On Extremal Rates of Storage Over Graphs

Zhou Li, *Member, IEEE*, and Hua Sun, *Member, IEEE*

***Abstract*— A storage code over a graph maps K independent source symbols, each of L_w bits, to N coded symbols, each of L_v bits, such that each coded symbol is stored in a node of the graph and each edge of the graph is associated with one source symbol. From a pair of nodes connected by an edge, the source symbol that is associated with the edge can be decoded. The ratio L_w/L_v is called the symbol rate of a storage code and the highest symbol rate is called the capacity. We show that the three highest capacity values of storage codes over graphs are $2, 3/2, 4/3$. We characterize all graphs over which the storage code capacity is 2 and $3/2$, and for capacity value of $4/3$, necessary condition and sufficient condition (that do not match) on the graphs are given.**

***Index Terms*— Capacity, extremal rates, storage codes.**

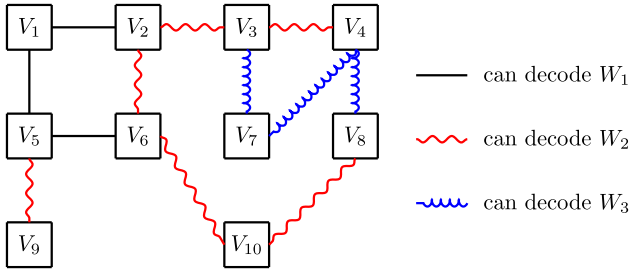

Fig. 1. An example graph of a storage code problem with $K = 3$ source symbols and $N = 10$ coded symbols, whose capacity turns out to be $4/3$ (refer to Theorem 7. See Fig. 11 for a code construction).

I. Introduction

MOTIVATED by the heterogeneity of modern distributed storage systems, a storage code problem over graphs is introduced in [1] and [2], where a storage code maps K independent source symbols, $W_1, \cdots, W_K$ to N coded symbols, $V_1, \cdots, V_N$, and the coded symbols are stored in the node set of a graph $\{V_1, \cdots, V_N\}$ (so that V_n denotes both the coded symbol and the node). The heterogeneous data recovery pattern is captured by the edges of the graph, where each edge $\{V_i, V_j\}$ is associated with one source symbol W_k and from (V_i, V_j), we can decode W_k. As the structure of the graph can be very diverse, versatile distributed storage and data access requirements can be accommodated. An example of the storage code problem over a graph is given in Fig. 1. The metric of pursuit is the capacity C of a storage code over a graph, i.e., the highest possible symbol rate, defined as L_w/L_v, where $L_w(L_v)$ is the number of bits contained in each source (coded) symbol and L_w/L_v represents the number of source symbol bits reliably stored in each coded symbol bit.

The graph based storage code problem is not new in the sense that it can be equivalently transformed to a network coding problem [1], [2], [3], [4] and adding further security constraints (i.e., beyond desired data decodability, leakage about other source symbols is prevented), it is intimately related to conditional disclosure of secrets [5], [6], [7], [8] and secret sharing [9], [10]. What is new is the view brought by [1] - finding extremal networks/graphs. Instead of first fixing the network/graph and then finding its highest rate, we focus on the extremal (highest) capacity values and aim to find the networks/graphs whose capacity is equal to the extremal values (see Fig. 2). This complementary view is useful in identifying critical combinatorial graph structures that limit the rate and in separating more tractable graph classes in terms of capacity characterization. Considering that networks are becoming more and more heterogeneous and solving each network instance becomes infeasible and impossible (as hard instances that require non-linear codes for achievability or non-Shannon information inequalities for converse are well known [11], [12], [13]), this extremal rate and extremal network approach might be a fruitful direction to produce new results and insights.

In this work, we start from the highest possible capacity values and for the two highest rates - 2 and $3/2$, all extremal graphs with corresponding extremal capacity values are easily characterized. For extremal rate of 2, absolute no interference is allowed as $L_w = 2L_v$, i.e., a pair of nodes can just store the desired source symbols. As long as there exists interference, the maximal capacity value drops to $3/2$, the next extremal rate, and all storage code instances with capacity $3/2$ only require intra-source symbol coding, i.e., mixing of symbols from the same source symbol. When rate of $3/2$ cannot be achieved, the next highest capacity value is shown to be $4/3$, which is our main focus and the corresponding graphs turn out to be highly technical. We identify necessary condition (converse required) and sufficient condition (achievability provided) for graphs with storage code capacity $4/3$ (see Fig. 2). The converse is based on delicate arguments on the intimate relation between the maximum amount of interference (undesired source symbols) allowed and the minimum amount of desired source symbols needed. The achievable scheme uses

Manuscript received 7 February 2023; revised 9 August 2023; accepted 18 October 2023. Date of publication 30 October 2023; date of current version 19 March 2024. This work was supported in part by NSF under Grant CCF-2007108, Grant CCF-2045656, and Grant CCF-2312228. An earlier version of this paper was presented in part at the 2023 IEEE International Symposium on Information Theory [DOI: 10.1109/ISIT54713.2023.10206619]. *(Corresponding author: Zhou Li.)*

The authors are with the Department of Electrical Engineering, University of North Texas, Denton, TX 76207 USA (e-mail: ZhouLi@my.unt.edu; hua.sun@unt.edu).

Communicated by I. Tamo, Associate Editor for Coding and Decoding.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3328434.

