Title

On the Necessity of Non-Shannon Information Inequalities for Storage Overhead Constrained PIR and Network Coding

Authors
Sun, Hua
Jafar, Syed

Publication Date
2018-10-24

On the Necessity of Non-Shannon Information Inequalities for Storage Overhead Constrained PIR and Network Coding

Hua Sun and Syed A. Jafar

Abstract

We show that to characterize the capacity of storage overhead constrained private information retrieval (PIR) with only 2 messages, and 2 databases, non-Shannon information inequalities are necessary.

As a by-product of this result, we construct the smallest instance, to our knowledge, of a network coding capacity problem that requires non-Shannon inequalities.

Hua Sun (email: huas2@uci.edu) and Syed A. Jafar (email: syed@uci.edu) are with the Center of Pervasive Communications and Computing (CPCC) in the Department of Electrical Engineering and Computer Science (EECS) at the University of California Irvine.

1 Introduction

There is much recent interest in exploring the fundamental limits of private information retrieval (PIR) protocols. PIR originates from theoretical computer science and its main theme is to protect the user's intention or preference, in contrast to the common theme of protecting the content of information in cryptography. The importance of PIR is further understored due to the discovery of its fundamental connections to a broad range of questions across a diverse array of fields, such as blind interference alignment, locally decodable codes, batch codes, and secure multi-party computation.

PIR is the problem of retrieving one out of K messages from N distributed databases (each stores all K message) in a way that each database learns no information about which message is being retrieved in the information theoretic sense. The rate of a PIR scheme is the ratio of the number of bits of the desired message over the total number of bits downloaded from all databases. The maximum rate is the capacity of PIR. The capacity of PIR and its variants is characterized recently, including LPIR, TPIR, RPIR, TRPIR, SPIR, MDS-PIR, MDS-SPIR, MPIR, MTPIR. Notably, the capacity of MDS-TPIR is still open and a conjectured is made in [1].

Precise capacity characterizations are rare not only in the broad area of network information theory but also within the special class of (noiseless) network coding problems for a variety of reasons. A common hurdle in finding optimal achievable schemes is the lack of understanding of non-linear coding schemes, which are often necessary [2, 3, 4]. A similar challenge on the converse side lies in the need for non-Shannon information inequalities [5, 6, 7, 8], of which there are infinitely many that remain unknown [9, 10]. Incidentally, the task of finding the strongest possible lower and upper bound on capacity, even when restricted to linear coding schemes and Shannon inequalities, quickly becomes prohibitively complex for larger networks due to the explosive growth in the number of parameters.

In light of the challenges, it is quite remarkable that precise capacity characterizations have been possible for the PIR problem [11], and for its variants such as LPIR [12], TPIR [13], RPIR [13], SPIR [14], MDS-PIR [15] and MDS-SPIR [16], MPIR [17], MTPIR [17]. Indeed, for these capacity results, linear coding schemes are sufficient, only Shannon information inequalities are required and even when the network size becomes arbitrarily large, the complexity of the proofs is kept in check by the symmetries inherent in the problem that facilitate inductive and recursive reasoning. It is a matter of some curiosity as to how far such a pattern of fortuitous outcomes could continue for capacity studies of other variants of PIR. Motivated by this curiosity, in this work we explore the capacity of an important variants of PIR – storage *overhead* constrained PIR (OPIR).

1.1 OPIR

Classical PIR assumes replicated databases, i.e., each database stores all the messages. For larger datasets, replication schemes incur substantial storage costs. Coding has been shown to be an effective way to reduce the storage costs in distributed data storage systems. Applications of coding to reduce the storage overhead for PIR have attracted attention recently [18, 19, 20, 21, 22, 23, 24, 15, 25, 16]. The approaches of these works fall into two decomposition based ones. One approach trivializes the storage code design and then studies optimal PIR protocols. That is, we fix the storage code and study the fundamental limits of all possible PIR schemes subject to the given storage code. One particular case that has been studied extensively is that each message is separately MDS-coded. Notably, subject to such separate MDS-coded storage, the capacity of both PIR and SPIR (MDS-PIR and MDS-SPIR problem) has been characterized [15, 16]. The

MDS-TPIR problem is studied in [21] and [1]. Intriguingly, the capacity of MDS-TPIR is still open and is conjectured in [1]. The other approach is to focus only on the storage codes and combine the codes with known PIR protocols. That is, we focus on the design of a special class of storage code that can be combined with any *linear* PIR protocols. The defining property of this class of storage code is that for each message, there exists a number of disjoint decoding groups (of databases), so that for each group, we may send the same query to the databases in that group (such that privacy is preserved because each query is independent of the desired message index), then because of the linearity of the PIR protocols, we can combine the answers using the decoding rule of the storage code to obtain the desired message (such that correctness is guaranteed as well). One downside of this approach is that when we have a small number of databases (say 2), the choices of separating the databases to disjoint groups are limited. Another fundamental hurdle may be the necessity of non-linear PIR protocols, which is not known yet, however, [17] contains an example where non-linear PIR protocols are useful, although in an extended form of PIR - MPIR. The strong aspect of this approach is that all linear PIR protocols can be used (even those that have not been invented yet) to combine with qualified storage codes (such a separation based approach is also flexible).

To understand the fundamental limits without restricting the class of coding schemes or to known PIR protocols, we consider the capacity of PIR with storage overhead constraints (OPIR). Here our contribution is based on the simple $K = N = 2$ setting, where we show that non-Shannon inequalities are necessary to characterize the capacity of OPIR.

1.2 Network Coding

As a by-product of the result for OPIR, we find the simplest (to our knowledge) example of a network coding capacity problem where non-Shannon information inequalities are necessary. This open problem could be of independent interest.

