2}	W_{K-p+3}	$\dots$	W_j	W_{j+1}	W_{j+2}	$\dots$	W_K	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-j+1}$	W_{K-j+1}	W_{K-j+1}	$\dots$	W_{K-j+1}	W_{K-j+1}	W_{K-j+1}	$\dots$	W_{2K-p-j}	$W_{2K-p-j+1}$	$W_{2K-p-j+2}$	$\dots$	W_{K-1}	W_K	$\star$	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-p-i+2}$	$W_{K-p-i+2}$	$W_{K-p-i+2}$	$\dots$	$W_{K-p-i+2}$	$W_{K-p-i+2}$	W_{K-p+2}	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$
$\boldsymbol{q}_{K-p-i+3}$	$W_{K-p-i+3}$	$W_{K-p-i+3}$	$\dots$	$W_{K-p-i+3}$	W_{K-p+2}	W_{K-p+3}	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-p}$	W_{K-p}	W_{K-p}	$\dots$	$W_{K-p+i-2}$	$W_{K-p+i-1}$	W_{K-p+i}	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$
$\boldsymbol{q}_{K-p+1}$	W_{K-p+1}	W_{K-p+2}	$\dots$	$W_{K-p+i-1}$	W_{K-p+i}	$W_{K-p+i+1}$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$
$\boldsymbol{q}_{K-p+2}$	W_{K-p+2}	W_{K-p+3}	$\dots$	W_{K-p+i}	$W_{K-p+i+1}$	$W_{K-p+i+2}$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-i}$	W_{K-i}	W_{K-i+1}	$\dots$	W_{K-2}	W_{K-1}	W_K	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$
$\boldsymbol{q}_{K-i+1}$	W_{K-i+1}	W_{K-i+2}	$\dots$	W_{K-1}	W_K	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-1}$	W_{K-1}	W_K	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\star$	$\dots$	$\star$	$\star$

Table 4.9: The set of demands $\{\boldsymbol{q}_i : 1 \leq i \leq K-1\}$

As $|\boldsymbol{Q}_i| = K - i$, $|\boldsymbol{B}_i| = |\boldsymbol{A}_{i-1}| = i - 1$ and $|\boldsymbol{C}| = N - K$, we have

$$\begin{aligned} M + (K-i)R_p + (N-K+i-1)R_1 &\geq H(Z_i) + H(X_{\boldsymbol{Q}_i}) + H(X_{\boldsymbol{B}_i}) + H(X_{\boldsymbol{C}}) \\ &\geq H(Z_i, X_{\boldsymbol{Q}_i}, X_{\boldsymbol{B}_i}, X_{\boldsymbol{C}}) \end{aligned} \tag{4.9}$$

$$\begin{aligned} M + (K-i)R_p + (N-K+i-1)R_1 &\geq H(Z_i) + H(X_{\boldsymbol{Q}_i}) + H(X_{\boldsymbol{A}_{i-1}}) + H(X_{\boldsymbol{C}}) \\ &\geq H(Z_i, X_{\boldsymbol{Q}_i}, X_{\boldsymbol{A}_{i-1}}, X_{\boldsymbol{C}}) \end{aligned} \tag{4.10}$$

We have the following result:

Theorem 17. *For the (N, K) cache network with cache size M, achievable rates R_1 and R_p must satisfy the following constraint*

$$\frac{K(K-1)}{2}R_p + \frac{K(2N-K-1)}{2}R_1 + KM \geq KN - 1$$

where $K \leq N$ and $2 \leq p \leq N$

Proof. We have

$$\begin{aligned} &\frac{K(K-1)}{2}R_p + \frac{K(2N-K-1)}{2}R_1 + KM \\ &\quad = \sum_{i=1}^{K-p+1} (M + (K-i)R_p + (N-K+i-1)R_1) + \sum_{i=K-p+2}^{K} (M + (K-i)R_p + (N-K+i-1)R_1) \\ &\quad \overset{(a)}{\geq} \sum_{i=1}^{K-p+1} H(Z_i, X_{\boldsymbol{Q}_i}, X_{\boldsymbol{B}_i}, X_{\boldsymbol{C}}) + \sum_{i=K-p+2}^{K} H(Z_i, X_{\boldsymbol{Q}_i}, X_{\boldsymbol{A}_{i-1}}, X_{\boldsymbol{C}}) \\ &\quad \overset{(b)}{=} \sum_{i=1}^{K-p+1} H(W_{[N]\backslash\{K\}}, Z_i, X_{\boldsymbol{Q}_i}, X_{\boldsymbol{B}_i}, X_{\boldsymbol{C}}) + \sum_{i=K-p+2}^{K} H(W_{[N]\backslash\{K\}}, Z_i, X_{\boldsymbol{Q}_i}, X_{\boldsymbol{A}_{i-1}}, X_{\boldsymbol{C}}) \\ &\quad \geq \sum_{i=1}^{K-p+1} H(W_{[N]\backslash\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) + \sum_{i=K-p+2}^{K} H(W_{[N]\backslash\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) \\ &\quad = H(W_{[N]\backslash\{K\}}, Z_1, X_{\boldsymbol{Q}_1}) + H(W_{[N]\backslash\{K\}}, Z_2, X_{\boldsymbol{Q}_2}) + \sum_{i=3}^{K} H(W_{[N]\backslash\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) \\ &\quad \geq H(W_{[N]\backslash\{K\}}, X_{\boldsymbol{Q}_1}) + H(W_{[N]\backslash\{K\}}, Z_2, X_{\boldsymbol{Q}_2}) + \sum_{i=3}^{K} H(W_{[N]\backslash\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) \\ &\quad \overset{(c)}{\geq} H(W_{[N]\backslash\{K\}}, Z_2, X_{\boldsymbol{Q}_1}) + H(W_{[N]\backslash\{K\}}, X_{\boldsymbol{Q}_2}) + \sum_{i=3}^{K} H(W_{[N]\backslash\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) \end{aligned}$$

$$\begin{aligned}
&\overset{(d)}{=} H(W_{[N]\setminus\{K\}}, Z_2, W_K, X_{\boldsymbol{Q}_1}) + H(W_{[N]\setminus\{K\}}, X_{\boldsymbol{Q}_2}) + \sum_{i=3}^{K} H(W_{[N]\setminus\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) \\
&\overset{(e)}{=} H(W_{[N]}) + H(W_{[N]\setminus\{K\}}, X_{\boldsymbol{Q}_2}) + H(W_{[N]\setminus\{K\}}, Z_3, X_{\boldsymbol{Q}_3}) + \sum_{i=4}^{K} H(W_{[N]\setminus\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) \\
&\overset{(c)}{\geq} H(W_{[N]}) + H(W_{[N]\setminus\{K\}}, Z_3, X_{\boldsymbol{Q}_2}) + H(W_{[N]\setminus\{K\}}, X_{\boldsymbol{Q}_3}) + \sum_{i=4}^{K} H(W_{[N]\setminus\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) \\
&\overset{(d)}{\geq} H(W_{[N]}) + H(W_{[N]\setminus\{K\}}, W_K, Z_3, X_{\boldsymbol{Q}_2}) + H(W_{[N]\setminus\{K\}}, X_{\boldsymbol{Q}_3}) + \sum_{i=4}^{K} H(W_{[N]\setminus\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) \\
&\overset{(e)}{\geq} 2H(W_{[N]}) + H(W_{[N]\setminus\{K\}}, W_K, X_{\boldsymbol{Q}_3}) + H(W_{[N]\setminus\{K\}}, Z_4, X_{\boldsymbol{Q}_4}) + \sum_{i=5}^{K} H(W_{[N]\setminus\{K\}}, Z_i, X_{\boldsymbol{Q}_i}) \\
&\overset{(f)}{\geq} (K-1)H(W_{[N]}) + H(W_{[N]\setminus\{K\}}, X_{\boldsymbol{Q}_K}) \\
&\overset{(g)}{=} (K-1)H(W_{[N]}) + H(W_{[N]\setminus\{K\}}) = (K-1)N + N - 1 = KN - 1
\end{aligned}$$

where

- (a) follows from (4.7), (4.9), and (4.10),
- (b) follows from (1.8),
- (c) follows from the submodularity propert of entropy and (4.8),
- (d) follows from the fact that user U_{i+1} requests the file W_K in $\boldsymbol{q}_{K-i} \in \boldsymbol{Q}_i$
- (e) follows from (1.9),
- (f) follows from repeated use of (c), (d) and (e),
- (g) follows from (4.7).