Digital Object Identifier 10.1109/TIT.2023.3328434

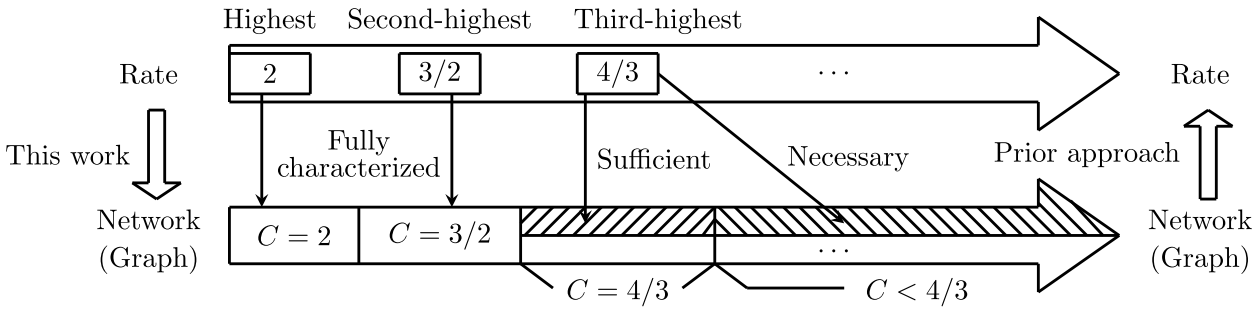

Fig. 2. The extremal rate and network approach of this work and results obtained.

vector linear codes that carefully control the alignment of interfering source symbols and the independence of desired source symbols. The conditions are stated in terms of the presence or absence of critical nodes and edges of the graph, whose combinatorial structure places constraints on the code rate.

A. Related Work and More Background

Before proceeding to the problem statement section, we give a more detailed account of related work using our terminology so that our paper is also put in perspective. In classical work on algebraic storage codes for distributed storage (e.g., RAID architectures [14], [15]), the studied regime is when the symbol rate is 1, i.e., $L_w = L_v$ (each source symbol has the same size as each coded symbol) and the recovery constraint is placed uniformly on *all* source symbols, e.g., minimum distance constraint (number of erasures that can be tolerated for no loss in recovering all sources). In contrast, we allow to *vary* the symbol rate (which is our main figure of merit) and our recovery constraint is stated with respect to *each* source symbol (instead of all) so that each source symbol may have different access patterns, in line with requirements of modern heterogeneous storage systems.

More recently, the distributed storage repair problem attracted much attention which mainly includes two lines of work - regenerating codes [16], [17] and locally repairable (recoverable) codes [18], [19], [20]. Both lines focus on how to efficiently recover lost (erased) *coded symbols* (server) while we focus on the recovery of *source symbols*. Regenerating codes use the communication cost (repair bandwidth or more generally the tradeoff between the storage cost and communication cost) as the performance metric while locally repairable codes use the number of coded symbols contacted during repair (called locality) as the performance metric. While early work considers the symmetric case where the number of coded symbols contacted is a constant, some recent work brings graphical topology into the picture [21], [22], [23], [24] where graphs are used to model the network connectivity (which links or coded symbols can be used for recovery). Note that these graphs, which describe coded symbol repair constraint, are different from ours, which describe source symbol recovery constraint although a similar term - storage codes on graphs is used.

Last but not least, locally decodable codes [25], [26], studied more in computer science literature, tackle how to efficiently recover source symbols by contacting a few coded symbols. Similar to classical coding theory work, the focus is mainly on minimizing the code length while fixing the symbol rate. Instead we focus on maximizing the symbol rate for fixed code length. The only locally decodable codes work that studies the symbol rate, to the best of our knowledge, is [27], where the recovery constraint is only on the number of coded symbols contacted and not specified by a graph.

II. Problem Statement and Definitions

Consider K independent uniform source symbols $W_1, \cdots, W_K$ of size L_w bits each.

$$H(W_1, \cdots, W_K) = H(W_1) + \cdots + H(W_K),$$
$$L_w = H(W_1) = \cdots = H(W_K). \tag{1}$$

Consider N coded symbols $V_1, \cdots, V_N$, each of L_v bits. Our interest lies in the relative size of L_w, L_v (see (3)) and coding over arbitrary finite fields is allowed, so L_w, L_v can take arbitrarily large values (that are not necessarily integers).

The source symbol recoverability constraint on the coded symbols is specified by a graph $G = (\mathcal{V}, \mathcal{E}, t)$, where the node[1] set $\mathcal{V} = \{V_1, \cdots, V_N\}$, the edge set $\mathcal{E}$ is a set of unordered pairs from $\mathcal{V}$, and the function t associates each edge $\{V_i, V_j\} \in \mathcal{E}$ with a source symbol $W_k, k \in \{1, 2, \cdots, K\} \triangleq [K]$, i.e., $t(\{V_i, V_j\}) = W_k$. For each edge $\{V_i, V_j\} \in \mathcal{E}$ such that $t(\{V_i, V_j\}) = W_k$, we can decode W_k with no error, i.e.,

$$H(W_k | V_i, V_j) = 0 \text{ if } t(\{V_i, V_j\}) = W_k. \tag{2}$$

Isolated nodes are trivial as they are not connected to any edges and thus involve no constraints. Without loss of generality, we assume in this work that any graph contains no isolated nodes.