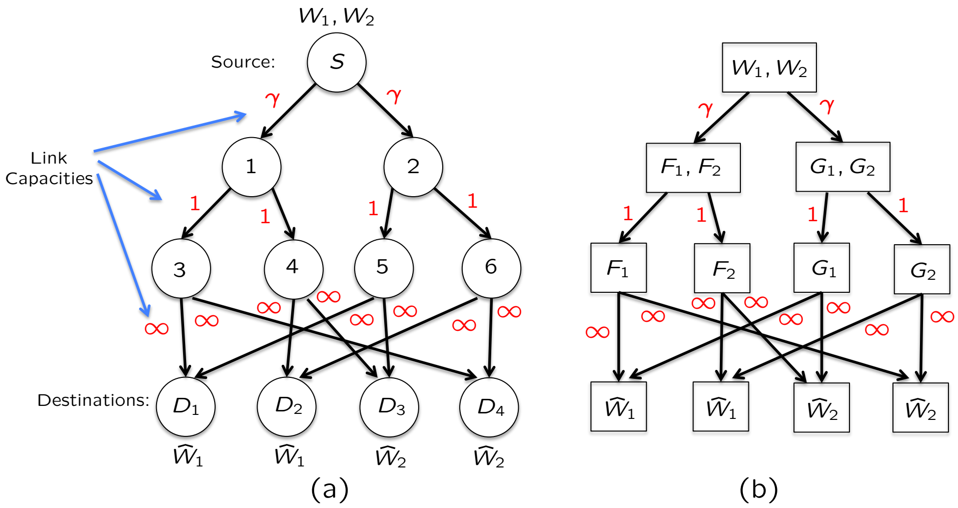

Figure 1: A simple network coding instance whose capacity characterization requires non-Shannon inequalities. $\gamma = 5/3$. (b) is an equivalent representation of (a), where the random variables inside a node represents that it has or wants.

As shown in Figure 1, the constructed network instance contains 2 messages, 1 source, 4 destinations and 6 intermediate nodes (in two layers), which can be described by only 6 random variables

(2 for messages and 4 for transmitted signals). The network structure (e.g., the number of messages, the number of links, the number of nodes, network topology) is the simplest known whose rate characterization requires non-Shannon inequalities [5, 7, 6, 8].

This simple network relates to a number of problems studied in prior network coding literature. For example, for a broadcast setting with 2 messages, 1 source and 2 destinations such that one destination is interested in one of the two messages, respectively, the capacity of arbitrary network topology is characterized in [26], where cut set outer bounds are tight. This result generalizes to the setting with common messages as well. Our example differs in that we have more destinations (4), out of which 2 have the same demand (compound setting). Another related class of problem is the combination network, where we have 1 source on the top, 1 layer of intermediate nodes and a number of destinations on the bottom. The capacity of each link is assumed to be 1. Combination network is first studied in the multicast setting, where unbounded gain of network coding versus routing is analyzed. More general message settings are studied in [27], and only Shannon inequalities are used in the converse arguments. Our example differs in that we have one more layer of intermediate nodes and the edge capacities are no longer restricted to 1. Note that for all networks discussed above, the converse only lies on Shannon inequalities. Therefore, the differences between our example and prior studied instances, e.g., diverse edge capacities, one more layer of intermediate nodes, overlapping demands, are essential for the necessity of non-Shannon inequalities.

Information inequalities lay down all rules that are obeyed by Shannon information measurements (entropy, joint entropy, conditional entropy and (conditional) mutual information). Shannon inequalities are a class of information inequalities that satisfy sub-modularity. All other inequalities are termed non-Shannon information inequalities, which remain largely a mystery. Beyond the natural application in characterizing network capacities, information inequalities are useful in counting in combinatorics, database theory and secret sharing in cryptography. The necessity of non-Shannon inequalities is a strong indicator for a hard problem as our understanding of non-Shannon inequalities is still very limited. The most well known approach of finding non-Shannon inequalities is to consider Shannon inequalities in a larger space and then project back to the original space. We do not understand the fundamental limits of this approach. All known non-Shannon inequalities essentially follow from this copy-and-project approach. A representative fact about non-Shannon inequalities that shows the depth is that there are an infinite number of non-Shannon inequalities and the region is not polyhedral. The network coding instance we find and its connection to OPIR shows the theoretical depth and central important of the PIR problem in information theory.

2 Storage Overhead Constrained Private Information Retrieval (OPIR)

In this work, we only consider the setting with $K = 2$ messages and $N = 2$ databases. We first introduce the problem formulation in this simple setting.

2.1 Problem Statement

Consider $K = 2$ independent messages W_1, W_2 of size L bits each.

$$H(W_1, W_2) = H(W_1) + H(W_2), \tag{1}$$

$$H(W_1) = H(W_2) = L. \tag{2}$$

There are $N = 2$ databases. Let S_n denote the information about the message realizations that is stored at the n^{th} database.

$$H(S_n|W_1, W_2) = 0, \forall n \in \{1, 2\} \tag{3}$$

Define the (normalized) storage overhead α_n for the n^{th} database as follows,

$$\alpha_n \quad \triangleq \quad \frac{H(S_n)}{L}. \tag{4}$$

For replication based schemes, each database stores all messages, so $S_n = (W_1, W_2)$, $H(S_n) = 2L$, and $\alpha_n = 2$. Replication is no longer possible if the storage overhead constraint requires $\alpha_n < 2$.

Let us use $\mathbb{F}$ to denote a random variable privately generated by the user, whose realization is not available to the databases. $\mathbb{F}$ represents the randomness in the strategies followed by the user. Similarly, $\mathbb{G}$ is a random variable that determines the random strategies followed by the databases, and whose realizations are assumed to be known to all the databases and the user without loss of generality. The random strategies are generated offline, i.e., before the realizations of the messages or the desired message index are known.

$$\begin{aligned} & H(\theta, \mathbb{F}, \mathbb{G}, W_1, W_2) \\ = \quad & H(\theta) + H(\mathbb{F}) + H(\mathbb{G}) + H(W_1) + H(W_2) \end{aligned} \tag{5}$$

A user privately generates θ uniformly from $\{1, 2\}$ and wishes to retrieve W_θ while keeping θ a secret from each database. Suppose $\theta = k$. In order to retrieve $W_k, k \in \{1, 2\}$ privately, the user privately generates $N = 2$ random queries, $Q_1^{[k]}, Q_2^{[k]}$.

$$H(Q_1^{[k]}, Q_2^{[k]}|\mathbb{F}) = 0, \quad \forall k \in \{1, 2\} \tag{6}$$

The user sends query $Q_n^{[k]}$ to the n^{th} database, $\forall n \in \{1, 2\}$. Upon receiving $Q_n^{[k]}$, the n^{th} database generates an answering string $A_n^{[k]}$. Without loss of generality, we assume that the answering string is a function of $Q_n^{[k]}$, the stored information S_n, and the random variable $\mathbb{G}$.

$$H(A_n^{[k]}|Q_n^{[k]}, S_n, \mathbb{G}) = 0. \tag{7}$$

Each database returns to the user its answer $A_n^{[k]}$.