□

Now, we focus on the case where $N \leq K$. Consider the demands $\boldsymbol{q}_i$, for $1 \leq i \leq K-1$, defined as below (and illustrated in Table 4.10):

$$\boldsymbol{q}_i = \begin{cases} \underbrace{W_i, W_i, \ldots, W_i}_{\text{first } (K-p-i+2) \text{ users}}, W_{N-p+2}, W_{N-p+3}, \ldots, W_N, \underbrace{\star, \star, \ldots, \star}_{\text{last } (i-1) \text{ users}} & \text{for } 1 \leq i \leq N-p+1 \\ \underbrace{W_i, W_i, \ldots, W_i}_{\text{first } (K-N) \text{ users}}, W_i, W_{i+1}, \ldots, W_N, \underbrace{\star, \star \ldots, \star, \ldots, \star \ldots, \star}_{\text{last } (i-1) \text{ users}} & \text{for } N-p+2 \leq i \leq N-1 \\ \underbrace{W_{N-1}, W_{N-1}, \ldots, W_{N-1}}_{\text{first } (K-i) \text{ users}}, W_N, \underbrace{\star, \star, \ldots, \star, \ldots, \star, \ldots, \star}_{\text{last } (i-1) \text{ users}} & \text{for } N \leq i \leq K-1 \end{cases}$$

where $\star$ can be any file such that $\boldsymbol{q}_i \in \boldsymbol{D}_p$.

Demand	U_1	$\dots$	U_l	$\dots$	U_{K-N+1}	U_{K-N+2}	$\dots$	U_i	U_{i+1}	$\dots$	U_{K-p+1}	U_{K-p+2}	$\dots$	U_j	$\dots$	U_{K-1}	U_K
$\boldsymbol{q}_1$	W_1	$\dots$	W_1	$\dots$	W_1	W_1	$\dots$	W_1	W_1	$\dots$	W_1	W_{N-p+2}	$\dots$	W_{N-K+j}	$\dots$	W_{N-1}	W_N
$\boldsymbol{q}_2$	W_2	$\dots$	W_2	$\dots$	W_2	W_2	$\dots$	W_2	W_2	$\dots$	W_{N-p+2}	W_{N-p+3}	$\dots$	$W_{N-K+j+1}$	$\dots$	W_N	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-j+1}$	W_{K-j+1}	$\dots$	W_{K-j+1}	$\dots$	W_{K-j+1}	W_{K-j+1}	$\dots$	W_{K-j+1}	W_{K-j+1}	$\dots$	$W_{N+K-j-p+1}$	$W_{N+K-j-p+2}$	$\dots$	W_N	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-p-i+2}$	$W_{K-p-i+2}$	$\dots$	$W_{K-p-i+2}$	$\dots$	$W_{K-p-i+2}$	$W_{K-p-i+2}$	$\dots$	$W_{K-p-i+2}$	W_{N-p+2}	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\boldsymbol{q}_{K-p-i+3}$	$W_{K-p-i+3}$	$\dots$	$W_{K-p-i+3}$	$\dots$	$W_{K-p-i+3}$	$W_{K-p-i+3}$	$\dots$	W_{N-p+2}	W_{N-p+3}	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{N-p}$	W_{N-p}	$\dots$	W_{N-p}	$\dots$	W_{N-p}	W_{N-p}	$\dots$	$W_{2N-K-p+i-1}$	$W_{2N-K-p+i}$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\boldsymbol{q}_{N-p+1}$	W_{N-p+1}	$\dots$	W_{N-p+1}	$\dots$	W_{N-p+1}	W_{N-p+2}	$\dots$	$W_{2N-K-p+i}$	$W_{2N-K-p+i+1}$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\boldsymbol{q}_{N-p+2}$	W_{N-p+2}	$\dots$	W_{N-p+2}	$\dots$	W_{N-p+2}	W_{N-p+3}	$\dots$	$W_{2N-K-p+i+1}$	$W_{2N-K-p+i+2}$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-i+1}$	W_{K-i+1}	$\dots$	W_{K-i+1}	$\dots$	W_{K-i+1}	W_{K-i+2}	$\dots$	W_N	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\boldsymbol{q}_{K-i+2}$	W_{K-i+2}	$\dots$	W_{K-i+2}	$\dots$	W_{K-i+2}	W_{K-i+3}	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_N$	W_{N-1}	$\dots$	W_{N-1}	$\dots$	W_N	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\boldsymbol{q}_{N+1}$	W_{N-1}	$\dots$	W_{N-1}	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-l+1}$	W_{N-1}	$\dots$	W_N	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\boldsymbol{q}_{K-1}$	W_{N-1}	$\dots$	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\star$	$\dots$	$\star$	$\dots$	$\star$	$\star$

Table 4.10: The set of demands $\{\boldsymbol{q}_i : 1 \leq i \leq K-1\}$

For $K - N + 1 \leq i \leq K$, let

Set	Demands	Number of Demands	Files Requested by User U_i	
			$K - N + 1 \leq i \leq K - p + 1$	$K - p + 2 \leq i \leq K - 1$
$\boldsymbol{S}_i$	$\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_{K-i}$	$K - i$	$W_1, \ldots, W_{K-p-i+2}, W_{N-p+2}, \ldots, W_{N-1}$	$W_{N-K+i}, \ldots, W_{N-1}$
$\boldsymbol{E}_i$	$\boldsymbol{a}_{K-p-i+3}, \boldsymbol{a}_{K-p-i+4}, \ldots, \boldsymbol{a}_{N-p+1}$	$N - K + i - 1$	$W_{K-p-i+3}, W_{K-p-i+4}, \ldots, W_{N-p+1}$	$W_{K-p-i+3}, W_{K-p-i+4}, \ldots, W_{N-p+1}$
$\boldsymbol{A}_{i-1}$	$\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{N-K+i-1}$	$N - K + i - 1$	$W_1, W_2, \ldots, W_{N-K+i-1}$	$W_1, W_2, \ldots, W_{N-K+i-1}$

For $1 \leq l \leq K - N$, let

Set	Demands	Number of Demands	Files Requested by User U_l
$\boldsymbol{G}_l$	$\boldsymbol{q}_N, \ldots, \boldsymbol{q}_{K-l}$	$K - N - l + 1$	W_{N-1}
$\boldsymbol{S}_{K-N+1}$	$\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N-1}$	$N - 1$	$W_1, \ldots, W_{N-1}$

Note that

$$\boldsymbol{E}_{K-N+1} = \boldsymbol{S}_K = \phi \tag{4.11}$$

$$\boldsymbol{G}_{l+1} \cap \boldsymbol{G}_l = \boldsymbol{G}_{l+1} \tag{4.12}$$

$$\boldsymbol{S}_{i+1} \cap \boldsymbol{S}_i = \boldsymbol{S}_{i+1} \tag{4.13}$$

$$\boldsymbol{G}_{K-N} = \boldsymbol{q}_N \tag{4.14}$$

As $|\boldsymbol{S}_i| = K - i$ and $|\boldsymbol{E}_i| = |\boldsymbol{A}_{i-1}| = i - 1$, we have

$$M + (K - i)R_p + (N - K + i - 1)R_1 \geq H(Z_i) + H(X_{\boldsymbol{S}_i}) + H(X_{\boldsymbol{E}_i}) \geq H(Z_i, X_{\boldsymbol{S}_i}, X_{\boldsymbol{E}_i}) \tag{4.15}$$

$$M + (K - i)R_p + (N - K + i - 1)R_1 \geq H(Z_i) + H(X_{\boldsymbol{S}_i}) + H(X_{\boldsymbol{A}_{i-1}}) \geq H(Z_i, X_{\boldsymbol{S}_i}, X_{\boldsymbol{A}_{i-1}}) \tag{4.16}$$