A mapping from the source symbols $W_1, \cdots, W_K$ to the coded symbols $V_1, \cdots, V_N$ that satisfies the decoding constraint (2) specified by a graph $G = (\mathcal{V}, \mathcal{E}, t)$ is called a storage code. The (achievable) symbol rate of a storage code is defined as

$$R \triangleq \frac{L_w}{L_v} \tag{3}$$

[1]Note that we abuse the notation by using V_n to denote both a coded symbol and a node of the graph, which will not cause confusion.

and the supremum of symbol rate is called the capacity, $C \triangleq \sup_{L_w} L_w/L_v = \lim_{L_w\to\infty} L_w/L_v$, as block codes are allowed.

Next we introduce some graph definitions to facilitate the presentation of our results.

A. Graph Definitions

Definition 1 (W_k-Edge, W_k-Path, and W_k-Component): An edge that is associated with W_k is called a W_k-edge. A sequence of distinct connecting W_k-edges is called a W_k-path. A W_k-component is a maximal subgraph wherein every edge is a W_k-edge and every two nodes are connected by a W_k-path (an isolated node is defined as a trivial component).

For example, in Fig. 1, $\{V_1, V_2\}$ (also all solid black edges) is a W_1-edge; the sequence of W_1-edges $(\{V_2, V_1\}, \{V_1, V_5\}, \{V_5, V_6\})$ is a W_1-path and also a W_1-component (V_2, V_1, V_5, V_6 are connected by W_1-edges/paths and there are no more W_1-edges to extend the connectivity).

Definition 2 (Internal Edge and Residing Path): A W_k-edge that connects two nodes (say V_i, V_j) in a $W_{k'}$-path, $k' \neq k$ is said to be internal and the $W_{k'}$-path with end nodes V_i, V_j is called the residing path of the internal W_k-edge $\{V_i, V_j\}$.

For example, in Fig. 1, the W_2-edge $\{V_2, V_6\}$ is an internal edge as it connects two nodes V_2, V_6 in the W_1-path $(\{V_2, V_1\}, \{V_1, V_5\}, \{V_5, V_6\})$, which is then its residing path.

Definition 3 (M-Color Node): A node whose connected edges are associated with M different source symbols is called an M-color node.

For example, in Fig. 1, V_1, V_9 are 1-color nodes and V_5, V_6 are 2-color nodes.

We need to further distinguish two types of 2-color nodes, defined as follows.

Definition 4 (Normal 2-Color Node and W_k-Special 2-Color Node): For a 2-color node V that is connected to W_k-edges and $W_{k'}$-edges, $k \neq k'$, if the nodes connected to V through W_k-edges are all 1-color, then V is called a W_k-special 2-color node (or just a special 2-color node when W_k does not need to be highlighted). A 2-color node that is not special is said to be normal.

For example, in Fig. 1, the 2-color node V_5 is W_2-special as V_9 is the only node that is connected to V_5 through W_2-edges and V_9 is 1-color; the 2-color node V_6 is normal as it is connected to a 2-color node V_2 through a W_2-edge and is connected to a 2-color node V_5 through a W_1-edge.

Definition 5 (Graph Class $\mathcal{G}_{C=R^*}, \mathcal{G}_{C\geq R^*}, \mathcal{G}_{C<R^*}$): The set of graphs whose storage code capacity is equal to\no smaller than\strictly smaller than R^* is denoted by $\mathcal{G}_{C=R^*} \backslash \mathcal{G}_{C\geq R^*} \backslash \mathcal{G}_{C<R^*}$.

III. Results

Our results are presented in this section, along with illustrative examples and observations.

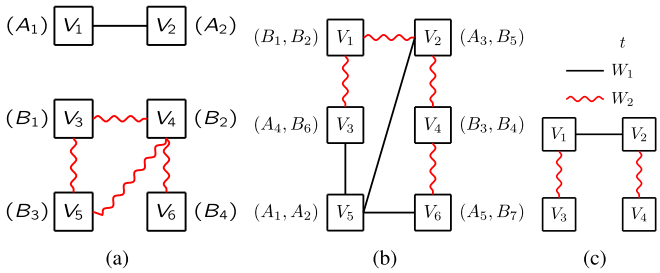

Fig. 3. (a) An example graph $G \in \mathcal{G}_{C=2}$. $W_1 = (a_1, a_2), W_2 = (b_1, b_2)$, and each $A_i \backslash B_j$ is a generic linear combination of $(a_1, a_2) \backslash (b_1, b_2)$. (b) An example graph $G \in \mathcal{G}_{C=3/2}$. $W_1 = (a_1, a_2, a_3), W_2 = (b_1, b_2, b_3)$, and each $A_i \backslash B_j$ is a generic linear combination of $(a_1, a_2, a_3) \backslash (b_1, b_2, b_3)$. (c) An example graph $G \in \mathcal{G}_{C<3/2}$ where two 2-color nodes V_1, V_2 are connected.

A. Extremal Graphs With Storage Code Capacity $2, 3/2$: $\mathcal{G}_{C=2}, \mathcal{G}_{C=3/2}$

The three highest extremal capacity values and the full extremal graph characterization for the two highest extremal capacity values are established in the following theorem.

Theorem 1: [$\mathcal{G}_{C=2}, \mathcal{G}_{C=3/2}$] The three highest storage code capacity values are $2, 3/2, 4/3$. The storage code capacity of a graph is equal to 2 ($G \in \mathcal{G}_{C=2}$) if and only if every node is 1-color. The storage code capacity of a graph is equal to $3/2$ ($G \in \mathcal{G}_{C=3/2}$) if and only if all nodes are 2-color or 1-color (and 2-color nodes exist) and there are no connected 2-color nodes.