From all the information that is now available to the user $(A_1^{[k]}, A_2^{[k]}, Q_1^{[k]}, Q_2^{[k]}, \mathbb{F})$, the user decodes the desired message W_k according to a decoding rule that is specified by the PIR scheme. Let P_e denote the probability of error achieved with the specified decoding rule.

To protect the user's privacy, the $K = 2$ strategies must be indistinguishable (identically distributed) from the perspective of any individual database, i.e., the following privacy constraint must be satisfied.

$$\text{[Privacy]} \quad (Q_n^{[1]}, A_n^{[1]}, \mathbb{G}, S_n) \quad \sim \quad (Q_n^{[2]}, A_n^{[2]}, \mathbb{G}, S_n), \forall n \in \{1, 2\} \tag{8}$$

where $X \sim Y$ denotes that random variables X and Y are identically distributed.

The PIR rate characterizes how many bits of desired information are retrieved per downloaded bit and is defined as follows.

$$R = \frac{L}{D} \tag{9}$$

where D is the expected value of the total number of bits downloaded by the user from all the databases.

A rate R is said to be ϵ-error achievable if there exists a sequence of PIR schemes, indexed by L, each of rate greater than or equal to R, for which $P_e \to 0$ as $L \to \infty$. Note that for such a sequence of PIR schemes, from Fano's inequality we must have

$$\begin{aligned} \text{[Correctness]} \quad o(L) &= \frac{1}{L} H(W_k | A_1^{[k]}, A_2^{[k]}, Q_1^{[k]}, Q_2^{[k]}, \mathbb{F}, \mathbb{G}) \\ &\overset{(6)}{=} \frac{1}{L} H(W_k | A_1^{[k]}, A_2^{[k]}, \mathbb{F}, \mathbb{G}), \quad \forall k \in \{1, 2\} \end{aligned} \tag{10}$$

where any function of L, say $f(L)$ is said to be $o(L)$ if $\lim_{L\to\infty} f(L)/L = 0$. The supremum of ϵ-error achievable rates is called the ϵ-error capacity C_ϵ.

A rate R is said to be zero-error achievable if there exists (for some L) a PIR scheme of rate greater than or equal to R for which $P_e = 0$. The supremum of zero-error achievable rates is called the zero-error capacity C_o. From the definitions, it is evident that

$$C_o \leq C_\epsilon \tag{11}$$

2.2 OPIR Capacity Characterization needs non-Shannon Inequalities

We consider the capacity of OPIR when we have $K = 2$ messages, $N = 2$ databases, and the storage overhead is $5/4$ per database (i.e., $\alpha_1 = \alpha_2 = 5/4$). We show that non-Shannon inequalities are necessary to settle this OPIR capacity problem. This result is stated in the following theorem.

Theorem 1 *For OPIR with $K = 2$ messages, $N = 2$ databases, and storage overhead $\alpha_1 = \alpha_2 = 5/4$, Shannon inequalities indicate that the rate of $2/3$ is achievable, while Zhang-Yeung non-Shannon inequality shows that the rate satisfies $R \leq 64/97 < 2/3$.*

The proof is presented next.

2.3 Proof of Theorem 1

We prove the result under ϵ-error framework and that under zero-error framework follows from (11). Let us make the following simplifying assumptions without loss of generality.

1. We assume that the PIR scheme is symmetric, in that

$$\begin{aligned} H(A_1^{[1]} | \mathbb{F}, \mathbb{G}) &= H(A_2^{[1]} | \mathbb{F}, \mathbb{G}) = H(A_2^{[2]} | \mathbb{F}, \mathbb{G}) & (12) \\ H(S_1) &= H(S_2) & (13) \end{aligned}$$

   Given any (asymmetric) PIR scheme that retrieves messages of size L, a symmetric PIR scheme with the same rate and storage overhead that retrieves messages of size NL is obtained by repeating the original scheme N times, and in the n^{th} repetition shifting the database indices cyclically by n. This symmetrization process is described in Theorem 3 of [17].

2. We assume that $Q_1^{[1]} = Q_1^{[2]}$, i.e., the query for the first database is chosen without the knowledge of the desired message index. There is no loss of generality in this assumption because of the privacy constraint, which requires that $Q_1^{[\theta]}$ is independent of θ. Note that this also means that $A_1^{[1]} = A_1^{[2]}$. Therefore, the PIR problem is described by 7 random variables $\mathbb{F}, \mathbb{G}, W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]}$. Note that the queries are functions of $\mathbb{F}$, from (6).

We then want to show that

1. $R = 2/3$ does not violate the constraints by Shannon inequalities, i.e., Shannon inequalities indicate that $R = 2/3$ is achievable.
2. However, Zhang-Yeung non-Shannon type inequality shows that $R \leq 64/97 < 2/3$. Therefore $R = 2/3$ is not achievable.

Next we prove the above two statements.

Proof of Statement 1

Statement 1 is proved by assuming that $R = 2/3$ and giving an assignment on the joint entropies of all subsets of random variables $(\mathbb{F}, \mathbb{G}, W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]})$ that satisfies all sub-modularity constraints. As a result, Shannon inequalities only can not rule out the possibility that $R = 2/3$.

The assignment is as follows. We assume $\mathbb{G}$ is deterministic so that we only need to consider the 6 remaining random variables. We denote a, a_1, a_2 each as an arbitrary element in $\{A_1^{[1]}, A_2^{[1]}, A_2^{[2]}\}$ (and a_1, a_2 are distinct), b, b_1, b_2 each as an arbitrary element in $\{W_1, W_2\}$ (and b_1, b_2 are distinct).

Terms in 1 variable: $H(\mathbb{F}) = 1, H(a) = 3L/4, H(b) = L$

Terms in 2 variables: $H(\mathbb{F}, A) = H(A) + 1, \text{where } |A| = 1, A \subset \{W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]}\}$
$H(b_1, b_2) = 2L, H(a, b) = 3L/2$
$H(a_1, a_2) = 5L/4$ if $(a_1, a_2) = (A_2^{[1]}, A_2^{[2]})$, and $3L/2$ otherwise

Terms in 3 variables: $H(\mathbb{F}, A) = H(A) + 1, \text{where } |A| = 2, A \subset \{W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]}\}$
$H(a_1, a_2, b) = 3L/2$, if from a_1, a_2, we can decode b
and all other terms equal $2L$

Terms in 4 variables: $H(\mathbb{F}, A) = H(A) + 1, \text{where } |A| = 3, A \subset \{W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]}\}$
all other terms equal $2L$.

