Towards the Exact Rate-Memory Trade-off for Uncoded Caching with Secure Delivery

Mohsen Bahrami Mohamed Adel Attia Ravi Tandon Bane Vasić
Department of Electrical and Computer Engineering
University of Arizona, Tucson, AZ 85721
E-mail:{*bahrami, madel, tandonr, vasic*}*@email.arizona.edu*

***Abstract*— We consider the problem of secure delivery in a single-hop caching network with a central server connected to multiple end-users via an insecure multi-cast link. The server has a database of a set of files (content) and each user, which is equipped with a local cache memory, requests access one of the files. In addition to delivering users' requests, the server needs to keep the files (information-theoretically) secure from an external eavesdropper observing the underlying communications between the server and users. We focus on an important class of content placement strategies where the prefetching is required to be *uncoded* and caches are filled with uncoded fragments of files. In this paper, we establish the exact characterization of secure rate-memory trade-off for the worst-case communication rate through a matching converse under the constraint of uncoded cache placement where the number of users is no larger than the number of files in the database.**

I. Introduction

The rapid growth of wireless data usage in the past two decades required researchers to look for innovative ways to avoid data congestion especially during peak traffic [1]. Video or Internet data traffic is typically characterized by *asynchronous content reuse*, i.e., there are few popular files which are in high demand among users, but are requested at arbitrary times. To cater such traffic demands, especially in applications such as video-on-demand, proactive caching of popular content during off-peak periods can reduce the communication load during peak time. This scenario of caching/storage and delivery, wherein some fragments of the requested files are cached locally, and the remaining fragments are delivered by the remote server via separate unicast transmissions, has been extensively studied in literature [2], [3]. However, traditional caching schemes are far from optimal and also, scalability becomes a concern as the number of users in the system increases.

Using opportunities of multicast delivery and coding, Maddah-Ali and Niesen proposed an information theoretic formulation of the caching problem, establishing the approximate rate-memory tradeoff between the transmission rate and the cache size [4]. In their scheme, the server communicates with a set of users with uniform cache size, over a shared link network, with the caching scheme objective to minimize the worst case transmission rate over all feasible user demands. Their coded caching scheme has a significant improvement over traditional caching schemes and was also proven to be optimal under the constraint of uncoded cache placemennt [5], [6]. This fundamental understanding of caching networks heralded a series of interesting results ranging from proposing improved lower bounds [7]–[9], and achievability schemes to the application of coded caching scheme on different topologies [10], changing parameters such as different cache [11], [12], and file size [13], investigating caching in Device-to-Device (D2D) communication networks [14] as well as decentralized variations [15]. Some of the notable developments, results and challenges in these directions are summarized in [16].

With the advent of new technologies and networks suited for 5G communication, the study of caching problems with security constraints in ultra-dense networks is important [17]. Consequently, secure transmission of messages, and their confidentiality become of one of the important aspects in caching problems and are investigated in recent works [18]–[21]. In [18], the concept of *secure delivery* is introduced with the goal of securely delivering the files to the users, while communicating over a public channel which overhears by an external eavesdropper. In [19], the notion of *secretive caching* is presented where any user should only be able to obtain information about the file it has requested and not other files requested by the other users. The secretive constraint models a video-on-demand system in which users need to pay each time for their requested content and no user should be able to obtain any information of other files.

Contributions: In this paper we characterize the information theoretically optimal secure delivery rate when the number of files in the server, N, is at least as large as the number of the users, K, i.e., $N \geq K$ in the existence of an external eavesdropper, and under uncoded cache placement constraint. We find that the secure delivery scheme introduced in [18] is in fact information-theoretically optimal. In our converse proof, we follow a novel bounding methodology similar to the recent result in [5], where the optimal uncoded cache placement problem is considered. The novel part in our converse proof is introducing the secure delivery constraint. This additional constraint together with the constraint that each user should be able to decode its requested file deliver a new linear programming. Then, we solve the resultant linear programming subject to problem constraints (file size and cache constraints) in order to find the best lower bounds over different regimes of the available storage. Furthermore, we demonstrate that the obtained lower bound matches with

The work of M. Attia and R. Tandon was supported by the NSF grant CAREER 1651492.

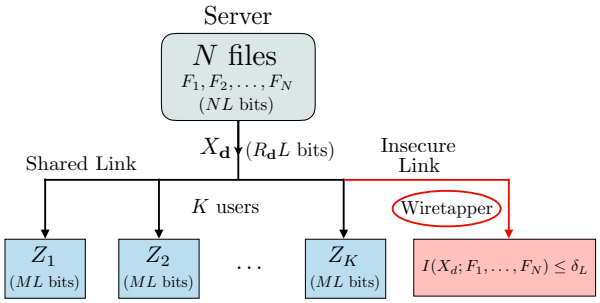

Fig. 1. System model for a single-hop caching network with secure delivery.

the achievable bound in [18].

Notations: $[n_1 : n_2]$ represents the set of all integers between n_1, and n_2, i.e., $\{n_1, n_1+1, \ldots, n_2\}$. Throughout the paper $\binom{n}{k} = 0$ for $k < n$, and $k, n < 0$. We denote the bold small letters for ordered sets. In order to describe subsets of ordered sets, we use the subscript to give the indexes of the elements being chosen from the set, e.g. for the ordered set $\mathbf{a} = (a_1 \ldots, a_n)$, $a_{[1:4]} = (a_1, a_2, a_3, a_4)$. We denote random variables (RVs) by capital letters. The subscript of a set of ordered RVs is used for short notation of a subset of a set of ordered RVs, e.g., for a set of RVs $X_1, \ldots, X_n$, we use $X_{[2:4]}$ to denote X_2, X_3, and X_4. For files, the sets in subscript is used to denote the partition of the file *only* stored by all the users indexed by the set, while the sets between parentheses in superscript is the total part of the file stored at the user indexed with that set, e.g., $F^{(\{1,2\})}$ is the union of the parts stored in users 1, and 2 about the file F, while $F_{\{1,2\}}$ is the part of F stored *at both* users 1, and 2.

II. Problem Setup

We consider a single-hop content delivery network with a central server connected to K users through a noiseless multicast link. An external wiretapper can observe communications from the server to the users and intends to eavesdrop the files as shown in Figure 1. We assume that a user can request access to any one of the files at a given time. In addition to satisfying users' demands, messages sent over the multicast link must be kept information-theoretically secure from the external wiretapper. The central server has a database of N independent files denoted by $\mathbf{F} = \{F_1, F_2, \ldots, F_N\}$, where each file is of size $L \in \mathbb{N}$ bits and is uniformly distributed over $\left[1 : 2^L\right]$. Each user is equipped with a cache memory of size MF bits for some $1 \leq M \leq N$, where M is the normalized cache memory size. In order to satisfy the security constraint, a randomness in the form of keys are introduced, which occupy a fragment of each user's cache. Subsequently, these keys can be used in the delivery phase to achieve information-theoretically secure delivery. For this, the server generates a set of orthogonal keys denoted by $\mathcal{K}$ (which are also independent of files) to be shared with the users.

A secure caching scheme has two key phases of operation, namely the content placement phase and the secure content delivery phase.