The proof of Theorem 1 is fairly straightforward and is deferred to Section IV-A. An example of the achievable scheme (code construction) is shown in Fig. 3.(a) and Fig. 3.(b). An example graph that does not belong to $\mathcal{G}_{C=2} \cup \mathcal{G}_{C=3/2}$ is shown in Fig. 3.(c). An intuitive explanation on why the rate is upper bounded by $4/3$ is as follows. V_3 can at most contribute L_v bits of information about W_2. $\{V_1, V_3\}$ is a W_2-edge so that V_1 has to provide at least the remaining $L_w - L_v$ bits of information about W_2, leaving at most $L_v - (L_w - L_v) = 2L_v - L_w$ bits of room for W_1. The same reasoning applies to V_2. Finally, $\{V_1, V_2\}$ is a W_1-edge so that the size of the remaining room must accommodate the L_w bits of W_1, i.e., $2(2L_v - L_w) \geq L_w$ so that $R = L_w/L_v \leq 4/3$.

B. Extremal Graphs With Capacity $4/3$: $\mathcal{G}_{C=4/3}$ With $K = 2$ Source Symbols

Next we focus on the storage code capacity value of $4/3$, whose extremal graph characterization turns out to be highly non-trivial. In this section, we consider the cases where there are $K = 2$ source symbols to illustrate the results in a simpler setting and defer the generalizations to more than 2 source symbols to the next section.

The obtained necessary and sufficient conditions are rather involved. To make the results more clear we give a summarizing chart in Fig. 4.

1) Sufficient Condition: Internal Edge and 1-Color Node: A crucial graphic structure for the achievability of rate $4/3$ is the absence of internal edges (or when they exist, the presence of 1-color nodes in their residing paths). This result is stated in the following theorem.

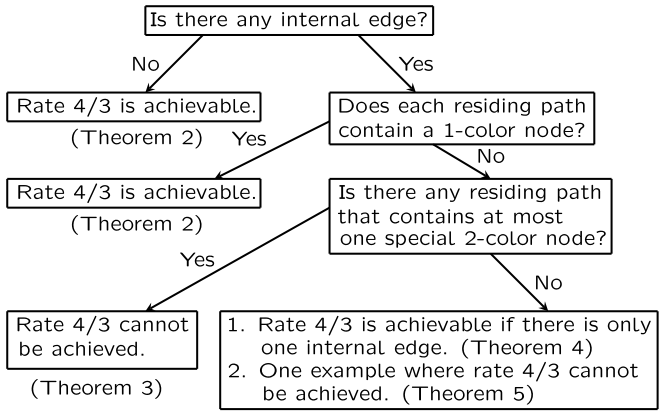

Fig. 4. A summary of sufficient and necessary conditions of $\mathcal{G}_{C=4/3}$ with $K=2$.

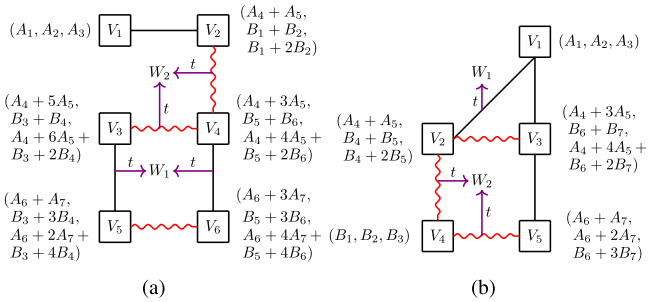

Fig. 5. Two example graphs $G \in \mathcal{G}_{C\geq 4/3}$ and code constructions for rate $4/3$. $W_1=(a_1,a_2,a_3,a_4), W_2=(b_1,b_2,b_3,b_4)$ and each $A_i\backslash B_j$ is a generic linear combination of $(a_1,\cdots,a_4)\backslash(b_1,\cdots,b_4)$.

Theorem 2: [Sufficient Condition of $\mathcal{G}_{C=4/3}$] With $K=2$ source symbols, a graph $G \in \mathcal{G}_{C\geq 4/3}$ if G contains no internal edge or for any internal edge, its residing path contains a 1-color node.

The proof of Theorem 2 is presented in Section IV-B. To illustrate the idea, two examples are shown in Fig. 5, where Example (a) contains no internal edge; Example (b) contains two internal edges $\{V_2,V_3\}$ and $\{V_3,V_5\}$. Internal W_2-edge $\{V_2,V_3\}$ resides in W_1-path $(\{V_2,V_1\},\{V_1,V_3\})$, which contains 1-color node V_1 and internal W_1-edge $\{V_3,V_5\}$ resides in W_2-path $(\{V_3,V_2\},\{V_2,V_4\},\ \{V_4,V_5\})$, which contains 1-color node V_4. So the condition of Theorem 2 is satisfied and rate $4/3$ is achievable. We next explain how to construct the code.

We are targeting at rate $L_w/L_v=4/3$ so that any pair of nodes connected by an edge contain $2L_v=3L_w/2$ bits. Except from L_w bits from the desired source, at most we can tolerate $2L_v-L_w=L_w/2$ undesired bits (i.e., interference). Then the key is to guarantee for any W_k-edge, $k\in\{1,2\}$, the interference from W_{3-k} has at most half source size. That is, W_{3-k} symbols shall be assigned according to W_k-edges (W_k-components). When there is no internal edge (or residing path contains 1-color nodes), such interference based assignment automatically ensures the independence (thus decodability) of desired source symbols. We now come back to the examples in Fig. 5 to see how to implement the above code design idea.