Terms in 5, 6 variables: $H(W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]}) = 2L$
$H(\mathbb{F}, A) = H(A) + 1, \text{where } |A| = 4, 5, A \subset \{W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]}\}$.

We are ready to prove that Shannon inequalities (sub-modularity constraints) are all satisfied. Shannon inequalities require that for arbitrary two sets of random variables A, B,

$$H(A) + H(B) \geq H(A \cup B) + H(A \cap B) \tag{14}$$

We need to verify that (14) holds for all choices of $A, B \subset \{\mathbb{F}, W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]}\}$.

Without loss of generality, we assume that the carnality of A is no greater than the cardinality of B, i.e., $|A| \leq |B|$. It is easily seen that (14) holds when $A \cap B = \emptyset$ or $A \subset B$. Henceforth, we consider only the cases where A intersects with, but does not belong to B. We then must have $|A| \geq 2$. Note also that when $H(B) = 2L$ or $H(A) = 2L$, we have $H(A \cup B) = 2L$ and (14) follows trivially. Henceforth, we consider only the cases where $H(A) \neq 2L$ and $H(B) \neq 2L$. All possible cases are listed in the following.

We first consider the cases where $\mathbb{F} \notin A$ and $\mathbb{F} \notin B$.

- $|A| = 2, |B| = 2, 3$: We must have $|A \cap B| = 1$ and $H(A \cap B) \leq L$.

  1. $H(A) = H(B) = 3L/2$: $H(A) + H(B) = 3L \geq H(A \cup B) + H(A \cap B)$.
  2. $H(A) = 5L/4, H(B) = 3L/2$: In this case, $H(A \cap B) = H(a) = 3L/4$, so that $H(A) + H(B) = 11L/4 = 2L + 3L/4 \geq H(A \cup B) + H(A \cap B)$.
  3. $H(A) = 3L/2, H(B) = 5L/4$: Same as 2.
  4. $H(A) = H(B) = 5L/4$: $A \cap B = \emptyset$ and (14) follows.

- $|A| = |B| = 3$: We must have $H(A) = H(B) = 3L/2$. In this case, $H(A \cup B) = 2L$ and $H(A \cap B) = H(a) = 3L/4$, so that (14) holds.

Next, we consider the cases where $\mathbb{F} \in A$ and $\mathbb{F} \notin B$. Then we wish to prove that

$$H(\mathbb{F}, A/\mathbb{F}) + H(B) \geq H(\mathbb{F}, A/\mathbb{F} \cup B) + H(A/\mathbb{F} \cap B) \tag{15}$$

where A/c represent the set that consists of elements in A, with $c \in A$ excluded. Note that $H(\mathbb{F}, C) = 1 + H(C)$ for arbitrary $C \subset \{W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]}\}$, so it is equivalent to prove that

$$H(A/\mathbb{F}) + H(B) \geq H(A/\mathbb{F} \cup B) + H(A/\mathbb{F} \cap B) \tag{16}$$

and we have the case where $\mathbb{F}$ does not belong to the two sets, so that the proof follows from that presented above. The cases where $\mathbb{F} \notin A$ and $\mathbb{F} \in B$ follow from symmetry.

Finally, we consider the cases where $\mathbb{F} \in A$ and $\mathbb{F} \in B$. Similarly, noting that $H(\mathbb{F}, C) = 1 + H(C)$ for arbitrary $C \subset \{W_1, W_2, A_1^{[1]}, A_2^{[1]}, A_2^{[2]}\}$, we wish to prove that

$$\begin{aligned} H(\mathbb{F}, A/\mathbb{F}) + H(\mathbb{F}, B/\mathbb{F}) &\geq H(\mathbb{F}, A/\mathbb{F} \cup B/\mathbb{F}) + H(A/\mathbb{F} \cap B/\mathbb{F}) & (17) \\ \iff \quad H(A/\mathbb{F}) + H(B/\mathbb{F}) &\geq H(A/\mathbb{F} \cup B/\mathbb{F}) + H(A/\mathbb{F} \cap B/\mathbb{F}) & (18) \end{aligned}$$

where again we boil down to the cases that have been considered. The proof is complete.

Proof of Statement 2

Let us start with a useful lemma.

Lemma 1

$$\begin{aligned} H(A_1^{[1]}|W_1, \mathbb{F}, \mathbb{G}), H(A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G}), H(A_2^{[2]}|W_2, \mathbb{F}, \mathbb{G}) &\leq L(1/R - 1) + o(L) & (19) \\ H(A_1^{[1]}|W_1, \mathbb{F}, \mathbb{G}), H(A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G}), H(A_2^{[2]}|W_2, \mathbb{F}, \mathbb{G}) &\geq L(2 - 1/R) + o(L) & (20) \\ H(A_2^{[1]}|W_1, A_2^{[2]}, \mathbb{F}, \mathbb{G}), H(A_2^{[2]}|W_2, A_2^{[1]}, \mathbb{F}, \mathbb{G}) &\geq L(5 - 3/R) + o(L) & (21) \\ I(A_2^{[1]}; A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G}), I(A_2^{[1]}; A_2^{[2]}|W_2, \mathbb{F}, \mathbb{G}) &\leq L(4/R - 6) + o(L) & (22) \end{aligned}$$

Proof: We first prove (19). From (47) in Lemma 2 of [17] we have

$$\begin{aligned}
L(1/R-1)+o(L) &\geq I(W_2; A_1^{[1]}, A_2^{[1]}, W_1, \mathbb{F}, \mathbb{G}) && (23)\\
&\overset{(5)}{=} I(W_2; A_1^{[1]}, A_2^{[1]} | W_1, \mathbb{F}, \mathbb{G}) && (24)\\
&\overset{(6)(7)(3)}{=} H(A_1^{[1]}, A_2^{[1]} | W_1, \mathbb{F}, \mathbb{G}) && (25)\\
\Rightarrow L(1/R-1)+o(L) &\geq H(A_1^{[1]} | W_1, \mathbb{F}, \mathbb{G}) && (26)\\
\text{and } L(1/R-1)+o(L) &\geq H(A_2^{[1]} | W_1, \mathbb{F}, \mathbb{G}) && (27)\\
&= H(A_2^{[2]} | W_1, \mathbb{F}, \mathbb{G}) && (28)
\end{aligned}$$

where (28) follows from the derivations in (76) to (81) in [17]. Symmetrically, from (26), it follows that $H(A_2^{[2]}|W_2, \mathbb{F}, \mathbb{G}) \leq L(1/R-1)+o(L)$.