Content Placement Phase: Each user $k \in [1 : K]$ stores a combination of bits from each file

$$Q_k \triangleq \phi_k \left(F_{[1:N]}\right), \tag{1}$$

where ϕ_k is the caching function at user k

$$\phi_k : \left[1 : 2^L\right]^N \to \left[1 : 2^{\lfloor M_D L \rfloor}\right], \tag{2}$$

as well as some keys shared with the server, denoted as $K^{(k)} \subseteq \mathcal{K}$ of size $M_k L$ bits, where $M = M_D + M_K$. Therefore, the cache content of user k is $Z_k = Q_k \cup K^{(k)}$. For uncoded placement of the files in which each user is allowed to cache any subset of bits of the files in an uncoded manner, Z_k is given as

$$Z_k = \left(F_1^{(k)}, \ldots, F_N^{(k)}, K^{(k)}\right), \tag{3}$$

where $F_n^{(k)}$ is the combination of bits of file F_n stored at the cache Z_k.

To generalize the cache placement, let us define $F_{n,\mathcal{W}}$ as the part of file F_n, for $n \in [1 : N]$, that is only stored at the cache of the users given by the set $\mathcal{W}$. Also, we define $K_{\mathcal{W}}$ as the key shared only among the cache of the users given by the set $\mathcal{W}$. Therefore, generally each file F_n, similarly the keys $\mathcal{K}$, is composed of 2^K partitions for $\mathcal{W} \in 2^{[1:K]}$, where $2^{[1:K]}$ is the power set containing all the possible subsets of the set $[1 : K]$ including the empty set, i.e.,

$$F_n = \bigcup_{\mathcal{W} \subseteq [1:K]} F_{n,\mathcal{W}}, \tag{4}$$

$$\mathcal{K} = \bigcup_{\mathcal{W} \subseteq [1:K]} K_{\mathcal{W}}. \tag{5}$$

Also, if we consider $F_{n,\mathcal{W}}$, and $K_{\mathcal{W}}$ as random variables, then the following entropies can be written

$$H(F_{n,\mathcal{W}}) = |F_{n,\mathcal{W}}|L, \quad H(K_{\mathcal{W}}) = |K_{\mathcal{W}}|L, \tag{6}$$

where $|F_{n,\mathcal{W}}|$, and $|K_{\mathcal{W}}|$ are the size of the partition $F_{n,\mathcal{W}}$, and the keys $K_{\mathcal{W}}$ normalized by the file size L.

For the file size constraint, we have

$$\begin{aligned} N = \frac{1}{L} H(F_{[1:N]}) &\overset{(a)}{=} \frac{1}{L} \sum_{n=1}^{N} \sum_{\mathcal{W} \subseteq [1:K]} H(F_{n,\mathcal{W}}) \\ &\overset{(b)}{=} \sum_{t=0}^{K} \sum_{n=1}^{N} \sum_{\substack{\mathcal{W} \subseteq [1:K]: \\ |\mathcal{W}|=t}} |F_{n,\mathcal{W}}| = \sum_{t=0}^{K} x_t, \end{aligned} \tag{7}$$

where (a), and (b) follow from (4), and (6), respectively, and $x_t \geq 0$ is defined as

$$x_t \triangleq \sum_{n=1}^{N} \sum_{\mathcal{W} \subseteq [1:K]:\ |\mathcal{W}|=t} |F_{n,\mathcal{W}}|, \quad t \in [0 : K]. \tag{8}$$

For the cache memory constraint, we have,

$$KM \geq \frac{1}{L} \sum_{k=1}^{K} H(Z_k) \overset{(a)}{=} \frac{1}{L} \sum_{k=1}^{K} H\left(F_{[1:N]}^{(k)}, K^{(k)}\right)$$

$$\overset{(b)}{=} \frac{1}{L}\sum_{k=1}^{K}\left(H\left(F_{[1:N]}^{(k)}\right)+H\left(K^{(k)}\right)\right)$$

$$\overset{(c)}{=} \frac{1}{L}\sum_{n=1}^{N}\sum_{\substack{\mathcal{W}\subseteq[1:K]:\\ \mathcal{W}\neq\phi}}|\mathcal{W}|\,H(F_{n,\mathcal{W}})+\frac{1}{L}\sum_{\substack{\mathcal{W}\subseteq[1:K]:\\ \mathcal{W}\neq\phi}}|\mathcal{W}|\,H(K_{\mathcal{W}})$$

$$\overset{(d)}{=} \sum_{t=1}^{K} t\left(\sum_{n=1}^{N}\sum_{\substack{\mathcal{W}\subseteq[1:K]:\\ |\mathcal{W}|=t}}|F_{n,\mathcal{W}}|+\sum_{\substack{\mathcal{W}\subseteq[1:K]:\\ |\mathcal{W}|=t}}|K_{\mathcal{W}}|\right)$$

$$\overset{(e)}{=} \sum_{t=0}^{K} t x_t+\sum_{t=1}^{K} t y_t, \tag{9}$$

where (a) follows from (3), (b) is due to the fact that the keys are independent of the files, (c) follows from (4), and the fact that when summing the contents in the caches, the partition $F_{j,\mathcal{W}}$ is summed $|\mathcal{W}|$ number of times, (d) follows from (6), and $y_t \geq 0$ in (e) is defined as

$$y_t \triangleq \sum_{\mathcal{W}\subseteq[1:K]:\,|\mathcal{W}|=t}|K_{\mathcal{W}}|, \quad t\in[0:K]. \tag{10}$$

Data Delivery Phase: Each user k reveals its request/demand d_k which can be any of the N files. The users' demand vector is given by $\mathbf{d}=(d_1,d_2,\ldots,d_K)$, where $\mathbf{d}\in[1:N]^K$. The server generates the input $X_{\mathbf{d}}$ as a function of the files and the keys

$$X_{\mathbf{d}} \triangleq \psi_{\mathbf{d}}\left(F_{[1:N]},K^{([1:K])}\right), \tag{11}$$

and sends it over the shared link to the users, where $\psi_{\mathbf{d}}$ for the demand vector $\mathbf{d}$

$$\psi_{\mathbf{d}}:\left[1:2^{L}\right]^{N}\times\mathcal{K}\to\left[1:2^{\lfloor R_{\mathbf{d}}L\rfloor}\right], \tag{12}$$

is the encoding function and $R_{\mathbf{d}}$ is the server transmission rate. A secure caching scheme consists of N^K encoding functions. Upon receiving $X_{\mathbf{d}}$, each user $k\in[1:K]$ generates an estimate of the requested file $\hat{F}_{d_k}$ as a function of its cache content Z_k, and the received messages from the server $X_{\mathbf{d}}$

$$\hat{F}_{d_k} \triangleq \mu_{\mathbf{d},k}(X_{\mathbf{d}},Z_k), \tag{13}$$

where $\mu_{\mathbf{d},k}$ is the decoding function

$$\mu_{\mathbf{d},k}:\left[1:2^{\lfloor ML\rfloor}\right]\times\left[1:2^{\lfloor R_{\mathbf{d}}L\rfloor}\right]\to\left[1:2^{L}\right]. \tag{14}$$

A caching scheme with secure delivery comprises of KN^K decoding functions. In the following, we define a secure achievable rate-memory tradeoff for a demand vector $\mathbf{d}$, and the rate-memory tradeoff for the worst case transmission rate.