Consider Example (a) first and Example (b) will follow similarly. We set $L_w/\log_2 p=4$ so that $W_1=(a_1,a_2,a_3,a_4)$ and $W_2=(b_1,b_2,b_3,b_4)$, where each symbol is from a sufficiently large finite field $\mathbb{F}_p$ (the exact field size will be given in the general proof in Section IV-B). To achieve rate $R=L_w/L_v=4/3$, we set $L_v=3\log_2 p$, i.e., each V_n contains three symbols from the same field. We generate a number of generic linear combinations of $(a_1,\cdots,a_4)\backslash(b_1,\cdots,b_4)$ and denote them as $(A_1,A_2,\cdots)\backslash(B_1,B_2,\cdots)$. For now, it suffices to view each $A_i\backslash B_j$ as a random linear combination of symbols from $W_1\backslash W_2$ and if we can collect four linearly independent combinations of $A_i\backslash B_j$, then we can recover $W_1\backslash W_2$. The detailed randomized construction is again deferred to the general proof. Each one of the three symbols in V_n will be a linear combination of some A_i and B_j symbols. We first assign the A_i and B_j symbols in each V_n and then linearly combine them to produce the final three symbols in V_n.

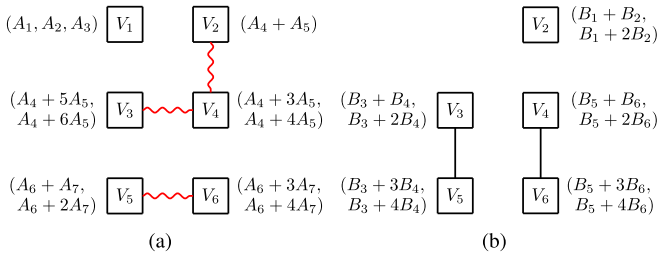

Fig. 6. (a) W_2-component decomposition of W_1-connected nodes in Fig. 5.(a), according to which A_i symbols are assigned. (b) W_1-component decomposition of W_2-connected nodes in Fig. 5.(a), according to which B_j symbols are assigned.

Consider nodes that are connected to W_1-edges so that some A_i symbols need to be assigned, i.e., all nodes $V_1,\cdots,V_6$. The 1-color nodes are trivial (i.e., V_1), and we just assign three distinct A_i symbols. Next, consider the remaining 2-color nodes $V_2,\cdots,V_6$ for which the A_i symbols are assigned according to W_2-components (see Fig. 6.(a)). $V_2,\cdots,V_6$ form two W_2-components - one consists of V_2,V_3,V_4 and the other consists of V_5,V_6. For each W_2-component, we assign generic linear combinations of the *same* $2=\frac{1}{2}L_w/\log_2 p$ A_i symbols (say A_{i_1},A_{i_2}) so that the interference dimension is limited to two. Further, a normal 2-color node and a W_2-special 2-color node will get two generic linear combinations of (A_{i_1},A_{i_2}) and a W_1-special 2-color node will get one generic linear combination of (A_{i_1},A_{i_2}). For example, consider W_2-component with nodes V_2,V_3,V_4, where the A_i symbols appeared are limited to A_4,A_5; V_2, as a W_1-special 2-color node, gets one combination A_4+A_5 and V_3,V_4, as normal 2-color nodes, each gets two generic combinations (e.g., V_3 gets A_4+5A_5, A_4+6A_5). The other W_2-component with nodes V_5,V_6 is assigned similarly - the A_i symbols are limited to A_6,A_7.

The assignment for nodes connected to W_2-edges is exactly the same (see Fig. 6.(b)). Nodes $V_2,\cdots,V_6$ are connected to W_2-edges and they are all 2-color. The B_j symbols are assigned according to W_1-components, i.e., V_2 (as a single-node component) gets generic linear combinations of B_1,B_2; V_3,V_5 form a W_1-component and get generic linear

combinations of B_3, B_4; V_4, V_6 form a W_1-component and the B_j symbols are limited to B_5, B_6.

The last step is to combine the A_i, B_j symbols so that each V_n has only three symbols. This step is simple, if a node gets at most three A_i, B_j symbols, then just set them as V_n (e.g., V_1, V_2); otherwise the node must be normal 2-color, which gets two generic combinations of A_i and two generic combinations of B_j and we just add one arbitrary combination (say the last) of A_i and B_j together to reduce the total number of symbols to three (e.g., V_3, V_4, V_5, V_6).

Finally, let us verify why the decoding constraints (2) are satisfied. An edge that contains 1-color node is straightforward, e.g., from W_1-edge $\{V_1, V_2\}$, we have $A_1, A_2, A_3, A_4 + A_5$, so as long as the A_i combinations are generic we can recover $W_1 = (a_1, \cdots, a_4)$. For edges that connect two 2-color nodes (e.g., W_2-edge $\{V_3, V_4\}$), we have 1) the interference dimension is limited to two as our assignment is based on components of interfering sources (e.g., we may decode A_4, A_5 and remove them, leaving us with only B_j symbols); 2) the four symbols from the desired source have full rank (e.g., B_3, B_4, B_5, B_6 are generic combinations) so that we can recover the desired source symbol. Note that because there is no internal edge, for any W_k-edge, the two nodes obtain distinct desired W_k symbols, e.g., for W_2-edge $\{V_3, V_4\}$, V_3 is assigned B_3, B_4 symbols and V_4 is assigned B_5, B_6 symbols as V_3, V_4 belong to distinct W_1-components (refer to Fig. 6.(b)). If V_3, V_4 belong to the same W_1-component, then the W_2-edge $\{V_3, V_4\}$ will be internal).