The following lemma are easy to obtain:

Lemma 19. *The demand sets* $\boldsymbol{S}_{K-N+1}$ *and* $\boldsymbol{G}_l$ *,* $1 \leq l \leq K - N$*, satisfy*

$$\left(|\boldsymbol{G}_l| + N - 1\right) R_p + \left(|\boldsymbol{G}_l|\right)(N - 1)R_1 + \left(|\boldsymbol{G}_l| + 1\right) M \geq H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) + \left(|\boldsymbol{G}_l|\right) N$$

Proof. We have,

$$\begin{aligned}
&\Big(|\boldsymbol{G}_l| + N - 1\Big) R_p + \left(|\boldsymbol{G}_l|\right)(N - 1)R_1 + \left(|\boldsymbol{G}_l| + 1\right) M \\
&= \Big[M + (N - 1)R_p\Big] + |\boldsymbol{G}_l|\Big[M + R_p\Big] + |\boldsymbol{G}_l|\Big[(N - 2)R_1 + R_1\Big] \\
&= \Big[M + (N - 1)R_p\Big] + |\boldsymbol{G}_l|\big[M + R_p + (N - 2)R_1\big] + |\boldsymbol{G}_l| R_1
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\geq} H(Z_l, X_{\boldsymbol{S}_{K-N+1}}) + \sum_{\boldsymbol{q}_j \in \boldsymbol{G}_l} \Big[H(Z_l) + H(X_{\boldsymbol{q}_j}) + H(X_{\boldsymbol{A}_{N-2}})\Big] + |\boldsymbol{G}_l| H(X_{\boldsymbol{a}_N}) \\
&\overset{(b)}{\geq} H(Z_l, X_{\boldsymbol{S}_{K-N+1}}) + \sum_{\boldsymbol{q}_j \in \boldsymbol{G}_l} \Big[H(Z_l, X_{\boldsymbol{q}_j}, X_{\boldsymbol{A}_{N-2}})\Big] + |\boldsymbol{G}_l| H(X_{\boldsymbol{a}_N}) \\
&\overset{(c)}{=} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}) + \sum_{\boldsymbol{q}_j \in \boldsymbol{G}_l} \Big[H(W_{[N-1]}, Z_l, X_{\boldsymbol{q}_j}, X_{\boldsymbol{A}_{N-2}})\Big] + |\boldsymbol{G}_l| H(X_{\boldsymbol{a}_N}) \\
&\geq H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}) + \sum_{\boldsymbol{q}_j \in \boldsymbol{G}_l} \Big[H(W_{[N-1]}, Z_l, X_{\boldsymbol{q}_j})\Big] + |\boldsymbol{G}_l| H(X_{\boldsymbol{a}_N}) \\
&\overset{(b)}{\geq} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}) + H(W_{[N-1]}, Z_l, X_{\boldsymbol{G}_l}) + (|\boldsymbol{G}_l| - 1)\, H(W_{[N-1]}, Z_l) + |\boldsymbol{G}_l| H(X_{\boldsymbol{a}_N}) \\
&\overset{(b)}{\geq} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) + H(W_{[N-1]}, Z_l) + (|\boldsymbol{G}_l| - 1)\, H(W_{[N-1]}, Z_l) + |\boldsymbol{G}_l| H(X_{\boldsymbol{a}_N}) \\
&= H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) + |\boldsymbol{G}_l| H(W_{[N-1]}, Z_l) + |\boldsymbol{G}_l| H(X_{\boldsymbol{a}_N}) \\
&\overset{(b)}{\geq} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) + |\boldsymbol{G}_l| H(W_{[N-1]}, Z_l, X_{\boldsymbol{a}_N}) \\
&\overset{(c)}{=} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) + |\boldsymbol{G}_l| H(W_{[N]}, Z_l, X_{\boldsymbol{a}_N}) \\
&\overset{(d)}{=} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) + |\boldsymbol{G}_l| H(W_{[N]}) = H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) + |\boldsymbol{G}_l| N
\end{aligned}
$$

where

- (a) follows from (4.15) and (4.11),
- (b) follows from the submodularity property of entropy,
- (c) follows from (1.8),
- (d) follows from (1.9).

□

Lemma 20. *The demand sets* $\boldsymbol{S}_{K-N+1}$ *and* $\boldsymbol{G}_l$ *,* $1 \leq l \leq K-N$*, satisfy*

$$
\sum_{l=1}^{K-N} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) \geq (K-N-1)N + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{q}_N}) \qquad (4.17)
$$

Proof. We have,

$$
\begin{aligned}
\sum_{l=1}^{K-N} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) &= H(W_{[N-1]}, Z_1, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_1}) + H(W_{[N-1]}, Z_2, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_2}) \\
&\quad + \sum_{l=3}^{K-N} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) \\
&\overset{(a)}{\geq} H(W_{[N-1]}, Z_1, Z_2, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_1}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_2})
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{l=3}^{K-N} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) \\
&\overset{(b)}{=} H(W_{[N-1]}, W_N, Z_1, Z_2, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_1}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_2}) \\
&+ \sum_{l=3}^{K-N} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) \\
&\overset{(c)}{=} H(W_{[N]}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_2}) + H(W_{[N-1]}, Z_3, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_3}) \\
&+ \sum_{l=4}^{K-N} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) \\
&\overset{(a)}{\geq} H(W_{[N]}) + H(W_{[N-1]}, Z_3, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_2}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_3}) \\
&+ \sum_{l=4}^{K-N} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) \\
&\overset{(b)}{=} H(W_{[N]}) + H(W_{[N-1]}, W_N, Z_3, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_2}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_3}) \\
&+ \sum_{l=4}^{K-N} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) \\
&\overset{(c)}{=} 2H(W_{[N]}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_3}) + H(W_{[N-1]}, Z_4, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_4}) \\
&+ \sum_{l=5}^{K-N} H(W_{[N-1]}, Z_l, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_l}) \\
&\overset{(d)}{\geq} (K-N-1)H(W_{[N]}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_{K-N}}) \\
&\overset{(e)}{=} (K-N-1)N + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{q}_N})
\end{aligned}
$$

where,

- (a) follows from the submodularity property of entropy and (4.12)
- (b) follows from the fact that user U_{l+1} requests the file W_N in $\boldsymbol{q}_{K-l} \in \boldsymbol{G}_l$,
- (c) follows from (1.9),
- (d) follows from repeated use of (a), (b) and (c),
- (e) follows from (4.14).

□

Lemma 21. *The demand sets $\boldsymbol{S}_i$, $K-N+1 \leq i \leq K$, satisfy*

$$
H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}) + \sum_{i=K-N+2}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \geq (N-1)N + N - 1
$$

Proof. We have,

$$H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}) + \sum_{i=K-N+2}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i})$$

$$=H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}) + H(W_{[N-1]}, Z_{K-N+2}, X_{\boldsymbol{S}_{K-N+2}}) + \sum_{i=K-N+3}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i})$$

$$\overset{(a)}{\geq} H(W_{[N-1]}, Z_{K-N+2}, X_{\boldsymbol{S}_{K-N+1}}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+2}}) + \sum_{i=K-N+3}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i})$$

$$\overset{(b)}{=} H(W_{[N]}, Z_{K-N+2}, X_{\boldsymbol{S}_{K-N+1}}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+2}}) + \sum_{i=K-N+3}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i})$$

$$\overset{(c)}{=} H(W_{[N]}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+2}}) + H(W_{[N-1]}, Z_{K-N+3}, X_{\boldsymbol{S}_{K-N+3}}) + \sum_{i=K-N+4}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i})$$

$$\overset{(a)}{\geq} H(W_{[N]}) + H(W_{[N-1]}, Z_{K-N+3}, X_{\boldsymbol{S}_{K-N+2}}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+3}}) + \sum_{i=K-N+4}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i})$$

$$\overset{(b)}{=} H(W_{[N]}) + H(W_{[N]}, Z_{K-N+3}, X_{\boldsymbol{S}_{K-N+2}}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+3}}) + \sum_{i=K-N+4}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i})$$

$$\overset{(c)}{=} 2H(W_{[N]}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+3}}) + H(W_{[N-1]}, Z_{K-M+4}, X_{\boldsymbol{S}_{K-N+4}}) + \sum_{i=K-N+5}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i})$$

$$\overset{(d)}{\geq} (N-1)H(W_{[N]}) + H(W_{[N-1]}, X_{\boldsymbol{S}_K})$$

$$\overset{(e)}{=} (N-1)H(W_{[N]}) + H(W_{[N-1]}) = (N-1)N + N - 1$$

where,

(a) follows from the submodularity property of entropy and (4.13),

(b) follows from the fact that user U_{K-N+l} requests the file W_N in $\boldsymbol{q}_{N-l+1} \in \boldsymbol{S}_{K-N+l-1}$,