Definition 1 (Achievable Rate-Memory Tradeoff): The rate-memory tradeoff $(M,R_{\mathbf{d}})$ is securely achievable for a demand vector $\mathbf{d}$ if there exists a caching scheme with secure delivery such that for some $\epsilon_L\to 0$, and $\delta_L\to 0$ as $L\to\infty$ (large enough file size L), the following two conditions are satisfied.

Decodability Constraint: Each user should be able to decode its requested file from the server transmission and its cache contents, which gives the following condition

$$H(F_k|Z_k,X_{\mathbf{d}})\leq\epsilon_L,\ \ \forall k\in[1:K]. \tag{15}$$

Secure Delivery Constraint: Messages sent over the shared link must not reveal any information about any of the requested files at the external eavesdropper, i.e.,

$$I(F_1,F_2,\ldots,F_N;X_{\mathbf{d}})\leq\delta_L. \tag{16}$$

Definition 2 (Worst-Case Transmission Rate): For any achievable scheme characterized by the functions (ϕ,ψ,μ), the rate-memory tradeoff for the worst-case transmission rate over all feasible demands $\mathbf{d}$ is defined as

$$R_{\text{wc}}^{(\phi,\psi,\mu)}(M)=\max_{\mathbf{d}\in[1:N]^K}R_{\mathbf{d}}(M). \tag{17}$$

We are interested in minimizing the worst-case transmission rate among all the caching scheme. Therefore, we can define the optimal worst-case transmission rate as

$$R_{\text{wc}}^{*}(M)=\min_{(\phi,\psi,\mu)}R_{\text{wc}}^{(\phi,\psi,\mu)}(M). \tag{18}$$

III. Main Results and Discussions

In this section, we present the main results and provide an example of a basic caching network with secure delivery constraint to explain the intuition behind the converse proof and highlight the main ideas.

Theorem 1: For a secure caching setting with K end users, a database of N files and with cache size constraint of M files at each user, the delivery and key distribution schemes presented in [18] under the constraint of uncoded cache placement and $N\geq K$ is optimal. The optimal worst-case communication load for this setting is

$$R_{wc}^{*}(M)=K\left(1-\frac{M-1}{N-1}\right)\left(\frac{1}{1+K\frac{M-1}{N-1}}\right), \tag{19}$$

for $M\in\left\{\frac{N-1}{K}t+1:t\in[0:K]\right\}$. Furthermore, for $M\notin\left\{\frac{N-1}{K}t+1:t\in[0:K]\right\}$, $R_{wc}^{*}(M)$ is the lower convex envelope of its values at $M\in\left\{\frac{N-1}{K}t+1:t\in[0:K]\right\}$.

The achievability follows from the results in [18], and hence we focus on the proof of the converse, i.e., a lower bound on $R_{\text{wc}}^{*}(M)$. The complete proof of the converse is presented in Section IV. In the following, we present the achievability and converse proofs for the example of $K=3$ users and database of $N=3$ files to highlight the key aspects and intuition behind the ideas involved in the proofs. Finally, we show that the worst case communication load $R_{\text{wc}}^{*}(M)$ in Theorem 1 is optimal for this setting.

Example 1: Consider the caching network with $K=3$ users, a database of $N=3$ files and with normalized cache memory size of $1\leq M\leq 3$ at each user. The server can communicate with the users over an insecure shared link and there is an external wiretapper intending to eavesdrop the files using the server transmission. In the following, we first state the achievability scheme for this setting and then we present the novel converse proof.

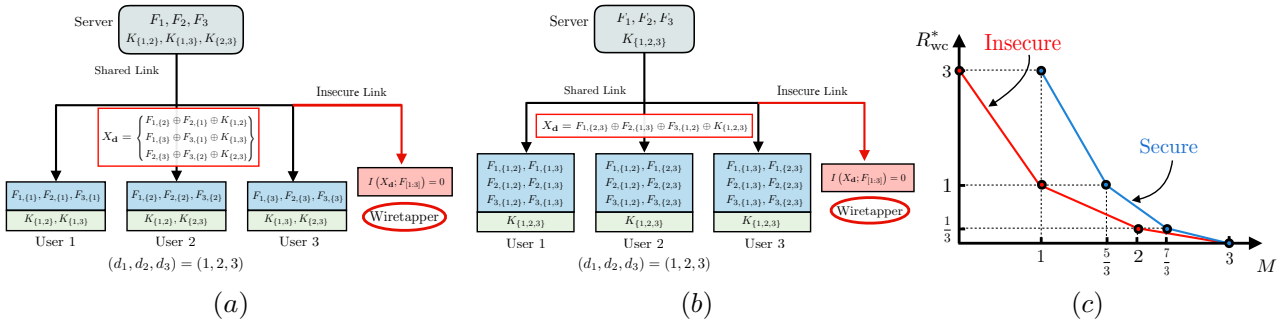

Fig. 2. Cache placement and delivery for $t = 1$ ($M = 5/3$) in (a), and $t = 2$ ($M = 7/3$) in (b). In (c), the optimal worst-case secure and non-secure rates are plotted versus the memory available at the users.

Achievability Scheme:

From Theorem 1, we have $M \in \frac{3-1}{3}t + 1$ for $t \in \{0, 1, 2\}$ which gives $M \in \{1, 5/3, 7/3, 3\}$, the possible cache sizes for each user. The cases $t = 0$ and $t = 3$ are trivial. When $t = 0$, or, equivalently $M = 1$, the server shares an individual key of size L bits with each user. Then, in the delivery phase, the server encrypts the requested file by each user using the key shared with that user and sends it over the shared link. Therefore, in this case, the pair $(M, R_{\mathbf{d}}) = (1, 3)$ is achievable. For the case $t = 3$ ($M = 3$), all the files can be stored in the cache memory of each user and subsequently $(M, R_{\mathbf{d}}) = (3, 0)$ is achievable. For $t \in \{1, 2\}$, we use the secure coded caching scheme in [18] as in Figures 2a, and 2b. Each file F_n is partitioned into $\binom{3}{t}$ fragments so that

$$F_n = \bigcup_{\mathcal{W} \subseteq [1:3] : |\mathcal{W}| = t} F_{n,\mathcal{W}}, \quad \forall n \in [1:3] \tag{20}$$

where the size of each subfile is $L/\binom{3}{t}$. In the cache placement phase, the server stores the subfiles $F_{n,\mathcal{W}}$ in the cache of user k if $k \in \mathcal{W}$. In order to satisfy the secure delivery constraint, the server generates an independent key of size $L/\binom{3}{t}$ for every subset $\mathcal{W}$ of $t+1$ users. Then, $K_{\mathcal{W}}$ is stored in the cache of user k if $k \in \mathcal{W}$. When $t = 1$, the contents of users' caches are

$$\begin{aligned} Z_1 &= \{F_{1,\{1\}}, F_{2,\{1\}}, F_{3,\{1\}}, K_{\{1,2\}}, K_{\{1,3\}}\}, \\ Z_2 &= \{F_{1,\{2\}}, F_{2,\{2\}}, F_{3,\{2\}}, K_{\{1,2\}}, K_{\{2,3\}}\}, \\ Z_3 &= \{F_{1,\{3\}}, F_{2,\{3\}}, F_{3,\{3\}}, K_{\{1,3\}}, K_{\{2,3\}}\}. \end{aligned} \tag{21}$$