The code construction for Example (b) in Fig. 5 follows from the same procedure as that of Example (a). That is, first consider 1-color nodes and assign generic combinations (e.g., V_1, V_4); for remaining 2-color nodes, assign W_k symbols according to W_{3-k}-components (e.g., the W_1 space of the W_2-edge $\{V_2, V_3\}$ is spanned by A_4, A_5, and the W_2 space of the W_1-edge $\{V_3, V_5\}$ is spanned by B_6, B_7); finally combine the four symbols to three for normal 2-color nodes (e.g., V_3). The decoding constraints (2) are easily verified as the interference dimension is strictly controlled and desired source symbols are sufficiently generic because after removing 1-color nodes, there no longer exist internal edges.

2) Necessary Condition: Residing Path and Special 2-Color Node: The sufficient condition of the achievability of rate $4/3$ in Theorem 2 requires the absence of internal edges or the presence of 1-color node in residing paths. Considering the complementary cases, we identify a crucial graphic structure for the unachievability of rate $4/3$ - the presence of at most one special 2-color node in a residing path. This result is stated in the following theorem.

Theorem 3: [Necessary Condition of $\mathcal{G}_{C=4/3}$] With $K = 2$ source symbols, a graph $G \in \mathcal{G}_{C<4/3}$ if G has a residing path which contains no 1-color node and at most one special 2-color node.

The proof of Theorem 3 is presented in Section IV-C. To illustrate the idea, an example is shown in Fig. 7, where the internal W_2-edge $\{V_1, V_2\}$ resides in the W_1-path $(\{V_1, V_3\}, \{V_3, V_4\}, \{V_4, V_2\})$ and this residing path contains only one special 2-color node V_3 and no 1-color node. So the condition of Theorem 3 is satisfied and rate $4/3$ cannot be achieved. To see why, we next give an intuitive explanation by contradiction.

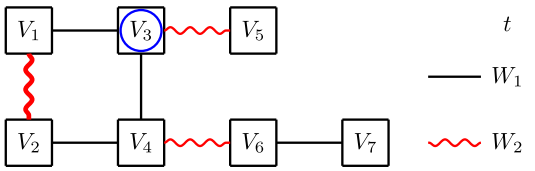

Fig. 7. An example graph $G \in \mathcal{G}_{C<4/3}$ where the internal edge $\{V_1, V_2\}$ is highlighted and the only special 2-color node V_3 in its residing path is highlighted.

Suppose rate $4/3$ is achievable, i.e., $L_w/L_v = 4/3$. Then we can show that for any 2-color node (e.g., V_3), it must contain at least $L_w/4$ bits of information about each of W_1 and W_2 (captured through conditional entropy. See Lemma 1 in Section IV-C). This is because the connecting node can provide at most $L_v = 3L_w/4$ bits of information about the desired source symbol (e.g., V_5 can contribute $L_v = 3L_w/4$ bits on W_2 at most and the remaining $L_w - L_v = L_w/4$ bits must come from V_3). Further, if the 2-color node is normal (e.g., V_4), it must contain exactly $L_w/2$ bits of information about each of W_1 and W_2 (see Lemma 2). The reason is that for two connecting 2-color nodes, the amount of interference allowed is at most $2L_v - L_w = L_w/2$ bits and a pair of nodes must contribute L_w bits of information about the desired source symbol (thus $L_w/2$ from each node). For example, consider W_1-edge $\{V_2, V_4\}$, where from an interference view, V_2 can contain at most $L_w/2$ bits on W_2; from the desired source view, V_2 must also contribute at least $L_w/2$ bits on W_2 because of the W_2-edge $\{V_1, V_2\}$.

We now consider the propagation of interference through the residing W_1-path $(\{V_2, V_4\}, \{V_4, V_3\}, \{V_3, V_1\})$. Start from the normal 2-color node V_2, which contains $L_w/2$ bits on W_2 and as a W_1-edge can tolerate at most $L_w/2$ bits on W_2, then the normal 2-color node V_4 must contain the same $L_w/2$ bits on W_2 (see Lemma 3). We are now at V_4 and continue the W_1-path through edge $\{V_3, V_4\}$, where V_3 is special so that V_3 contains at least $L_w/4$ bits on W_2 and this $L_w/4$ bits are contained in the total $L_w/2$ interference bits in V_4. Continue further the W_1-path through edge $\{V_3, V_1\}$, where the $L_w/4$ bits on W_2 in V_3 must be contained in the $L_w/2$ bits on W_2 in V_1. This in turn means that the $L_w/2$ bits on W_2 in V_1 must overlap with the $L_w/2$ bits on W_2 in V_2 (in the $L_w/4$ bits on W_2 in V_3), thus the internal W_2-edge $\{V_1, V_2\}$ cannot contribute $L_w/2 + L_w/2 = L_w$ independent bits for the desired W_2 source and we have arrived at a contradiction.