(c) follows from (1.9),

(d) follows from repeated use of (a), (b) and (c),

(d) follows from (4.11)

□

We have the following result:

Theorem 18. *For the (N, K) cache network with cache size M, achievable rates R_1 and R_p must satisfy the following constraint*

$$\frac{K(K-1)}{2}R_p + \frac{(N-1)((K-N)^2+K)}{2}R_1 + \frac{(K-N)^2+3K-N}{2}M \geq KN - 1 + \frac{(K-N)(K-N+1)}{2}N$$

where $K \leq N$ *and* $2 \leq p \leq N$.

Proof. We have,

$$
\begin{aligned}
&\frac{K(K-1)}{2}R_p + \frac{(N-1)((K-N)^2+K)}{2}R_1 + \frac{(K-N)^2+3K-N}{2}M \\
&\overset{(a)}{=} \sum_{i=1}^{K-N}\Big[\left(|\boldsymbol{G}_i|+N-1\right)R_p + \left(|\boldsymbol{G}_i|\right)(N-1)R_1 + \left(|\boldsymbol{G}_i|+1\right)M\Big] \\
&\quad + \sum_{i=K-N+1}^{K-p+1}(M+(K-i)R_p+(N-K+i-1)R_1) \\
&\quad + \sum_{i=K-p+2}^{K}(M+(K-i)R_p+(N-K+i-1)R_1) \\
&\overset{(b)}{\geq} \sum_{i=1}^{K-N}\Big[\left(|\boldsymbol{G}_i|+N-1\right)R_p + \left(|\boldsymbol{G}_i|\right)(N-1)R_1 + \left(|\boldsymbol{G}_i|+1\right)M\Big] \\
&\quad + \sum_{i=K-N+1}^{K-p+1} H(Z_i, X_{\boldsymbol{S}_i}, X_{\boldsymbol{C}_i}) + \sum_{i=K-p+2}^{K} H(Z_i, X_{S_i}, X_{\boldsymbol{A}_{N-K+i-1}}) \\
&\overset{(c)}{\geq} \sum_{i=1}^{K-N}\Big[H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_i}) + |\boldsymbol{G}_l|N\Big] + \sum_{i=K-N+1}^{K-p+1} H(Z_i, X_{\boldsymbol{S}_i}, X_{\boldsymbol{C}_i}) \\
&\quad + \sum_{i=K-p+2}^{K} H(Z_i, X_{S_i}, X_{\boldsymbol{A}_{N-K+i-1}}) \\
&\overset{(d)}{=} \sum_{i=1}^{K-N}\Big[H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_i}) + |\boldsymbol{G}_l|N\Big] + \sum_{i=K-N+1}^{K-p+1} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}, X_{\boldsymbol{C}_i}) \\
&\quad + \sum_{i=K-p+2}^{K} H(W_{[N-1]}, Z_i, X_{S_i}, X_{\boldsymbol{A}_{N-K+i-1}}) \\
&\geq \sum_{i=1}^{K-N}\Big[H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_i}) + |\boldsymbol{G}_l|N\Big] + \sum_{i=K-N+1}^{K-p+1} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \\
&\quad + \sum_{i=K-p+2}^{K} H(W_{[N-1]}, Z_i, X_{S_i}) \\
&= \sum_{i=1}^{K-N}\Big[H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{G}_i})\Big] + \sum_{i=K-N+1}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) + \sum_{i=1}^{K-N} |\boldsymbol{G}_i|N \\
&\overset{(e)}{\geq} (K-N-1)N + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{q}_N}) + \sum_{i=K-N+1}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \\
&\quad + \sum_{i=1}^{K-N} |\boldsymbol{G}_i|N
\end{aligned}
$$

$$
\begin{aligned}
&= H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{q}_N}) + H(W_{[N-1]}, Z_{K-N+1}, X_{\boldsymbol{S}_{K-N+1}}) + \sum_{i=K-N+2}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \\
&\quad + (K-N-1)N + \sum_{i=1}^{K-N} |\boldsymbol{G}_i| N \\
&\overset{(f)}{\geq} H(W_{[N-1]}, Z_{K-N+1}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{q}_N}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}) + \sum_{i=K-N+2}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \\
&\quad + (K-N-1)N + \sum_{i=1}^{K-N} |\boldsymbol{G}_i| N \\
&\overset{(g)}{=} H(W_{[N]}, Z_{K-N+1}, X_{\boldsymbol{S}_{K-N+1}}, X_{\boldsymbol{q}_N}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}) \\
&\quad + \sum_{i=K-N+2}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) + (K-N-1)N + \sum_{i=1}^{K-N} |\boldsymbol{G}_i| N \\
&\overset{(h)}{=} H(W_{[N]}) + H(W_{[N-1]}, X_{\boldsymbol{S}_{K-N+1}}) + \sum_{i=K-N+2}^{K} H(W_{[N-1]}, Z_i, X_{\boldsymbol{S}_i}) \\
&\quad + (K-N-1)N + \sum_{i=1}^{K-N} |\boldsymbol{G}_i| N \\
&\overset{(i)}{\geq} H(W_{[N]}) + (N-1)N + N - 1 + (K-N-1)N + \sum_{i=1}^{K-N} |\boldsymbol{G}_i| N \\
&= N + (N-1)N + N - 1 + (K-N-1)N + \sum_{i=1}^{K-N} (K-N-i+1)N \\
&= (K-1)N + N - 1 + \frac{(K-N)(K-N+1)}{2} N = KN - 1 + \frac{(K-N)(K-N+1)}{2} N
\end{aligned}
$$

where,

- (a) follows from (A.4)
- (b) follows from (4.11), (4.16), and (4.15),
- (c) follows from lemma 19,
- (d) follows from (1.8),
- (e) follows from Lemma 20,
- (f) follows from the submodularit property of entropy,
- (g) follows from the fact that user U_{K-N+1} requests the file W_N in $\boldsymbol{q}_N$,
- (h) follows from (1.9),
- (i) follows from Lemma 21.

□

When cache size $M \in \left[\frac{N}{K}(K-2), \frac{N}{K}(K-1)\right]$, the rate achieved by the YMA scheme for any demand

$\boldsymbol{d} \in \boldsymbol{D}_p$ (where $1 \leq p \leq \min(N, K)$) is given in Table 4.11 (see Appendix A.2). The constraint of Theorem 17 and Theorem 18 are satisfied with an equality by the scheme and any attempt at modifying the scheme to obtain better performance for R_p would entail a performance loss for R_1. Thus for $0 \leq M \leq \frac{N}{K}$, no scheme can dominate the YMA scheme.

Number of files	$R_1(M)$	$R_p(M)$	Constraints
$N \leq K$	$1 - \frac{M}{N}$	$\frac{(K^2+K-2)}{K(K-1)} - \frac{(K+1)M}{N(K-1)}$	$\frac{K(K-1)}{2}R_p + \frac{(N-1)((K-N)^2+K)}{2}R_1 + \frac{(K-N)^2+3K-N}{2}M = KN - 1 + \frac{(K-N)(K-N+1)}{2}N$
$K \leq N$	$1 - \frac{M}{N}$	$\frac{(K^2+K-2)}{K(K-1)} - \frac{(K+1)M}{N(K-1)}$	$\frac{K(K-1)}{2}R_p + \frac{K(2N-K-1)}{2}R_1 + KM = KN - 1$

Table 4.11: Rate achieved by YMA scheme when $M \in \left[\frac{N(K-2)}{K}, \frac{N(K-1)}{K}\right]$

4.6.3 Case III: $\frac{N}{K}(K-1) \leq M \leq N$

Interestingly, for the cache size $M \in \left[\frac{N}{K}(K-1), N\right]$, the YMA scheme achieves optimal rates for all the demand sets simultaneously. For the sake of completeness, we derive a memory rate constraint involving the the rates R_p and R_1. First, let us consider the case where $N \leq K$.