We next prove (20). From (73) in [17], we have

$$L \leq H(A_1^{[1]}|W_1, \mathbb{F}, \mathbb{G}) + H(A_2^{[1]}|W_1, \mathbb{F}, \mathbb{G}) \qquad (29)$$

Combining (26), (27) and (29), we have shown that

$$\begin{aligned}
L(2-1/R)+o(L) &\leq H(A_1^{[1]}|W_1, \mathbb{F}, \mathbb{G}) && (30)\\
L(2-1/R)+o(L) &\leq H(A_2^{[1]}|W_1, \mathbb{F}, \mathbb{G}) \overset{(28)}{=} H(A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G}) && (31)
\end{aligned}$$

Symmetrically, from (30), it follows that $H(A_2^{[2]}|W_2, \mathbb{F}, \mathbb{G}) \geq L(2-1/R)+o(L)$.

We proceed to prove (21). We only prove $H(A_2^{[1]}|W_1, A_2^{[2]}, \mathbb{F}, \mathbb{G}) \geq L(5-3/R)+o(L)$ and the other inequality follows from symmetry. From Shannon inequalities (sub-modularity), we have

$$\begin{aligned}
& H(A_2^{[1]}, A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G})\\
\geq\ & -H(A_1^{[1]}, A_2^{[1]}|W_1, \mathbb{F}, \mathbb{G}) + H(A_1^{[1]}, A_2^{[1]}, A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G}) + H(A_2^{[1]}|W_1, \mathbb{F}, \mathbb{G}) && (32)\\
\overset{(25)(10)(31)}{\geq}\ & -L(1/R-1) + H(A_1^{[2]}, A_2^{[1]}, A_2^{[2]}, W_2|W_1, \mathbb{F}, \mathbb{G}) + L(2-1/R) + o(L) && (33)\\
\geq\ & H(W_2|W_1, \mathbb{F}, \mathbb{G}) + L(3-2/R) + o(L) && (34)\\
\overset{(5)(2)}{=}\ & L(4-2/R) + o(L) && (35)\\
\Rightarrow\ & H(A_2^{[1]}|W_1, A_2^{[2]}, \mathbb{F}, \mathbb{G}) = H(A_2^{[1]}, A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G}) - H(A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G}) && (36)\\
\overset{(28)}{\geq}\ & L(4-2/R) - L(1/R-1) + o(L) = L(5-3/R) + o(L) && (37)
\end{aligned}$$

(22) are direct consequences of (19) to (21). ■

Using Zhang-Yeung non-Shannon Inequality

Equipped with Lemma 1, we are ready to call Zhang-Yeung non-Shannon inequality [28] to produce the desired bound on R.

$$\begin{aligned}
& I(A_2^{[1]}; A_2^{[2]}|\mathbb{F}, \mathbb{G})\\
\leq\ & \frac{1}{2}I(W_1; W_2|\mathbb{F}, \mathbb{G}) + \frac{1}{2}I(W_1; A_2^{[1]}, A_2^{[2]}|\mathbb{F}, \mathbb{G}) + \frac{3}{2}I(A_2^{[1]}; A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G}) + \frac{1}{2}I(A_2^{[1]}; A_2^{[2]}|W_2, \mathbb{F}, \mathbb{G})
\end{aligned}$$

$$\begin{aligned}
&\overset{(22)(5)}{\leq} && \frac{1}{2}I(W_1; A_2^{[1]}, A_2^{[2]}|\mathbb{F}, \mathbb{G}) + 2L(4/R-6) + o(L) && (38)\\
\Rightarrow & && H(A_2^{[1]}|\mathbb{F}, \mathbb{G}) + H(A_2^{[2]}|\mathbb{F}, \mathbb{G}) - H(A_2^{[1]}, A_2^{[2]}|\mathbb{F}, \mathbb{G}) && (39)\\
&\leq && \frac{1}{2}H(A_2^{[1]}, A_2^{[2]}|\mathbb{F}, \mathbb{G}) - \frac{1}{2}H(A_2^{[1]}, A_2^{[2]}|W_1, \mathbb{F}, \mathbb{G}) + 2L(4/R-6) + o(L) && (40)\\
&\overset{(35)}{\leq} && \frac{1}{2}H(A_2^{[1]}, A_2^{[2]}|\mathbb{F}, \mathbb{G}) - L(2-1/R) + 2L(4/R-6) + o(L) && (41)\\
&= && \frac{1}{2}H(A_2^{[1]}, A_2^{[2]}|\mathbb{F}, \mathbb{G}) + L(9/R-14) + o(L) && (42)\\
\Rightarrow & && H(A_2^{[1]}, A_2^{[2]}|\mathbb{F}, \mathbb{G}) && (43)\\
&\geq && \frac{2}{3}\Big[L(14-9/R) + H(A_2^{[1]}|\mathbb{F}, \mathbb{G}) + H(A_2^{[2]}|\mathbb{F}, \mathbb{G})\Big] + o(L) && (44)\\
&\overset{(9)}{=} && L(28-16/R)/3 + o(L) && (45)\\
\overset{(4)}{\Rightarrow} & && 5L/4 \geq L(28-16/R)/3 + o(L) && (46)
\end{aligned}$$

Let L go to infinity, then we have $R \leq 64/97 < 2/3$.

3 Network Coding

Inspired by the OPIR capacity result presented as above, we find the simplest (to our knowledge) example of a network coding capacity problem (shown in Figure 2) where non-Shannon information inequalities are necessary. The network coding instance is constructed by translating the simplest decoding structure for the OPIR problem. Let us start with the problem formulation, specified for the network instance in Figure 2.