From the above reasoning, we can now illuminate the role of special and normal 2-color nodes in a residing path. For an internal W_k-edge, its residing W_{3-k}-path made up of 2-color nodes must have two normal 2-color end nodes, each of which contains $L_w/2$ independent bits of information about the desired source W_k (e.g., V_1, V_2 about W_2). In the residing W_{3-k}-path, a normal 2-color node will keep the interference on W_k to the same $L_w/2$ dimensions (e.g., V_2, V_4 have the same $L_w/2$ dimensions about W_2 and V_1, V_3 have the same

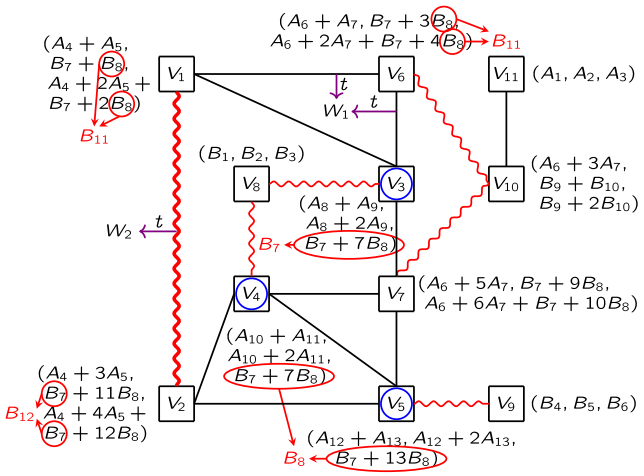

Fig. 8. An example graph $G \in \mathcal{G}_{C\geq 4/3}$ with only one internal edge $\{V_1, V_2\}$ (highlighted). Special 2-color nodes V_3, V_4, V_5 are highlighted and each residing path has at least two of them. The code produced by the assignment of Theorem 2 and updates needed to produce the final code with rate $4/3$ are shown.

$L_w/2$ dimensions about W_2) while a special 2-color node will inherit at least $L_w/4$ interference dimensions on W_k (e.g., V_3 gets at least $L_w/4$ dimensions of W_2 from V_3). Conversely, a special 2-color node in a residing W_{3-k}-path can change at most $L_w/4$ dimensions of the interference on W_k (which is the desired source for the internal W_k-edge), so to ensure the independence of the desired source at the internal edge we need at least two special 2-color nodes in the residing path. This case is exactly our focus in the next section (along this line, we can also see the role of 1-color node in a residing path, i.e., it completely stops the propagation of interference. See the 1-color node V_4 in the residing W_2-path $(\{V_3, V_2\}, \{V_2, V_4\}, \{V_4, V_5\})$ of Fig. 6.(b), where V_3, V_5 can hold independent W_1 bits although the W_1-edge $\{V_3, V_5\}$ is internal).

3) Remaining Cases: Rate 4/3 May or May Not Be Achievable: Continuing the discussion in the previous paragraph, the cases that are not covered by Theorem 2 and Theorem 3 are those where each residing path contains at least two special 2-color nodes (and no 1-color node). This setting turns out to be quite intricate and is not fully understood. In the following, we show that here $4/3$ may or may not be achievable, depending on the structure of other parts of the graph.

On the one hand, we show that if there is only one internal edge, then the presence of two special 2-color nodes in the residing path is sufficient to achieve rate $4/3$. This result is stated in the following theorem.

Theorem 4: With $K = 2$ source symbols, a graph $G \in \mathcal{G}_{C\geq 4/3}$ if G contains only one internal edge and its every residing path has at least two special 2-color nodes.

The proof of Theorem 4 is presented in Section IV-D. To illustrate the idea, an example is shown in Fig. 8, where for the only internal edge $\{V_1, V_2\}$, three special 2-color nodes V_3, V_4, V_5 ensure that at least two of them are contained in any residing path.

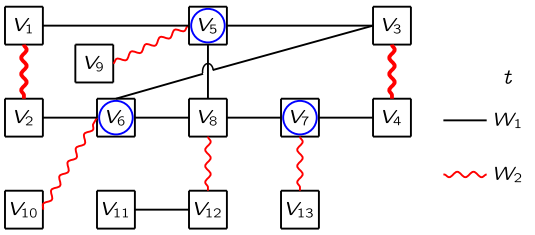

Fig. 9. A graph $G \in \mathcal{G}_{C<4/3}$ albeit each residing path has two special 2-color nodes. G contains two internal edges $\{V_1, V_2\}$, $\{V_3, V_4\}$ and three special 2-color nodes V_5, V_6, V_7 in residing paths.