Theorem 19. *For the (N, K) cache network with case size M, achievable rates R_1 and R_p must satisfy the following constraint*

$$M + (N-1)R_p + R_1 \geq N$$

where $N \leq K$ and $2 \leq p \leq N$.

Proof. We have,

$$\begin{aligned}
M + (N-1)R_p + R_1 &\overset{(a)}{\geq} H(Z_1) + H(X_{\boldsymbol{S}_{K-N+1}}) + H(X_{\boldsymbol{a}_N}) \\
&\overset{(b)}{\geq} H(Z_1, X_{\boldsymbol{S}_{K-N-1}}) + H(X_{\boldsymbol{a}_N}) \\
&\overset{(c)}{=} H(W_{[N-1]}, Z_1, X_{\boldsymbol{S}_{K-N-1}}) + H(X_{\boldsymbol{a}_N}) \\
&\overset{(b)}{\geq} H(W_{[N-1]}, Z_1, X_{\boldsymbol{S}_{K-N-1}}, X_{\boldsymbol{a}_N}) \\
&\overset{(d)}{=} H(W_{[N]}, Z_1, X_{\boldsymbol{S}_{K-N-1}}, X_{\boldsymbol{a}_N}) \\
&\overset{(e)}{=} H(W_{[N]}) = N
\end{aligned}$$

where,

- (a) follows from the definitions of $\boldsymbol{S}_i$ and $\boldsymbol{a}_i$,
- (b) follows from the submodularity property of entropy,
- (c) follows from the fact that in demand set $\boldsymbol{S}_{K-N+1}$ user U_1 requests files $W_{[N-1]}$,
- (d) follows from the fact that user U_1 requests the file W_N in demand $\boldsymbol{a}_N$,
- (e) follows from (1.9).

□

Now consider the case $K \leq N$.

Theorem 20. *For the (N, K) cache network with case size M, achievable rates R_1 and R_p must satisfy the following constraint*

$$M + (N-1)R_p + (N-K+1)R_1 \geq N$$

where $K \leq N$ and $2 \leq p \leq K$.

Proof. We have,

$$\begin{aligned}
M + (N-1)R_p + (N-K+1)R_1 &\overset{(a)}{\geq} H(Z_1) + H(X_{\boldsymbol{Q}_1}) + H(X_{\boldsymbol{C}}) + H(X_{\boldsymbol{a}_K}) \\
&\overset{(b)}{\geq} H(Z_1, X_{\boldsymbol{Q}_1}) + H(X_{\boldsymbol{C}}) + H(X_{\boldsymbol{a}_K}) \\
&\overset{(c)}{=} H(W_{[K-1]}, Z_1, X_{\boldsymbol{Q}_1}) + H(X_{\boldsymbol{C}}) + H(X_{\boldsymbol{a}_K}) \\
&\overset{(b)}{\geq} H(W_{[K-1]}, Z_1, X_{\boldsymbol{Q}_1}, X_{\boldsymbol{C}}) + H(X_{\boldsymbol{a}_K}) \\
&\overset{(d)}{=} H(W_{[N]\setminus\{K\}}, Z_1, X_{\boldsymbol{Q}_1}, X_{\boldsymbol{C}}) + H(X_{\boldsymbol{a}_K}) \\
&\overset{(b)}{\geq} H(W_{[N]\setminus\{K\}}, Z_1, X_{\boldsymbol{Q}_1}, X_{\boldsymbol{C}}, X_{\boldsymbol{a}_K}) \\
&\overset{(e)}{=} H(W_{[N]\setminus\{K\}}, W_K, Z_1, X_{\boldsymbol{Q}_1}, X_{\boldsymbol{C}}, X_{\boldsymbol{a}_K}) \\
&\overset{(f)}{\geq} H(W_{[N]}) = N
\end{aligned}$$

where,

(a) follows from the definitions of $\boldsymbol{Q}_i$, $\boldsymbol{C}$, and a_i (4.3),

(b) follows from the submodularity property of entropy,

(c) follows from the fact that in demand set $\boldsymbol{Q}_1$ user U_1 requests files $W_{[K-1]}$,

(d) follows from the fact that in demand set $\boldsymbol{C}$ user U_1 requests files $\{W_{K+1}, W_{K+2}, \ldots, W_N\}$,

(f) follows from (1.9).

□

When cache size $M \in \left[\frac{N}{K}(K-1), N\right]$, the rate achieved by YMA scheme for any demand $\boldsymbol{d} \in \boldsymbol{D}_p$ (where $1 \leq p \leq \min(N, K)$) is given in Table 4.12. The constraint of Theorem 19 and Theorem 20 are satisfied with an equality by the scheme and any attempt at modifying the scheme to obtain better performance for R_p would entail a performance loss for R_1. Thus for $\frac{N}{K}(K-1) \leq M \leq N$, no scheme can dominate the YMA scheme.

Number of files	R_p	Constraints
$N \leq K$	$1 - \frac{M}{N}$	$M + (N-1)R_p + R_1 = N$
$K \leq N$	$1 - \frac{M}{N}$	$M + (N-1)R_p + (N-K+1)R_1 = N$

Table 4.12: Rate achieved by YMA scheme when $M \in \left[\frac{N(K-1)}{K}, N\right]$

The consequence of Theorems 16-20 for the YMA scheme is summarized in the following result:

Theorem 21. *For the (N, K) cache network, the YMA scheme is Pareto optimal when $M \in \left[0, \frac{N}{K}\right]$ and $M \in \left[\frac{N}{K}(K-2), N\right]$.*

4.7 Conclusions

In this chapter we introduce new kind of lower bounds which jointly constrain the rates achievable by the different demand sets. These lower bounds give us a better insight into how the performance for one demand set affects the performance for the other demand sets. We use such bounds to demonstrate that there is no universal scheme when coding is permitted in both the placement and the delivery phase. They are also used to demonstrate that the CFL scheme operates at the Pareto optimal frontier when $N \leq K$ and $0 \leq M \leq \frac{1}{K}$. In a similar fashion, the YMA scheme was shown to be Pareto optimal when $0 \leq M \leq \frac{N}{K}$ and $\frac{N}{K}(K-2) \leq M \leq N$.

5

Summary and Future Work

Contents

5.1 Summary of The Thesis

In this thesis, we studied different aspects of optimality in coded caching. We noted that the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_N$ was known for small caches where $0 \leq M \leq \frac{1}{K}$ and large caches where $\frac{N}{K}(K-1) \leq M \leq N$. New lower bounds and a new coded caching scheme were presented which characterized the optimal rate memory tradeoff for small caches where $0 \leq M \leq \frac{1}{K} + \frac{1}{K(N-1)}$ and for large caches where $\frac{N}{K}(K-1) - \frac{N-1}{K(K-1)} \leq M \leq N$. These results are summarized in Figure 5.1.