Therefore, the total used memory at each user is $M = \frac{5}{3}L$ bits which does not violate the cache memory constraint. Similarly, for the case $t = 2$, the contents of each user's cache is as follows

$$\begin{aligned} Z_1 &= \{F_{1,\{1,2\}}, F_{2,\{1,2\}}, F_{3,\{1,2\}}, F_{1,\{1,3\}}, F_{2,\{1,3\}}, F_{3,\{1,3\}}, \\ &\qquad K_{\{1,2,3\}}\}, \\ Z_2 &= \{F_{1,\{1,2\}}, F_{2,\{1,2\}}, F_{3,\{1,2\}}, F_{1,\{2,3\}}, F_{2,\{2,3\}}, F_{3,\{2,3\}}, \\ &\qquad K_{\{1,2,3\}}\}, \\ Z_3 &= \{F_{1,\{1,3\}}, F_{2,\{1,3\}}, F_{3,\{1,3\}}, F_{1,\{2,3\}}, F_{2,\{2,3\}}, F_{3,\{2,3\}}, \\ &\qquad K_{\{1,2,3\}}\}, \end{aligned} \tag{22}$$

where the size of each fragment of files and each key is $L/3$, and the total used memory at each user is $M = \frac{7}{3}L$ bits. In the delivery phase, the demands are revealed to the server. Without loss of generality, let us assume that $\mathbf{d} = (1, 2, 3)$. Then, the server needs to securely deliver the fragments of file requested by that user and not stored in its cache. When $t = 1$, the server sends the messages

$$X_{\mathbf{d}} = \begin{cases} F_{1,\{2\}} \oplus F_{2,\{1\}} \oplus K_{\{1,2\}}, \\ F_{1,\{3\}} \oplus F_{3,\{1\}} \oplus K_{\{1,3\}}, \\ F_{2,\{3\}} \oplus F_{3,\{2\}} \oplus K_{\{2,3\}}, \end{cases}$$

over the shared link. The total number of bits sent by the server over the shared link in this case is $3 \times L/3 = L$ bits. Therefore, the rate-memory pair $(M, R_{\mathbf{d}}) = (5/3, 1)$ is achievable. Now, for $t = 2$, the server sends the following message over the shared link

$$X_{\mathbf{d}} = F_{1,\{2,3\}} \oplus F_{2,\{1,3\}} \oplus F_{3,\{1,2\}} \oplus K_{\{1,2,3\}}, \tag{23}$$

to the users. In this case, the total number of bits that the server sends over the shared link is $L/3$ bits and the rate-memory pair $(M, R_{\mathbf{d}}) = (7/3, 1/3)$ is achievable. In the above, we showed that the set of rate-memory pairs $\{(1, 3), (5/3, 1), (7/3, 1/3), (3, 0)\}$ are achievable. As shown in Figure 2c, using the memory-sharing, we can achieve the linear curves $6 - 3M$, $8/3 - M$, and $3/2 - M/2$ between the rate-memory pairs $\{(1, 3), (5/3, 1)\}$, $\{(5/3, 1), (7/3, 1/3)\}$, and $\{(7/3, 1/3), (3, 0)\}$, respectively.

Converse Proof:

The file size, and the cache size constraints are given in (7), and (9), respectively, for this example as

$$\sum_{t=0}^{3} x_t = 3, \tag{24}$$

$$\sum_{t=0}^{3} t x_t + \sum_{t=1}^{3} t y_t \leq 3M. \tag{25}$$

We define $\sigma : (1, 2, 3) \to (\sigma_1, \sigma_2, \sigma_3)$ as a permutation of the ordered set $(1, 2, 3)$. Using the decoding constraint in (15), we can establish the following bound

$$\begin{aligned} H(F_{[1:3]}) &= I(X_{\mathbf{d}}, Z_{[1:3]}; F_{[1:3]}) + H(F_{[1:3]} | X_{\mathbf{d}}, Z_{[1:3]}) \\ &\overset{(a)}{\leq} H(X_{\mathbf{d}}, Z_{[1:3]}) - H(X_{\mathbf{d}}, Z_{[1:3]} | F_{[1:3]}) + 3\epsilon_L \end{aligned}$$

$$
\begin{aligned}
&= H(X_{\mathbf{d}}, Z_{[1:3]}) - H(X_{\mathbf{d}}, F_{[1:3]}^{([1:3])}, K^{([1:3])}|F_{[1:3]}) + 3\epsilon_L \\
&= H(X_{\mathbf{d}}, Z_{[1:3]}) - H(K^{([1:3])}|F_{[1:3]}) \\
&\quad - H(X_{\mathbf{d}}|K^{([1:3])}, F_{[1:3]}) + 3\epsilon_L \\
&\overset{(b)}{=} H(X_{\mathbf{d}}, Z_{[1:3]}) - H(K^{([1:3])}) + 3\epsilon_L \\
&\overset{(c)}{=} H(X_{\mathbf{d}}) + H(Z_{\sigma_1}|X_{\mathbf{d}}) + H(Z_{\sigma_2}|X_{\mathbf{d}}, Z_{\sigma_1}) \\
&\quad + H(Z_{\sigma_3}|X_{\mathbf{d}}, Z_{\sigma_1}, Z_{\sigma_2}) - H(K^{([1:3])}) + 3\epsilon_L \\
&\overset{(d)}{\leq} R_{\mathbf{d}}^* L - H(K^{([1:3])}) + H(F_{[1:3]}^{(\sigma_1)}) + H(K^{(\sigma_1)}) \\
&\quad + H(F_{d_{\sigma_{[2:3]}}}^{(\sigma_2)}|F_{d_{\sigma_{[2:3]}}}^{(\sigma_1)}) + H(K^{(\sigma_2)}|K^{(\sigma_1)}) \\
&\quad + H(F_{d_{\sigma_3}}^{(\sigma_3)}|F_{d_{\sigma_3}}^{(\sigma_{[1:2]})}) + H(K^{(\sigma_3)}|K^{(\sigma_{[1:2]})}) + 3\epsilon_L \\
&\overset{(e)}{=} R_{\mathbf{d}}^* L - H(K^{([1:3])}) + H(K^{([1:3])}) + H(F_{d_{\sigma_1}}|F_{d_{\sigma_1}}^{(\sigma_1)}) \\
&\quad + H(F_{d_{\sigma_2}}|F_{d_{\sigma_2}}^{(\sigma_{[1:2]})}) + H(F_{d_{\sigma_3}}|F_{d_{\sigma_3}}^{([1:3])}) + 3\epsilon_L, \qquad (26)
\end{aligned}
$$