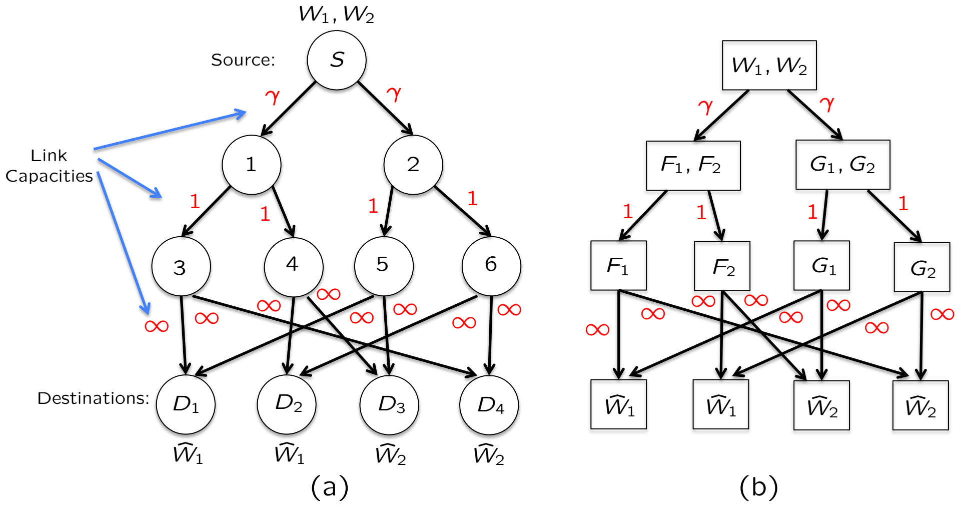

Figure 2: (a) A network coding instance. $\gamma = 5/3$. (b) An equivalent representation. A square represents a node. The random variables inside a node represent those it has (wants).

3.1 Problem Statement

The problem statement and definitions are standard Shannon theoretic [29], which we briefly summarize below. There are two independent messages W_1, W_2, each of which is uniform over the index set $\{1, 2, \cdots, 2^{nR}\}$. The source node S wishes to send W_1 to destination nodes D_1, D_2 and W_2 to D_3, D_4, respectively, through n channel uses. The network topology is depicted in Figure 2(a). Over each channel use, a random variable is sent over each link, and the entropy of the random variable can not exceed the capacity of that link. $\gamma = 5/3$. The random variable sent on the outgoing link of a node is a deterministic function of the random variables that are received through the incoming links over current and previous channel uses. From all the random variables received over all n channel uses, each destination can recover the desired message either with vanishing probability of error or zero-error. The supreme of the achievable rate R is the capacity $C_o(C_\epsilon)$ under zero error (vanishing probability of error).

3.2 Necessity of Non-Shannon Inequalities

We show that to characterize the capacity of the network coding instance in Figure 2, non-Shannon inequalities are necessary. This result is stated in the following theorem.

Theorem 2 *For the Network Coding instance shown in Figure 2, Shannon inequalities indicate that $R = 4/3$ is achievable, while Zhang-Yeung non-Shannon inequality shows that $R \leq 37/28 < 4/3$.*

The proof is presented next.

3.3 Proof of Theorem 2

We prove the result under ϵ-error framework and that under zero-error framework follows. It is easy to see that the network coding instance can be equivalently represented as shown in Figure 2(b), where the random variables that a node has (wants) are shown inside the node. F_1^n represents the stack of F_1 over all n channel uses and other notations are defined similarly. The following entropy constraints are satisfied without loss of generality.

$$\begin{aligned}
H(W_1, W_2) &= H(W_1) + H(W_2) && (47)\\
H(W_1) = H(W_2) &= nR && (48)\\
H(F_1^n, F_2^n, G_1^n, G_2^n | W_1, W_2) &= 0 && (49)\\
H(W_1 | F_1^n, G_1^n) &= o(n) && (50)\\
H(W_1 | F_2^n, G_2^n) &= o(n) && (51)\\
H(W_2 | F_1^n, G_2^n) &= o(n) && (52)\\
H(W_2 | F_2^n, G_1^n) &= o(n) && (53)\\
H(F_1^n, F_2^n) = H(G_1^n, G_2^n) &= 5n/3 && (54)\\
H(F_1^n) = H(G_1^n) = H(F_2^n) = H(G_2^n) &= n && (55)
\end{aligned}$$

We want to show that

1. $R = 4/3$ does not violate the constraints by Shannon inequalities, i.e., Shannon inequalities indicate that $R = 4/3$ is achievable.

2. However, Zhang-Yeung non-Shannon type inequality shows that $R \leq 21/16 < 4/3$. Therefore $R = 4/3$ is not achievable.

Next we prove the above two statements.

Proof of Statement 1

Statement 1 is proved by assuming that $R = 4/3$ and giving an assignment on the joint entropies of all subsets of random variables $(W_1, W_2, F_1^n, F_2^n, G_1^n, G_2^n)$ that satisfy all sub-modularity constraints. Therefore Shannon inequalities only can not rule out the possibility that $R = 4/3$.

The assignment is as follows. We denote a, a_1, a_2 each as an arbitrary element in $\{F_1^n, G_1^n, F_2^n, G_2^n\}$ (and a_1, a_2 are distinct), b, b_1, b_2 each as an arbitrary element in $\{W_1, W_2\}$ (and b_1, b_2 are distinct).

Terms in 1 variable: $H(a) = n, H(b) = 4n/3$

Terms in 2 variables: $H(b_1, b_2) = 8n/3, H(a, b) = 2n$

$H(a_1, a_2) = 5n/3$ if $(a_1, a_2) = (f_1, f_2)$ or (g_1, g_2), and $2n$ otherwise

Terms in 3 variables: $H(a_1, a_2, b) = 2n$, if from a_1, a_2, we can decode b

and all other terms equal $8n/3$

Terms in 4 or more variables: all terms equal $8n/3$.

We are ready to prove that Shannon inequalities (sub-modularity constraints) are all satisfied. Shannon inequalities require that for arbitrary two sets of random variables A, B,

$$H(A) + H(B) \geq H(A \cup B) + H(A \cap B) \tag{56}$$

We need to verify that (56) holds for all choices of $A, B \subset \{W_1, W_2, f_1, f_2, g_1, g_2\}$. Without loss of generality, we assume that the carnality of A is smaller than the cardinality of B, i.e., $|A| \leq |B|$. It is easily seen that (56) holds when $A \cap B = \emptyset$ or $A \subset B$. Henceforth, we consider only the cases where A intersects with, but does not belong to B. We then must have $|A| \geq 2$. Note also that when $H(B) = 8n/3$ or $H(A) = 8n/3$, we have $H(A \cup B) = 8n/3$ and (56) follows trivially. Henceforth, we consider only the cases where $H(A) \neq 8n/3$ and $H(B) \neq 8n/3$. All possible cases are listed in the following.

- $|A| = 2, |B| = 2, 3$: We must have $|A \cap B| = 1$ and $H(A \cap B) \leq 4n/3$.