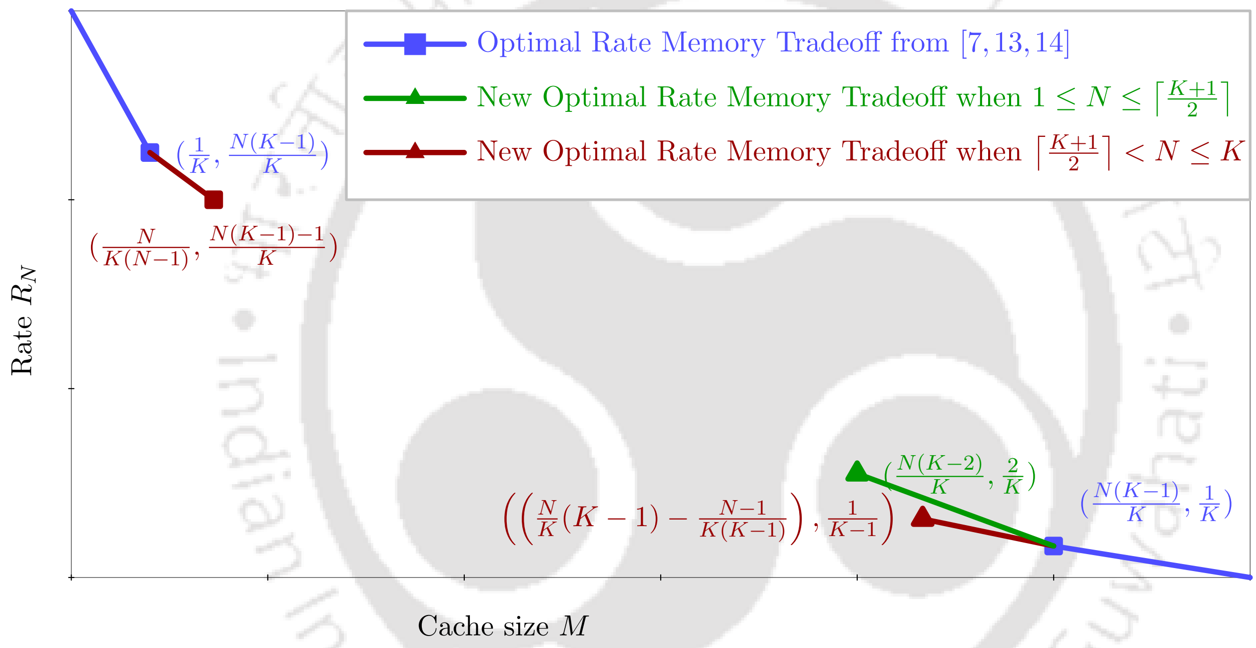

Figure 5.1: Optimal rate memory tradeoff for the (N, K) cache network

We then studied the problem of coded caching from a multi-objective perspective and derived new lower bounds that jointly constrained the achievable rates for the demand sets $\boldsymbol{D}_p$, $1 \leq p \leq \min\{N, K\}$. These bounds were used to argue that universal schemes do not exist for the general problem of coded caching and to motivate the study of Pareto optimal schemes. We considered the CFL scheme when $M \in \left[0, \frac{N}{K}\right]$ and the YMA scheme when $M \in \left[0, \frac{N}{K}\right] \cup \left[\frac{N}{K}(K-2), N\right]$ and demonstrated that they operate on the Pareto optimal frontier of the problem of coded caching.

5.2 Future Scope

The main contributions of the thesis is in extending the characterization of the optimal rate memory tradeoff for the demand set $\boldsymbol{D}_N$ and the demonstration of the Pareto optimal nature of the CFL and YMA scheme. The following problems are worth investigating in light of these presented results:

- Demand Set $\boldsymbol{D}_N$: The optimal rate memory tradeoff for this demand set is still not known when $M \in \left[\frac{1}{K} + \frac{1}{K(N-1)}, \frac{N}{K}(K-1) - \frac{N-1}{K(K-1)}\right]$. Can a characterization be obtained by extending the techniques presented in this thesis?

- Demand Set $\boldsymbol{D}_p$: The optimal rate memory tradeoff for the demand set $\boldsymbol{D}_p$, $1 < p < N$ has received much less attention in past work. Can results similar to the ones obtained in this thesis be arrived at using similar techniques ?

- The YMA Scheme: The Pareto optimality of the YMA scheme was demonstrated in this thesis when $M \in \left[0, \frac{N}{K}\right] \cup \left[\frac{N}{K}(K-2), N\right]$. Can this result be extended to when $M \in \left[\frac{N}{K}, \frac{N}{K}(K-2)\right]$ using similar techniqies?

- The CFL scheme which is known to be optimal for the demand set $\boldsymbol{D}_N$, was also shown to be Pareto optimal in this thesis. In this context it is natural to ask whether the scheme proposed by Gómez-Vilardebó, which in this thesis was shown to be optimal for the demand set $\boldsymbol{D}_N$, is a Pareto optimal scheme? Similarly, is the scheme we proposed in Chapter 3 (and shown to be optimal for the demand set $\boldsymbol{D}_N$) Pareto optimal?

- Can the study of Pareto optimal schemes be extended to further variations of the problem of coded caching and other index coding problems?

A

Appendix

Contents

A.1 Regarding the Proof of Theorem 18

For $1 \leq i \leq K-N$, consider the set $\boldsymbol{G}_i$, such that $|\boldsymbol{G}_i| = K-N-i+1$. We have the following results:

$$\sum_{i=1}^{K-N}\Big((N-1)+|\boldsymbol{G}_i|\Big)+\sum_{i=K-N+1}^{K}(K-i)=\sum_{i=1}^{K-N}\Big((N-1)+K-N-i+1\Big)+\sum_{i=K-N+1}^{K}(K-i)$$
$$=\sum_{i=1}^{K}(K-i)=\frac{K(K-1)}{2} \tag{A.1}$$

$$\sum_{i=1}^{K-N}(N-1)|\boldsymbol{G}_i|+\sum_{i=K-N+1}^{K}(N-K+i-1)=\sum_{i=1}^{K-N}(N-1)(K-N-i+1)+\sum_{i=K-N+1}^{K}(N-K+i-1)$$
$$=\frac{(N-1)(K-N)(K-N+1)}{2}+\frac{N(N-1)}{2}$$
$$=\frac{(N-1)((K-N)^2+K)}{2} \tag{A.2}$$

$$\sum_{i=1}^{K-N}(|\boldsymbol{G}_i|+1)+\sum_{K-N+1}^{K}1=\sum_{i=1}^{K-N}(K-N-i+1+1)+\sum_{K-N+1}^{K}1$$
$$=\frac{(K-N)(K-N+3)}{2}+N$$
$$=\frac{(K-N)^2+3K-N}{2} \tag{A.3}$$

Now using (A.1), (A.2) and (A.3) we have,

$$\sum_{i=1}^{K-N}\Big[(|\boldsymbol{G}_i|+N-1)R_p+(|\boldsymbol{G}_i|)\,(N-1)R_1+(|\boldsymbol{G}_i|+1)\,M\Big]$$
$$+\sum_{i=K-N+1}^{K-p+1}(M+(K-i)R_p+(N-K+i-1)R_1)$$
$$+\sum_{i=K-p+2}^{K}(M+(K-i)R_p+(N-K+i-1)R_1)$$
$$=\left[\sum_{i=1}^{K-N}\Big((N-1)+|\boldsymbol{G}_i|\Big)+\sum_{i=K-N+1}^{K}(K-i)\right]R_p+\left[\sum_{i=1}^{K-N}(N-1)|\boldsymbol{G}_i|+\sum_{i=K-N+1}^{K}(N-K+i-1)\right]R_1$$
$$+\left[\sum_{l=1}^{K-N}(|\boldsymbol{G}_l|+1)+\sum_{K-N+1}^{K}1\right]M$$
$$=\frac{K(K-1)}{2}R_p+\left[\frac{(N-1)((K-N)^2+K)}{2}\right]R_1+\left[\frac{(K-N)^2+3K-N}{2}\right]M \tag{A.4}$$