where (a) follows from the decodability constraint in (15), (b) is due to the facts that the keys are independent of files, and $X_{\mathbf{d}}$ is a function of the files and keys as given in (11), (c) from the chain rule of entropy, (d) follows from the cache content in (3), and the fact that conditioning reduced entropy, and (e) follows from the chain rule of entropy. Hence, following Remark 1 we obtain a lower bound on R_{wc}^* as follows

$$
\begin{aligned}
R_{\text{wc}}^* + 3\epsilon_L/L &\geq \frac{1}{L}H(F_{d_{\sigma_1}}|F_{d_{\sigma_1}}^{(\sigma_1)}) + \frac{1}{L}H(F_{d_{\sigma_2}}|F_{d_{\sigma_2}}^{(\sigma_{[1:2]})}) \\
&\quad + \frac{1}{L}H(F_{d_{\sigma_3}}|F_{d_{\sigma_3}}^{([1:3])}) \\
&\overset{(a)}{=} \frac{1}{L}H(F_{d_1,\phi}, F_{d_1,\sigma_2}, F_{d_1,\sigma_3}, F_{d_1,\sigma_{[2:3]}}) \\
&\quad + \frac{1}{L}H(F_{d_2,\phi}, F_{d_2,\sigma_3}) + \frac{1}{L}H(F_{d_3,\phi}) \\
&\overset{(b)}{=} |F_{d_1,\phi}| + |F_{d_1,\sigma_2}| + |F_{d_1,\sigma_3}| + |F_{d_1,\sigma_{[2:3]}}| \\
&\quad + |F_{d_2,\phi}| + |F_{d_2,\sigma_3}| + |F_{d_3,\phi}|, \qquad (27)
\end{aligned}
$$

where (a), and (b) follow from (4), and (6), respectively.

For all the 3! permutations σ, and all the 3! possible demand vectors $\mathbf{d}$, we get $3! \times 3! = 36$ lower bounds. Summing them up, we obtain the following bound

$$
\begin{aligned}
36R_{\text{wc}}^* &\geq 36\sum_{j=1}^{3}(|F_{j,\phi}|) + 12\sum_{j=1}^{3}(|F_{j,1}| + |F_{j,2}| + |F_{j,3}|) \\
&\quad + 4\sum_{j=1}^{3}\left(|F_{j,\{1,2\}}| + |F_{j,\{2,3\}}| + |F_{j,\{1,3\}}|\right) - 3 \times 36\epsilon_L/L, \qquad (28)
\end{aligned}
$$

which can also be written using the definition in (8) as

$$
R_{\text{wc}}^* \geq x_0 + \frac{1}{3}x_1 + \frac{1}{9}x_2 - 3\epsilon_L/L. \qquad (29)
$$

For secure delivery, we obtain the following constraint for some demand vector $\mathbf{d} = (d_1, d_2, d_3)$

$$
\begin{aligned}
H(F_{d_j}|F_{[1:3]}^{(j)}) &= I(F_{d_j}; K^{(j)}, X_{\mathbf{d}}|F_{[1:3]}^{(j)}) + H(F_{d_j}|Z_j, X_{\mathbf{d}}) \\
&\overset{(a)}{\leq} I(F_{d_j}; X_{\mathbf{d}}|F_{[1:3]}^{(j)}) + I(F_{d_j}; K^{(j)}|X_{\mathbf{d}}, F_{[1:3]}^{(j)}) + \epsilon_L \\
&\overset{(b)}{\leq} \delta_L + H(K^{(j)}|X_{\mathbf{d}}, F^{(j)}) + \epsilon_L \leq H(K^{(j)}) + \delta_L + \epsilon_L, \qquad (30)
\end{aligned}
$$

which is also known as the Shannon constraint for perfect secrecy [22], where the size of the message needed to be transmitted securely to a user j, i.e., $H(F_{d_j}|F_{[1:3]}^{(j)})$, must be less than the size of the keys stored at that user, i.e., $H(K^{(j)})$, where (a) follows from the decodability constraint in (15), and (b) is obtained using the secrecy constraint in (16). Therefore, from (4), and (6) we have

$$
\sum_{\mathcal{W}\subseteq[1:3]\setminus j}|F_{d_j,\mathcal{W}}| \leq \sum_{\mathcal{W}\subseteq[1:3]:\, j\in\mathcal{W}}|K_{\mathcal{W}}| + \frac{\delta_L + \epsilon_L}{L}. \qquad (31)
$$

This bound can be obtained for $d_j \in [1:3]$, and $j \in [1:3]$ to get 9 different constraints, and then by summing them up we arrive to

$$
\sum_{j=1}^{3}\sum_{\mathcal{W}\subseteq[1:3]\setminus j}\sum_{k=1}^{3}|F_{k,\mathcal{W}}| \leq 3\sum_{j=1}^{3}\sum_{\substack{\mathcal{W}\subseteq[1:3]:\\ j\in\mathcal{W}}}|K_{\mathcal{W}}| + 9\frac{\delta_L + \epsilon_L}{L}, \qquad (32)
$$

which can be simplified using (8), and (10) to

$$
3x_0 + 2x_1 + x_2 \leq 3(y_1 + 2y_2 + 2y_3) + 9\frac{\delta_L + \epsilon_L}{L}. \qquad (33)
$$

By using (33) in (25), we get

$$
\begin{aligned}
3M + 3\frac{\delta_L + \epsilon_L}{L} &\geq x_1 + 2x_2 + 3x_3 + \frac{1}{3}(3x_0 + 2x_1 + x_2) \\
&= \frac{1}{3}(3x_0 + 5x_1 + 7x_2 + 9x_3) \\
&\overset{(a)}{=} 3 + \frac{2}{3}(x_1 + 2x_2 + 3x_3), \qquad (34)
\end{aligned}
$$

where (a) follows directly from (24), which gives the following constraint

$$
x_1 + 2x_2 + 3x_3 \leq \frac{9}{2}(M-1) + \frac{9}{2}\frac{\delta_L + \epsilon_L}{L}, \qquad (35)
$$

To summarize, the constraints (24), (29), and (35) provide 3 different lower bounds over R_{wc}^* by eliminating the variable pairs (x_0, x_1), (x_1, x_2), and (x_2, x_3). To eliminate (x_0, x_1), we first substitute by x_0 in (24) into (29) to get the following constraint

$$
R_{\text{wc}}^* + 3\epsilon_L/L \geq 3 - \frac{2}{3}x_1 - \frac{8}{9}x_2 - x_3. \qquad (36)
$$

Next, by bounding x_1 using (35), we can bound R_{wc}^* in (36) as follows

$$
\begin{aligned}
R_{\text{wc}}^* + 3\frac{\delta_L + \epsilon_L}{L} &\geq 6 - 3M + \frac{4}{9}x_2 + x_3 \\
&\overset{(a)}{\geq} 6 - 3M, \qquad (37)
\end{aligned}
$$

where (a) since $x_2, x_3 \geq 0$, which gives the bound $R_{\text{wc}}^* \geq 6 - 3M$ as $L \to \infty$. Similarly, by eliminating the pairs (x_1, x_2), and (x_2, x_3), we obtain two more bounds $R_{\text{wc}}^* \geq 8/3 - M$, and $R_{\text{wc}}^* \geq 3/2 - M/2$, respec-

tively, which match the achievable scheme of [18] and is shown in Figure 2c. For comparison, we also plot the optimal rate-memory tradeoff for the insecure setting (i.e., no secrecy constraint) in Figure 2c.