  1. $H(A) = H(B) = 2n$: $H(A) + H(B) = 4n \geq H(A \cup B) + H(A \cap B)$.
  2. $H(A) = 5n/3, H(B) = 2n$: In this case, $H(A \cap B) = H(a) = n$, so that $H(A) + H(B) = 11n/3 = 8n/3 + n \geq H(A \cup B) + H(A \cap B)$.
  3. $H(A) = 2n, H(B) = 5n/3$: Same as 2.
  4. $H(A) = H(B) = 5n/3$: $A \cap B = \emptyset$ and (56) follows.

- $|A| = |B| = 3$: We must have $H(A) = H(B) = 2n$. In this case, $H(A \cup B) = 8n/3$ and $H(A \cap B) = H(b) = 4n/3$ or $H(A \cap B) = H(a) = n$, so that (56) holds.

Proof of Statement 2

Let us start with a useful lemma.

Lemma 2

$$
\begin{aligned}
H(G_1^n|W_1), H(G_2^n|W_1), H(F_1^n|W_1), H(F_2^n|W_1) &\leq n(2-R)+o(n) \qquad (57)\\
H(G_1^n|W_1), H(G_2^n|W_1), H(F_1^n|W_1), H(F_2^n|W_1) &\geq n(2R-2)+o(n) \qquad (58)\\
H(G_1^n|W_2), H(G_2^n|W_2), H(F_1^n|W_2), H(F_2^n|W_2) &\leq n(2-R)+o(n) \qquad (59)\\
H(G_1^n|W_2), H(G_2^n|W_2), H(F_1^n|W_2), H(F_2^n|W_2) &\leq n(2R-2)+o(n) \qquad (60)\\
H(G_1^n|W_1, G_2^n), H(G_2^n|W_2, G_1^n) &\geq n(5R-6)+o(n) \qquad (61)\\
I(G_1^n; G_2^n|W_1), I(G_1^n; G_2^n|W_2) &\leq n(8-6R)+o(n) \qquad (62)
\end{aligned}
$$

Proof: We first prove (57).

$$
\begin{aligned}
2n &\overset{(55)}{=} H(F_1^n)+H(G_1^n) \qquad (63)\\
&\geq H(F_1^n, G_1^n) \qquad (64)\\
&\overset{(50)}{=} H(F_1^n, G_1^n, W_1)+o(n) \qquad (65)\\
&= H(W_1)+H(F_1^n, G_1^n|W_1)+o(n) \qquad (66)\\
&\overset{(48)}{=} nR+H(F_1^n, G_1^n|W_1)+o(n) \qquad (67)\\
\Rightarrow H(F_1^n|W_1) &\leq n(2-R)+o(n),\ H(G_1^n|W_1) \leq n(2-R)+o(n) \qquad (68)\\
\Rightarrow (\text{Symmetry}):\ H(F_2^n|W_1) &\leq n(2-R)+o(n),\ H(G_2^n|W_1) \leq n(2-R)+o(n) \qquad (69)
\end{aligned}
$$

We next prove (58).

$$
\begin{aligned}
H(F_1^n|W_1)+H(G_2^n|W_1) &\geq H(F_1^n, G_2^n|W_1) \qquad (70)\\
&\overset{(52)}{=} H(F_1^n, G_2^n, W_2|W_1)+o(n) \qquad (71)\\
&\geq H(W_2|W_1)+o(n) \qquad (72)\\
&\overset{(47)(48)}{=} nR+o(n) \qquad (73)\\
\Rightarrow (\text{Symmetry}):\ H(F_2^n|W_1)+H(G_1^n|W_1) &\geq nR+o(n) \qquad (74)
\end{aligned}
$$

Combining (68), (69) and (73), (74), we have the desired equality (58). From (57)(58), we have (59)(60), by symmetry.

We proceed to prove (61). We only prove $H(G_1^n|W_1, G_2^n) \geq n(5R-6)+o(n)$ and the other inequality follows from symmetry. From Shannon inequalities (sub-modularity), we have

$$
\begin{aligned}
H(G_1^n, G_2^n|W_1) &\geq -H(G_1^n, F_1^n|W_1)+H(F_1^n, G_1^n, G_2^n|W_1)+H(G_1^n|W_1) \qquad (75)\\
&\overset{(67)(52)(58)}{\geq} -n(2-R)+H(F_1^n, G_1^n, G_2^n, W_2|W_1)+n(2R-2)+o(n) \qquad (76)\\
&\geq H(W_2|W_1)+n(3R-4)+o(n) \qquad (77)\\
&\overset{(47)(48)}{=} 4n(R-1)+o(n) \qquad (78)\\
\Rightarrow H(G_1^n|W_1, G_2^n) &= H(G_1^n, G_2^n|W_1)-H(G_2^n|W_1) \qquad (79)\\
&\overset{(57)}{\geq} 4n(R-1)-n(2-R)+o(n) = n(5R-6)+o(n) \qquad (80)
\end{aligned}
$$

(62) are direct consequences of (57) to (61). ■

Using Zhang-Yeung non-Shannon Inequality

Equipped with Lemma 1, we are ready to call Zhang-Yeung non-Shannon inequality [28] to produce the desired bound on R.

$$I(G_1^n; G_2^n) \leq \frac{1}{2}I(W_1; W_2) + \frac{1}{2}I(W_1; G_1^n, G_2^n) + \frac{3}{2}I(G_1^n; G_2^n|W_1) + \frac{1}{2}I(G_1^n; G_2^n|W_2)$$

$$\overset{(62)(47)}{\Longrightarrow} I(G_1^n; G_2^n) \leq \frac{1}{2}I(W_1; G_1^n, G_2^n) + 2n(8-6R) + o(n) \tag{81}$$

$$\Rightarrow H(G_1^n) + H(G_2^n) - H(G_1^n, G_2^n) \leq \frac{1}{2}H(G_1^n, G_2^n) - \frac{1}{2}H(G_1^n, G_2^n|W_1) + 2n(8-6R) + o(n) \tag{82}$$

$$\overset{(78)}{\leq} \frac{1}{2}H(G_1^n, G_2^n) - 2n(R-1) + 2n(8-6R) + o(n) \tag{83}$$

$$= \frac{1}{2}H(G_1^n, G_2^n) + 2n(9-7R) + o(n) \tag{84}$$

$$\Rightarrow H(G_1^n, G_2^n) \geq \frac{2}{3}\left[2n(7R-9) + H(G_1^n) + H(G_2^n)\right] + o(n) \tag{85}$$

$$\overset{(55)}{=} 2n(14R-16)/3 + o(n) \tag{86}$$

$$\overset{(54)}{\Rightarrow} 5n/3 \geq 2n(14R-16)/3 + o(n) \tag{87}$$

Let n go to infinity, then we have $R \leq 37/28 < 4/3$.