A.2 Regarding Constraints in Table 4.11

In this section we give details of constraints presented in Table 4.11. The rates achieved by the YMA scheme for the (N, K) cache network when cache size $M \in \left[\frac{N}{K}(K-2), \frac{N}{K}(K-1)\right]$ are:

$$R_1 = 1 - \frac{1}{N}M \tag{A.5}$$

$$R_p = \frac{(K^2+K-2)}{K(K-1)} - \frac{(K+1)}{N(K-1)}M \tag{A.6}$$

where $2 \leq p \leq \min\{N, K\}$. Thus we have,

$$\begin{aligned}
\frac{K(K-1)}{2}R_p + \frac{K(2N-K-1)}{2}R_1 + KM &= \frac{K(K-1)}{2}\left(\frac{(K^2+K-2)}{K(K-1)} - \frac{(K+1)}{N(K-1)}M\right) \\
&+ \frac{K(2N-K-1)}{2}\left(1 - \frac{1}{N}M\right) + KM \\
&= \left(\frac{(K^2+K-2)}{2} - \frac{K(K+1)}{2N}M\right) + \left(\frac{K(2N-K-1)}{2} - \frac{K(2N-K-1)}{2N}M\right) + KM \\
&= \left(\frac{(K^2+K-2)}{2} + \frac{K(2N-K-1)}{2}\right) + \left(K - \frac{K(K+1)}{2N} - \frac{K(2N-K-1)}{2N}\right)M \\
&= \left(\frac{(K^2+K-2)+K(2N-K-1)}{2}\right) + \left(K - \frac{K(K+1)+K(2N-K-1)}{2N}\right)M \\
&= \left(\frac{K^2+K-2+2NK-K^2-K}{2}\right) + \left(K - \frac{K^2+K+2NK-K^2-K}{2N}\right)M \\
&= \left(\frac{2NK-2}{2}\right) + \left(K - \frac{2NK}{2N}\right)M \\
&= KN - 1 + (K-K)\,M = KN - 1
\end{aligned}$$

We also have,

$$\begin{aligned}
&\frac{K(K-1)}{2}R_p + \frac{(N-1)((K-N)^2+K)}{2}R_1 + \frac{(K-N)^2+3K-N}{2}M \\
&= \frac{K(K-1)}{2}\left(\frac{(K^2+K-2)}{K(K-1)} - \frac{(K+1)}{N(K-1)}M\right) + \frac{(N-1)((K-N)^2+K)}{2}\left(1 - \frac{1}{N}M\right) \\
&+ \frac{(K-N)^2+3K-N}{2}M \\
&= \left(\frac{(K^2+K-2)}{2} + \frac{(N-1)((K-N)^2+K)}{2}\right) \\
&+ \left(\frac{(K-N)^2+3K-N}{2} - \frac{K(K+1)}{2N} - \frac{(N-1)((K-N)^2+K)}{2N}\right)M \\
&= \left(\frac{(K^2+K-2)+(N-1)(K^2+N^2-2NK+K)}{2}\right) \\
&+ \left(\frac{(K-N)^2+3K-N}{2} - \frac{K^2+K+(N-1)(K^2+N^2-2NK+K)}{2N}\right)M
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{2NK - 2 + N(K^2 + N^2 - 2NK + K) - N^2}{2}\right) \\
&\quad + \left(\frac{(K-N)^2 + 3K - N}{2} - \frac{N(K^2 + N^2 - 2NK + K) - N^2 + 2NK}{2N}\right) M \\
&= \left(\frac{2NK - 2 + N(K^2 + N^2 - 2NK + K - N)}{2}\right) \\
&\quad + \left(\frac{(K-N)^2 + 3K - N}{2} - \frac{N(K^2 + N^2 - 2NK + K + 2K - N)}{2N}\right) M \\
&= \left(\frac{2NK - 2 + N((K-N)^2 + K - N)}{2}\right) \\
&\quad + \left(\frac{(K-N)^2 + 3K - N}{2} - \frac{N((K-N)^2 + 3K - N)}{2N}\right) M \\
&= \left(\frac{2KN - 2 + N(K-N)((K-N)+1)}{2}\right) = KN - 1 + \frac{(K-N)(K-N+1)}{2} N
\end{aligned}
$$

Bibliography

[1] A. Luotonen and K. Altis, "World-wide web proxies," *Computer Networks and ISDN systems*, vol. 27, no. 2, pp. 147–154, 1994.

[2] J. Wang, "A survey of web caching schemes for the internet," *ACM SIGCOMM Computer Communication Review*, vol. 29, no. 5, pp. 36–46, 1999.

[3] S. Podlipnig and L. Böszörmenyi, "A survey of web cache replacement strategies," *ACM Computing Surveys*, vol. 35, no. 4, pp. 374–398, 2003.

[4] A. Vakali, "Proxy cache replacement algorithms: A history-based approach," *World Wide Web*, vol. 4, no. 4, pp. 277–297, 2001.

[5] Y. Li, J. Li, J. Chen, M. Lu, and C. Li, "Seed selection for data offloading based on social and interest graphs," *Computers, Materials and Continua*, 2018.

[6] F. You, T. Liu, X. Peng, J. Liang, B. Zhang, and Y. Zhou, "An efficient web caching replacement algorithm," in *International Conference on Artificial Intelligence and Security*. Springer, 2020, pp. 479–488.

[7] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2856–2867, 2014.

[8] C. Tian, "Symmetry, outer bounds, and code constructions: A computer-aided investigation on the fundamental limits of caching," *MDPI Entropy*, vol. 20, no. 8, p. 603, 2018.

[9] K. Wan, D. Tuninetti, and P. Piantanida, "On caching with more users than files," in *International Symposium on Information Theory*. IEEE, 2016, pp. 135–139.

[10] M. M. Amiri, Q. Yang, and D. Gündüz, "Coded caching for a large number of users," in *Information Theory Workshop*. IEEE, 2016, pp. 171–175.

[11] K. Wan, D. Tuninetti, and P. Piantanida, "A novel index coding scheme and its application to coded caching," in *Information Theory and Applications Workshop*. IEEE, 2017, pp. 1–6.

[12] K. Wan, D. Tuninetti, and P. Piantanida, "An index coding approach to caching with uncoded cache placement," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1318–1332, 2020.

[13] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1281–1296, 2017.

[14] Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: Improved bounds for users with small buffers," *IET Communications*, vol. 10, no. 17, pp. 2315–2318, 2016.

[15] S. Sahraei and M. Gastpar, "K users caching two files: An improved achievable rate," in *Conference on Information Science and Systems* . IEEE, 2016, pp. 620–624.

[16] M. M. Amiri and D. Gündüz, "Fundamental limits of coded caching: Improved delivery rate-cache capacity tradeoff," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 806–815, 2017.

[17] C. Tian and J. Chen, "Caching and delivery via interference elimination," in *International Symposium on Information Theory*. IEEE, 2016, pp. 830–834.

[18] J. Gómez-Vilardebó, "Fundamental limits of caching: Improved rate-memory trade-off with coded prefetching," *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4488–4497, 2018.

[19] S. Shao, J. Gomez-Vukardebo, K. Zhang, and C. Tian, "On the fundamental limit of coded caching systems with a single demand type," in *Information Theory Workshop*, IEEE, 2016, pp.1–5.

[20] S. Shao, J. Gómez-Vilardebó, K. Zhang, and C. Tian, "On the fundamental limits of coded caching systems with restricted demand types," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 863–873, 2020.

[21] A. Sengupta, R. Tandon, and T. C. Clancy, "Improved approximation of storage-rate tradeoff for caching via new outer bounds." in *International Symposium on Information Theory*, IEEE, 2015, pp. 1691–1695.

[22] C.-Y. Wang, S. H. Lim, and M. Gastpar, "A new converse bound for coded caching," in *Information Theory and Applications Workshop*, IEEE, 2016, pp. 1–6.

[23] S. H. Lim, C.-Y. Wang, and M. Gastpar, "Information-theoretic caching: The multi-user case," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7018–7037, 2017.

[24] C.-Y. Wang, S. S. Bidokhti, and M. Wigger, "Improved converses and gap results for coded caching," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7051–7062, 2018.

[25] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4388–4413, 2017.

[26] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "Characterizing the rate-memory tradeoff in cache networks within a factor of 2," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 647–663, 2018.

[27] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," *IEEE/ACM Transactions on Networking*, vol. 23, no. 4, pp. 1029–1040, 2015.

[28] S. A. Saberali, L. Lampe, and I. F. Blake, "Decentralized coded caching without file splitting," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 1289–1303, 2019.

[29] U. Niesen and M. A. Maddah-Ali, "Coded caching with nonuniform demands," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 1146–1158, 2017.

[30] J. Zhang, X. Lin, and X. Wang, "Coded caching under arbitrary popularity distributions," *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 349–366, 2017.