IV. General Proof of Converse

In this section, we present an information theoretic lower bound for the worst-case communication rate for any N, and K where $N \geq K$. We show that the lower bound matches with the upper bound in (19) and the delivery and key distribution schemes introduced in [18] for secure delivery caching is optimal under the uncoded cache placement.

*Remark 1 (**Basic idea for the converse**):* We assume a sequence of different demands. A lower bound over the optimal rate for a certain demand vector $\mathbf{d}$, $R^*_{\mathbf{d}}$, serves also as a lower bound for the worst-case since the optimal worst-case rate is larger than or equal to the rate for any demand, i.e., $R^*_{\text{wc}} \geq R^*_{\mathbf{d}}$. We then average out all the obtained lower bounds based on the chosen demands The novel part in our proof is choosing the right demands which lead to the optimal lower bound.

We start by considering the demand vector $\mathbf{d} = (d_1, d_2, \ldots, d_K)$ of different files, such that $d_i \in [1:N]$, and $d_i \neq d_j$, $\forall i \neq j$. This is true when $N \geq K$. We also consider a permutation of the user indexes given as $\sigma : (1:K) \to (\sigma_1, \ldots, \sigma_K)$, the user of index σ_k is requesting the file $F_{d_{\sigma_k}}$. We first use the decodability constraint in (15) to get the following bound:

$$
\begin{aligned}
H(F_{[1:N]}) &= I(F_{[1:N]}; X_{\mathbf{d}}, Z_{[1:K]}) + H(F_{[1:N]}|X_{\mathbf{d}}, Z_{[1:K]}) \\
&\leq I(F_{[1:N]}; X_{\mathbf{d}}, Z_{[1:K]}) + H(F_{[1:N]}|X_{\mathbf{d}}, Z_{[1:K]}, F_{\mathbf{d}}) + K\epsilon_L \\
&\leq I(F_{[1:N]}; X_{\mathbf{d}}, Z_{[1:K]}) + H(F_{[1:N]\setminus\mathbf{d}}|Z_{[1:K]}) + K\epsilon_L \\
&\leq H(X_{\mathbf{d}}, Z_{[1:K]}) - H(X_{\mathbf{d}}, Z_{[1:K]}|F_{[1:N]}) \\
&\quad + H(F_{[1:N]\setminus\mathbf{d}}|F^{([1:K])}_{[1:N]\setminus\mathbf{d}}) + K\epsilon_L \\
&= H(X_{\mathbf{d}}, Z_{[1:K]}) - H(K^{([1:K])}) \\
&\quad + H(F_{[1:N]\setminus\mathbf{d}}|F^{([1:K])}_{[1:N]\setminus\mathbf{d}}) + K\epsilon_L \\
&= H(X_{\mathbf{d}}) + \sum_{i=1}^{K} H(Z_{\sigma_i}|Z_{\sigma_{[1:i-1]}}, X_{\mathbf{d}}) - H(K^{([1:K])}) \\
&\quad + H(F_{[1:N]\setminus\mathbf{d}}|F^{([1:K])}_{[1:N]\setminus\mathbf{d}}) + K\epsilon_L \\
&\leq R^*_{\mathbf{d}}L + \sum_{i=1}^{K} H(Z_{\sigma_i}|Z_{\sigma_{[1:i-1]}}, X_{\mathbf{d}}, F_{[d_{\sigma_1}:d_{\sigma_{i-1}}]}) \\
&\quad - H(K^{([1:K])}) + H(F_{[1:N]\setminus\mathbf{d}}|F^{([1:K])}_{[1:N]\setminus\mathbf{d}}) + K\epsilon_L \\
&\leq R^*_{\mathbf{d}}L + \sum_{i=1}^{K} H(F^{(\sigma_i)}_{[1:N]\setminus d_{\sigma_{[1:i-1]}}}, K^{(\sigma_i)}|Z_{\sigma_{[1:i-1]}}) \\
&\quad - H(K^{([1:K])}) + H(F_{[1:N]\setminus\mathbf{d}}|F^{([1:K])}_{[1:N]\setminus\mathbf{d}}) + K\epsilon_L \\
&\leq R^*_{\mathbf{d}}L + \sum_{i=1}^{K} \sum_{j\in[1:N]\setminus d_{\sigma_{[1:i-1]}}} H(F^{(\sigma_i)}_j|F^{(\sigma_{[1:i-1]})}_j) \\
&\quad + \sum_{i=1}^{K} H(K^{(\sigma_i)}|K^{(\sigma_{[1:i-1]})}) - H(K^{([1:K])}) \\
&\quad + H(F_{[1:N]\setminus\mathbf{d}}|F^{([1:K])}_{[1:N]\setminus\mathbf{d}}) + K\epsilon_L \\
&= R^*_{\mathbf{d}}L + \sum_{j\in[1:N]\setminus\mathbf{d}} \sum_{i=1}^{K} H(F^{(\sigma_i)}_j|F^{(\sigma_{[1:i-1]})}_j) + K\epsilon_L \\
&\quad + \sum_{j=1}^{K}\sum_{i=1}^{j} H(F^{(\sigma_i)}_{d_{\sigma_j}}|F^{(\sigma_{[1:i-1]})}_{d_{\sigma_j}}) + H(F_{[1:N]\setminus\mathbf{d}}|F^{([1:K])}_{[1:N]\setminus\mathbf{d}}) \\
&= R^*_{\mathbf{d}}L + H(F^{([1:K])}_{[1:N]\setminus\mathbf{d}}) + \sum_{j=1}^{K} H(F^{(\sigma_{[1:j]})}_{d_{\sigma_j}}) \\
&\quad + H(F_{[1:N]\setminus\mathbf{d}}|F^{([1:K])}_{[1:N]\setminus\mathbf{d}}) + K\epsilon_L \\
&= R^*_{\mathbf{d}}L + H(F_{[1:N]\setminus\mathbf{d}}) + \sum_{j=1}^{K} H(F^{(\sigma_{[1:j]})}_{d_{\sigma_j}}) + K\epsilon_L, \qquad (38)
\end{aligned}
$$

which gives a lower bound over $R^*_{\mathbf{d}}$, which is also a lower bound over R^*_{wc} following Remark 1 as follows

$$
\begin{aligned}
R^*_{\text{wc}} + K\epsilon_L/L &\geq \frac{1}{L}\sum_{j=1}^{K} H\left(F_{d_{\sigma_j}}|F^{(\sigma_{[1:j]})}_{d_{\sigma_j}}\right) \\
&\overset{(a)}{=} \sum_{j=1}^{K} \sum_{\mathcal{W}\subseteq[\sigma_{j+1}:\sigma_K]} |F_{d_{\sigma_j},\mathcal{W}}|, \qquad (39)
\end{aligned}
$$

where (a) follows from (4), and (6). Next, for each j we obtain N different bounds for $d_j \in [1:N]$. Then, by summing up all the N bounds we have

$$
R^*_{\text{wc}} + K\epsilon_L/L \geq \frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{K} \sum_{\mathcal{W}\subseteq[\sigma_{j+1}:\sigma_K]} |F_{n,\mathcal{W}}|. \qquad (40)
$$