4 Conclusion

We show that the capacity characterization of OPIR requires non-Shannon inequalities in general. We also construct a closely related network coding instance that requires non-Shannon inequalities. These results, along with the connection between PIR and network coding, indicate the central importance of PIR in information theory.

The dual of the OPIR capacity problem is the minimum storage overhead problem of PIR for a given rate. In this context, we may similarly show the necessity of non-Shannon inequalities. In particular, consider the PIR problems with $K = 2$ messages, $N = 2$ databases, and rate $R = 2/3$ (capacity achieving), the best outer bound by all Shannon inequalities on the minimum storage overhead is that $\alpha_1(\alpha_2) \geq 5/4$ and Zhang-Yeung non-Shannon inequality produces a tighter bound, $\alpha_1(\alpha_2) \geq 4/3$. Further, [17] shows that the minimum storage overhead of all linear PIR schemes is that $\alpha_1(\alpha_2) = 3/2$. As a consequence, the minimum storage overhead problem is still open and we either need non-linear schemes, or to prove the optimality of linear schemes, we need stronger non-Shanonn inequalities.

References

[1] R. Freij-Hollanti, O. Gnilke, C. Hollanti, and D. Karpuk, "Private Information Retrieval from Coded Databases with Colluding Servers," *arXiv preprint arXiv:1611.02062*, 2016.

[2] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2745 – 2759, Aug. 2005.

[3] S. Rouayheb, A. Sprintson, and C. Georghiades, "On the Index Coding Problem and Its Relation to Network Coding and Matroid Theory," *IEEE Trans. on Inf. Theory*, vol. 56, no. 7, pp. 3187–3195, July 2010.

[4] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Lexicographic products and the power of non-linear network coding," *ArXiv:1108.2489*, Aug. 2011. [Online]. Available: http://arxiv.org/abs/1108.2489

[5] R. Dougherty, C. Freiling, and K. Zeger, "Networks, matroids, and non-Shannon information inequalities," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 1949 – 1969, Jun. 2007.

[6] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Broadcasting with side information: Bounding and approximating the broadcast rate," *IEEE Trans. on Inf. Theory*, vol. 59, no. 9, pp. 5811–5823, 2013.

[7] R. Baber, D. Christofides, A. N. Dang, S. Riis, and E. R. Vaughan, "Multiple unicasts, graph guessing games, and non-shannon inequalities," in *International Symposium on Network Coding (NetCod)*, 2013, pp. 1–6.

[8] H. Sun and S. A. Jafar, "Index Coding Capacity: How far can one go with only Shannon Inequalities?" *IEEE Trans. on Inf. Theory*, vol. 61, no. 6, pp. 3041–3055, 2015.

[9] F. Matus, "Infinitely many information inequalities," in *Proceedings of International Symposium on Information Theory (ISIT)*, 2007, pp. 41 – 44.

[10] T. H. Chan, "Recent Progresses in Characterising Information Inequalities," *Entropy*, vol. 13, pp. 379 – 401, 2011.

[11] H. Sun and S. A. Jafar, "The Capacity of Private Information Retrieval," *arXiv preprint arXiv:1602.09134*, 2016.

[12] ——, "Optimal Download Cost of Private Information Retrieval for Arbitrary Message Length," *arXiv preprint arXiv:1610.03048*, 2016.

[13] ——, "The Capacity of Robust Private Information Retrieval with Colluding Databases," *arXiv preprint arXiv:1605.00635*, 2016.

[14] ——, "The Capacity of Symmetric Private Information Retrieval," *arXiv preprint arXiv:1606.08828*, 2016.

[15] K. Banawan and S. Ulukus, "The Capacity of Private Information Retrieval from Coded Databases," *arXiv preprint arXiv:1609.08138*, 2016.

[16] Q. Wang and M. Skoglund, "Symmetric Private Information Retrieval For MDS Coded Distributed Storage," *arXiv preprint arXiv:1610.04530*, 2016.

[17] H. Sun and S. A. Jafar, "Multiround Private Information Retrieval: Capacity and Storage Overhead," *arXiv preprint arXiv:1611.02257*, 2016.

[18] N. Shah, K. Rashmi, and K. Ramchandran, "One Extra Bit of Download Ensures Perfectly Private Information Retrieval," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2014, pp. 856–860.

[19] T. H. Chan, S.-W. Ho, and H. Yamamoto, "Private Information Retrieval for Coded Storage," *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, pp. 2842–2846, 2015.

[20] A. Fazeli, A. Vardy, and E. Yaakobi, "Codes for distributed PIR with low storage overhead," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 2852–2856.

[21] R. Tajeddine and S. E. Rouayheb, "Private Information Retrieval from MDS Coded Data in Distributed Storage Systems," *arXiv preprint arXiv:1602.01458*, 2016.

[22] S. Rao and A. Vardy, "Lower Bound on the Redundancy of PIR Codes," *arXiv preprint arXiv:1605.01869*, 2016.

[23] S. Blackburn and T. Etzion, "PIR Array Codes with Optimal PIR Rate," *arXiv preprint arXiv:1607.00235*, 2016.

[24] T. E. Simon R. Blackburn and M. B. Paterson, "PIR schemes with small download complexity and low storage requirements," *arXiv preprint arXiv:1609.07027*, 2016.

[25] Y. Zhang, X. Wang, H. Wei, and G. Ge, "On private information retrieval array codes," *arXiv preprint arXiv:1609.09167*, 2016.

[26] C. Ngai and R. Yeung, "Multisource network coding with two sinks," in *Communications, Circuits and Systems, 2004. ICCCAS 2004. 2004 International Conference on*, vol. 1. IEEE, 2004, pp. 34–37.

[27] S. S. Bidokhti, V. M. Prabhakaran, and S. N. Diggavi, "Capacity results for multicasting nested message sets over combination networks," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 4968–4992, 2016.

[28] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1440 – 1452, Jul. 1998.

[29] R. W. Yeung, *Information Theory and Network Coding.* Springer, 2008.