[31] J. Hachem, N. Karamchandani, and S. N. Diggavi, "Coded caching for multi-level popularity and access," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 3108–3141, 2017.

[32] N. Karamchandani, U. Niesen, M. A. Maddah-Ali, and S. N. Diggavi, "Hierarchical coded caching," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3212–3229, 2016.

[33] S. P. Shariatpanahi, S. A. Motahari, and B. H. Khalaj, "Multi-server coded caching," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 7253–7271, 2016.

[34] V. Ravindrakumar, P. Panda, N. Karamchandani, and V. M. Prabhakaran, "Private coded caching," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 685–694, 2017.

[35] A. M. Daniel and W. Yu, "Optimization of heterogeneous coded caching," *IEEE Transactions on Information Theory*, 2019.

[36] M. M. Amiri, Q. Yang, and D. Gündüz, "Decentralized caching and coded delivery with distinct cache capacities," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 4657–4669, 2017.

[37] C. Li, "On rate region of caching problems with non-uniform file and cache sizes," *IEEE Communications Letters*, vol. 21, no. 2, pp. 238–241, 2016.

[38] E. Parrinello, A. Unsal, and P. Elia, "Fundamental limits of caching in heterogeneous networks with uncoded prefetching," *arXiv preprint arXiv:1811.06247*, 2018.

[39] M. Cheng, J. Jiang, Q. Yan, and X. Tang, "Constructions of coded caching schemes with flexible memory size," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4166–4176, 2019.

[40] D. Cao, D. Zhang, P. Chen, N. Liu, W. Kang, and D. Gündüz, "Coded caching with asymmetric cache sizes and link qualities: The two-user case," *IEEE Transactions on Communications*, vol. 67, no. 9, pp. 6112–6126, 2019.

[41] A. M. Ibrahim, A. A. Zewail, and A. Yener, “Coded caching for heterogeneous systems: An optimization perspective,” *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5321–5335, 2019.

[42] E. Parrinello and P. Elia, “Coded caching with optimized shared-cache sizes,” in *Information Theory Workshop*, IEEE, 2019, pp. 1–5.

[43] A. Malik, B. Serbetci, E. Parrinello, and P. Elia, “Fundamental limits of stochastic caching networks,” *arXiv preprint arXiv:2005.13847*, 2020.

[44] S. Sasi and B. S. Rajan, “An improved multi-access coded caching with uncoded placement,” *arXiv preprint arXiv:2009.05377*, 2020.

[45] K. S. Reddy and N. Karamchandani, “Rate-memory trade-off for multi-access coded caching with uncoded placement,” *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3261–3274, 2020.

[46] H. Ghasemi and A. Ramamoorthy, “Asynchronous coded caching with uncoded prefetching,” *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2146–2159, 2020.

[47] H. Ghasemi and A. Ramamoorthy, “Algorithms for asynchronous coded caching,” in *Asilomar Conference on Signals, Systems, and Computers*, IEEE, 2017, pp. 636–640.

[48] H. Ghasemi and A. Ramamoorthy, “Asynchronous coded caching,” in *International Symposium on Information Theory*, IEEE, 2017, pp. 2438–2442.

[49] M. M. Amiri and D. Gündüz, “Cache-aided data delivery over erasure broadcast channels,” in *International Conference on Communications*, IEEE, 2017, pp. 1–6.

[50] N. S. Karat, A. Thomas, and B. S. Rajan, “Error correction in coded caching with symmetric batch prefetching,” *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5264–5274, 2019.

[51] A. Sengupta, R. Tandon, and T. C. Clancy, “Fundamental limits of caching with secure delivery,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 355–370, 2014.

[52] Y.-P. Wei, K. Banawan, and S. Ulukus, “Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits,” *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1126–1139, 2018.

[53] S. Kamath, J. Ravi, and B. K. Dey, “Demand-private coded caching and the exact trade-off for N= K= 2,” in *National Conference on Communications*, IEEE, 2020, pp. 1–6.

[54] Ç. 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 Transactions on Communications*, vol. 67, no. 12, pp. 8179–8192, 2019.

[55] M. Ji, G. Caire, and A. F. Molisch, “Fundamental limits of caching in wireless D2D networks,” *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 849–869, 2015.

[56] K. Wan, D. Tuninetti, M. Ji, and P. Piantanida, “Fundamental limits of distributed data shuffling,” in *Allerton Conference on Communication, Control, and Computing* , IEEE, 2018, pp. 662–669.

[57] M. A. Attia and R. Tandon, “Near optimal coded data shuffling for distributed learning,” *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7325–7349, 2019.

[58] K. Wan, D. Tuninetti, M. Ji, G. Caire, and P. Piantanida, “Fundamental limits of decentralized data shuffling,” *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3616–3637, 2020.

[59] K. P. Vijith Kumar, B. Kumar Rai, and T. Jacob, “Pareto optimal schemes in coded caching,” in *International Symposium on Information Theory*, IEEE, 2019, pp. 2629–2633.

[60] K. V. Kumar, B. K. Rai, and T. Jacob, “Towards the exact memory rate tradeoff for the (4, 5) cache network,” in *International Conference on Signal Processing and Communications*, IEEE, 2020, pp. 1–5.

[61] K. P. Vijith Kumar, B. K. Rai, and T. Jacob, “Towards the exact rate memory tradeoff in coded caching,” in *National Conference on Communications*, IEEE, 2019, pp. 1–6.

[62] K. P. Vijith Kumar, B. K. Rai, and T. Jacob, “Fundamental limits of coded caching: The memory rate pair (K-1-1/K, 1/(K-1)),” in *International Symposium on Information Theory* , IEEE, 2019, pp. 2624–2628.

[63] K. Wan, D. Tuninetti, and P. Piantanida, "On the optimality of uncoded cache placement," in *Information Theory Workshop*, IEEE, 2016, pp. 161–165.

[64] N. Ajaykrishnan, N. S. Prem, V. M. Prabhakaran, and R. Vaze, "Critical database size for effective caching," in *National Conference on Communications*, IEEE, 2015, pp. 1–6.

List of Publications

Journals

Published:

- K. P. Vijith Kumar, B. K. Rai, and T. Jacob, "Towards The Optimal Rate Memory Tradeoff in Caching with Coded Placement," accepted in *IEEE Transactions on Information Theory*, 2022. (**Chapter 2** and **Chapter 3**)

Under preparation:

- K. P. Vijith Kumar, B. K. Rai, and T. Jacob, "Pareto Optimal Schemes in Coded Caching: The (N, 4) cache network," to be submitted to *IEEE Communication letters.* (**Chapter 4**)
- K. P. Vijith Kumar, B. K. Rai, and T. Jacob, "Pareto Optimal Schemes in Coded Caching: The (N, K) cache network," to be submitted to *IEEE Transactions on Information Theory.* (**Chapter 4**)

Conferences

Published:

- K. P. Vijith Kumar, B. K. Rai, and T. Jacob, "Towards the exact rate memory tradeoff in coded caching," in *Proc. IEEE National Conf. Commun. (NCC)*, Bangalore, India, February 2019, pp.1–6. (**Chapter 3**)
- K. P. Vijith Kumar, B. K. Rai, and T. Jacob, "Pareto optimal schemes in coded caching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, July 2019, pp. 2629–2633.(**Chapter 4**)
- K. P. Vijith Kumar, B. K. Rai, and T. Jacob, "Fundamental limits of coded caching: The memory rate pair (K-1-1/K, 1/(K-1))," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, July 2019, pp. 2624–2628.(**Chapter 3**)

- K. P. Vijith Kumar, B. K. Rai, and T. Jacob, "Towards the Exact Memory Rate Tradeoff for the (4, 5) Cache Network" in *Proc. Int. Conf. on Signal Proc. and Commun. (SPCOM)* 2020 Jul 19, pp. 1–5. (**Chapter 2** and **Chapter 3**)
- K. P. Vijith Kumar, B. K. Rai, and T. Jacob, "Pareto Optimal Schemes in Coded Caching: Uncoded Prefetching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, July 2021, pp. 2624–2628.(**Chapter 4**)