As we obtain the above bound for some permutation of the users σ, we can obtain a total of $K!$ different bounds for all different permutations, and then average them to get the following bound

$$
\begin{aligned}
R^*_{\text{wc}} + K\epsilon_L/L &\geq \frac{1}{NK!}\sum_{n=1}^{N}\sum_{j=1}^{K} \sum_{\sigma\in[K!]} \sum_{\mathcal{W}\subseteq[\sigma_{j+1}:\sigma_K]} |F_{n,\mathcal{W}}| \\
&= \frac{1}{NK!}\sum_{t=0}^{K}\sum_{n=1}^{N}\sum_{j=1}^{K} \sum_{\sigma\in[K!]} \sum_{\substack{\mathcal{W}\subseteq[\sigma_{j+1}:\sigma_K]: \\ |\mathcal{W}|=t}} |F_{n,\mathcal{W}}|. \qquad (41)
\end{aligned}
$$

Due to symmetry, for each value of $t \in [0:K]$ in the above summation, the coefficients of $|F_{j,\mathcal{W}}|$ are equal for each $j \in [1:N]$, and $|\mathcal{W}| = t$. Therefore, the coefficient of $|F_{1,[1:t]}|$ is the same as the coefficient of $x_t \geq 0$ defined as

$$
x_t = \sum_{j=1}^{N} \sum_{\mathcal{W}\subseteq[1:K]:\ |\mathcal{W}|=t} |F_{j,\mathcal{W}}|, \quad t \in [0:K]. \qquad (42)
$$

Next, we find the coefficient of $|F_{1,[1:t]}|$. We first notice that for a fixed value of t, and $n = 1$, to obtain $[1:t] \subseteq \sigma_{[j+1:K]}$, we must have $K - j \geq t$. Then, we have $\sigma_1, \sigma_2, \ldots, \sigma_j \notin [1:t]$, which gives the number of permutations σ where $[1:t] \subseteq \sigma_{[j+1:K]}$ given by $\frac{K-t!}{K-t-j!}K - j! = \binom{K-j}{t}/\binom{K}{t}$.

Therefore the coefficient of $|F_{1,[1:t]}|$ (hence x_t) is given as

$$\frac{1}{NK!}\sum_{j=1}^{K-t}\binom{K-j}{t}\Big/\binom{K}{t} = \frac{1}{NK!}\sum_{i=t}^{K-1}\binom{i}{t}\Big/\binom{K}{t} \overset{(a)}{=} \frac{1}{NK!}\binom{K}{t+1}\Big/\binom{K}{t} = \frac{K-t}{N(t+1)}, \tag{43}$$

where (a) follows from the Pascal's triangle. Therefore, we obtain the following bound

$$R^*_{\text{wc}} + K\epsilon_L/L \geq \frac{1}{N}\sum_{t=0}^{K}\frac{K-t}{t+1}x_t. \tag{44}$$

For secure delivery, we get the following constraint in steps similar to (30) for a demand vector $\mathbf{d}$:

$$\begin{aligned} H(F_{d_j}|F^{(j)}_{[1:N]}) &= I(F_{d_j};K^{(j)},X_{\mathbf{d}}|F^{(j)}_{[1:N]}) + H(F_{d_j}|Z_j,X_{\mathbf{d}}) \\ &\overset{(a)}{\leq} I(F_{d_j};X_{\mathbf{d}}|F^{(j)}_{[1:N]}) + I(F_{d_j};K^{(j)}|X_{\mathbf{d}},F^{(j)}_{[1:N]}) + \epsilon_L \\ &\overset{(b)}{\leq} \delta_L + H(K^{(j)}|X_{\mathbf{d}},F^{(j)}) + \epsilon_L \leq H(K^{(j)}) + \delta_L + \epsilon_L, \end{aligned} \tag{45}$$

where (a) follows from the decodability constraint in (15), and (b) follows from the secrecy constraint in (16). Therefore, from (4), and (6) we have

$$\sum_{\mathcal{W}\subseteq[1:K]\setminus j}|F_{d_j,\mathcal{W}}| \leq \sum_{\mathcal{W}\subseteq[1:K]:\, j\in\mathcal{W}}|K_{\mathcal{W}}| + \frac{\delta_L+\epsilon_L}{L}. \tag{46}$$

Summing up over $j\in[1:K]$ we obtain

$$\begin{aligned} \sum_{j=1}^{K}\sum_{\mathcal{W}\subseteq[1:K]\setminus j}|F_{d_j,\mathcal{W}}| - K\frac{\delta_L+\epsilon_L}{L} &\leq \sum_{j=1}^{K}\sum_{\mathcal{W}\subseteq[1:K]:\, j\in\mathcal{W}}|K_{\mathcal{W}}| \\ &= \sum_{t=1}^{K} t y_t. \end{aligned} \tag{47}$$

Now, summing up over $d_j\in[1:N]$,

$$\sum_{t=1}^{K} t y_t \geq \frac{1}{N}\sum_{t=0}^{K}\sum_{j=1}^{K}\sum_{n=1}^{N}\sum_{\substack{\mathcal{W}\subseteq[1:K]\setminus j:\\ |\mathcal{W}|=t}}|F_{n,\mathcal{W}}| - K\frac{\delta_L+\epsilon_L}{L}. \tag{48}$$

Due to symmetry, finding the coefficient of x_t in (48) is equivalent to finding the coefficient of $|F_{1,[1:t]}|$, which is non-zero only when $j\notin[1:t]$. Therefore, we obtain

$$\begin{aligned} \sum_{t=1}^{K} t y_t + K\frac{\delta_L+\epsilon_L}{L} &\geq \frac{1}{N}\sum_{t=0}^{K}\sum_{j=t+1}^{K} x_t \\ &= \sum_{t=0}^{K}\frac{K-t}{N}x_t \\ &\overset{(a)}{=} K - \frac{1}{N}\sum_{t=0}^{K} t x_t, \end{aligned} \tag{49}$$

where (a) follows from the file size constraint in (7). Using (49) in the cache size constraint in (9), we arrive at the following constraint:

$$\sum_{t=0}^{K} t x_t \leq NK\frac{M-1}{K-1} + \frac{KN}{N-1}\frac{\delta_L+\epsilon_L}{L}. \tag{50}$$

Finally, letting $L\to\infty$, then $\epsilon_L/L\to 0$, and $\delta_L/L\to 0$ in (44), and (50). Therefore, we have the following two constraints as well as the file size constraint in (7)

$$R^*_{\text{wc}} \geq \frac{1}{N}\sum_{t=0}^{K}\frac{K-t}{t+1}x_t, \tag{51}$$

$$\sum_{t=0}^{K} t x_t \leq NK\frac{M-1}{K-1}. \tag{52}$$

To derive the necessary lower bounds on the optimal worst-case transmission rate R^*_{wc}, we eliminate the pairs (x_j, x_{j+1}), for each $j\in[0:K-1]$, in the equation (51) using the equations (7) and (52). First, we use (7) to write x_j as follows

$$x_j = N - \sum_{t\in[0:K]\setminus j} x_t, \tag{53}$$

and use it in the bounds (51) and (52) to obtain

$$\begin{aligned} R^*_{\text{wc}} &\geq \frac{1}{N}\sum_{t\in[0:K]\setminus j}\frac{K-t}{t+1}x_t + \frac{K-j}{N(j+1)}\left(N - \sum_{t\in[0:K]\setminus j} x_t\right) \\ &= \frac{K-j}{j+1} + \frac{1}{N}\sum_{t\in[0:K]\setminus j}\left(\frac{K-t}{t+1} - \frac{K-j}{j+1}\right)x_t, \end{aligned} \tag{54}$$

and

$$\begin{aligned} \sum_{t\in[0:K]\setminus j} t x_t + j\left(N - \sum_{t\in[0:K]\setminus j} x_t\right) &\leq KN\left(\frac{M-1}{N-1}\right) \\ \sum_{t\in[0:K]\setminus j}(t-j)\,x_t &\leq -jN + KN\left(\frac{M-1}{N-1}\right). \end{aligned} \tag{55}$$

Now, we need to eliminate x_{j+1} from (54). We use (55) to bound x_{j+1} as

$$x_{j+1} \leq KN\left(\frac{M-1}{N-1}\right) - jN - \sum_{t\in[0:K]\setminus\{j,j+1\}}(t-j)\,x_t. \tag{56}$$

Then, we use this bound in (54) as follows

$$\begin{aligned} R^*_{\text{wc}} &\geq \frac{K-j}{j+1} + \frac{1}{N}\sum_{t\in[0:K]\setminus\{j,j+1\}}\left(\frac{K-t}{t+1} - \frac{K-j}{j+1}\right)x_t \\ &\quad - \frac{1}{N}\left(\frac{K-j}{j+1} - \frac{K-j-1}{j+2}\right)x_{j+1} \\ &\overset{(a)}{\geq} \frac{K-j}{j+1} + \frac{K(K+1)}{(j+1)(j+2)}\left(j - \frac{M-1}{N-1}\right) \\ &\quad + \frac{1}{N}\sum_{t\in[0:K]\setminus\{j,j+1\}}\lambda_t x_t \\ &\overset{(b)}{\geq} \frac{K-j}{j+1} + \frac{K(K+1)}{(j+1)(j+2)}\left(j - \frac{M-1}{N-1}\right) \end{aligned} \tag{57}$$

where (a) follows because the coefficient of x_{j+1} is negative for all $j \in [0:K-1]$, and (b) since the coefficient, λ_t, of $x_t > 0$ is positive, which can be shown in the following:

$$\begin{aligned}\lambda_t &= \frac{K-t}{t+1} - \frac{K-j}{j+1} + \frac{K+1}{(j+1)(j+2)}(j-t) \\ &= \frac{(K+1)(j-t)(j-t+1)}{(j+1)(j+2)(t+1)}, \end{aligned} \tag{58}$$

where $K+1, j+1, j+2, t+1 > 0$ for $t \in [0:K-1]$, then we only need to show that $(j-t)(j-t+1) > 0$ for $t \in [0:K] \setminus \{j, j+1\}$. This can be easily checked by assuming $y = j - t$, then $y(y+1)$ is only negative in the range $-1 < y < 0$, or $j < t < j+1$, which is not in the range of t in the summation.

In order to show that the lower bound in (57) matches with the achievable rate given by (19), we consider two values of memory, $M_1 = \frac{N-1}{K} j + 1$, and $M_2 = \frac{N-1}{K}(j+1) + 1$, for some $j \in [0:K-1]$, which gives the achievable rates $R_1 = \frac{K-j}{j+1}$, and $R_2 = \frac{K-j-1}{j+2}$. By memory sharing, the line joining these two achievable points is also achieved, which gives the following upper bounds over R^*_{wc}

$$R^*_{\text{wc}} \leq \frac{K-j}{j+1} + \frac{K(K+1)}{(j+1)(j+2)} \left(j - \frac{M-1}{N-1} \right), \tag{59}$$

for $j \in [0:K-1]$, and $M_1 \leq M \leq M_2$, which matches the lower bounds in (57), and completes the proof of Theorem 1

V. Conclusions

In this paper, we considered the secure delivery problem for the uncoded caching problem. We characterized the information theoretically optimal rate-memory trade-off, where the number of users is no larger than the number of files in the server. The achievability part followed the scheme developed in [18], while the converse proof followed a novel bounding methodology similar to the recent result in [5], where our novel contribution was to add the secure delivery constraint. Future directions include extending our result to the case when the number of users can be larger than the number of files.

References

[1] Cisco, "The Zettabyte Era: Trends and Analysis," Technical Report 2014.

[2] L. W. Dowdy and D. V. Foster, "Comparative models of the file assignment problem," *ACM Computing Surveys*, vol. 14, no. 2, pp. 287–313, Jun. 1982.

[3] D. Wessels, *Web Caching: Reducing Network Traffic*. Ed. O'Reilly, 2001.

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

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

[6] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," *CoRR*, vol. abs/1609.07817, 2016. [Online]. Available: http://arxiv.org/abs/1609.07817

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

[8] A. Sengupta and R. Tandon, "Improved approximation of storage-rate tradeoff for caching with multiple demands," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 1940–1955, May 2017.

[9] C. Tian, "Symmetry, demand types and outer bounds in caching systems," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 825–829.

[10] 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, June 2016.

[11] S. Wang, W. Li, X. Tian, and H. Liu, "Fundamental limits of heterogenous cache," *CoRR*, vol. abs/1504.01123, 2015. [Online]. Available: http://arxiv.org/abs/1504.01123

[12] M. M. Amiri, Q. Yang, and D. Gündüz, "Decentralized coded caching with distinct cache capacities," *CoRR*, vol. abs/1611.01579, 2016. [Online]. Available: http://arxiv.org/abs/1611.01579

[13] J. Zhang, X. Lin, C. C. Wang, and X. Wang, "Coded caching for files with distinct file sizes," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1686–1690.

[14] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of distributed caching in D2D wireless networks," in *Proc. IEEE Information Theory Workshop (ITW)*, Sept 2013, pp. 1–5.

[15] 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, Aug 2015.

[16] ——, "Coding for caching: fundamental limits and practical challenges," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 23–29, August 2016.

[17] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 4, pp. 2522–2545, May 2016.

[18] 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, Feb 2015.

[19] V. Ravindrakumar, P. Panda, N. Karamchandani, and V. Prabhakaran, "Fundamental limits of secretive coded caching," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 425–429.

[20] A. A. Zewail and A. Yener, "Coded caching for resolvable networks with security requirements," in *Proc. IEEE Conference on Communications and Network Security (CNS)*, Oct 2016, pp. 621–625.

[21] Z. H. Awan and A. Sezgin, "Fundamental limits of caching in D2D networks with secure delivery," in *Proc. IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 464–469.

[22] C. Shannon, "Communication Theory of Secrecy Systems," *Bell Labs Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.