To assign the code, we first follow the same procedure in Theorem 2 to assign the coded symbols (see Fig. 8). Because of the presence of the internal W_2-edge $\{V_1, V_2\}$ (and absence of 1-color nodes in the residing paths), the desired W_2 symbols are not independent (i.e., only B_7, B_8 appears in V_1, V_2 while we need four B_j symbols to recover W_2). So we need to expand the dimension of the W_2 symbols to satisfy the decoding constraint (2) for the internal edge $\{V_1, V_2\}$. This is done by replacing $B_8 \backslash B_7$ in $V_1 \backslash V_2$ with another generic $B_{11} \backslash B_{12}$ symbol (see Fig. 8), but now the interference on W_2 in the residing path will not be limited to only two dimensions. A final update is required - starting from V_2, we visit each residing path to find its closest special 2-color nodes, which turn out to be V_4, V_5 and remove B_7 therein to ensure the interference along this path is limited to B_8, B_{12} two dimensions (see Fig. 8). Repeat the same for V_1, i.e., visit each residing path starting from V_1, find the closest special 2-color nodes, which turn out to be V_3, and only keep B_7 (remove B_8) at V_3. Such special 2-color nodes are guaranteed to exist as each residing path has at least two special 2-color nodes. Also replace B_8 by B_{11} for each node visited along the residing paths so that now again the interference dimension is limited to B_7, B_{11} (i.e., V_6. See Fig. 8). The update is complete and decoding constraints (2) are all satisfied (refer to Fig. 8 for a verification). Indeed, we may see that the role of each special 2-color node along the residing W_k-path is to replace $L_w/4$ dimensions of W_{3-k} so that with two special 2-color nodes we may have fully independent $L_w/2$ dimensions of W_{3-k} for the two nodes in the internal $W_{3-k}-$edge.

On the other hand, we show that for the graph in Fig. 9, rate $4/3$ cannot be achieved even if each residing path contains two special 2-color nodes. This result is stated in the following theorem.

Theorem 5: The storage code capacity of the graph G in Fig. 9 is strictly smaller than $4/3$.

The proof of Theorem 5 is presented in Section IV-E. An intuitive explanation, which builds upon and generalizes the converse arguments in Theorem 3, on the unachievablity of rate $4/3$ is given here. Suppose rate $4/3$ can be achieved. Consider the internal W_2-edge $\{V_1, V_2\}$, where V_1, V_2 are normal 2-color and each must contain independent $L_w/2$ bits of information about W_2. Due to the W_1-edges $\{V_1, V_5\}$ and $\{V_2, V_6\}$, the special 2-color node V_5 must inherit $L_w/4$ bits on W_2 from V_1 (because the total amount of interference about W_2 in any W_1-edge cannot exceed $L_w/2$ bits) and the special 2-color node V_6 must inherit $L_w/4$ bits on W_2 from V_2. Note

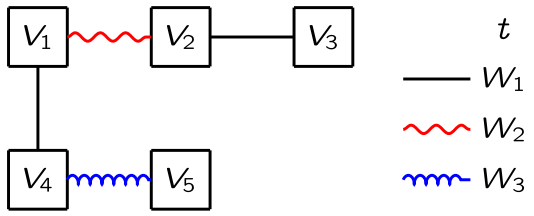

Fig. 10. An example graph $G \in \mathcal{G}_{C<4/3}$, where a normal 2-color node V_1 is connected to a special 2-color node V_4 and V_1, V_4 are connected to different types of edges.

now that V_3 is connected to both V_5, V_6 with W_1-edges, so the $L_w/2$ bits of information about W_2 in the normal 2-color node V_3 must contain the $L_w/4$ bits of information about W_2 in V_5 and V_6. Further, V_5 and V_6 contain independent information about W_2. So the $L_w/2$ bits on W_2 in V_3 is exactly the union of the $L_w/4$ bits on W_2 in V_5 and V_6. From the same reasoning, as V_8 is connected to both V_5 and V_6 with W_1-edges, V_8 contains exactly the same $L_w/2$ bits of information about W_2 as V_3. This will cause a contradiction because $(\{V_8, V_7\}, \{V_7, V_4\})$ belongs to a residing path and V_7 is a special 2-color node, so V_4 must share $L_w/4$ bits on W_2 with V_8 (thus V_3), which contradicts the fact that $\{V_3, V_4\}$ is an internal W_2-edge, i.e., V_3, V_4 must contain independent $L_w/2$ bits of information about W_2.

C. Extremal Graphs With Capacity 4/3: $\mathcal{G}_{C=4/3}$ With $K > 2$ Source Symbols

We now generalize the results on $\mathcal{G}_{C=4/3}$ from $K = 2$ source symbols to $K > 2$ source symbols. Let us start from necessary conditions, which include some new graphic structures with more than 2 sources that place rate constraints, and state the result in the following theorem.

Theorem 6: [Necessary Condition of $\mathcal{G}_{C=4/3}$] A graph $G \in \mathcal{G}_{C<4/3}$ if G contains 1) an M-color node, where $M \geq 4$, or 2) a 3-color code that is connected to an M-color code, where $M \geq 2$, or 3) a normal 2-color node V that is connected to a 2-color node whose connected edges are associated with a different set of source symbols from that connected to V.

The set of graphs that satisfy the conditions in Theorem 6 is denoted as $\mathcal{G}^{\text{Thm 6}}_{C<3/4}$. The first two conditions are easily seen and an example for the third condition is shown in Fig. 10. The proof of Theorem 6 is deferred to Section IV-F. We give an intuitive explanation here on why $R < 4/3$ for the graph G in Fig. 10. From (V_1, V_4, V_5), we can decode W_1, W_3, i.e., $2L_w$ bits. V_4, V_5 can contribute at most $2L_v$ bits on W_1, W_3 so that the remaining $2L_w - 2L_v$ bits must come from V_1, leaving only $L_v - (2L_w - 2L_v) = 3L_v - 2L_w$ bits of room for W_2. Similarly, V_2 has at most $L_v - (L_w - L_v) = 2L_v - L_w$ bits of room for W_2 because at least $L_w - L_v$ bits of W_1 must come from V_2 (consider the W_1-edge $\{V_2, V_3\}$). The W_2-edge $\{V_1, V_2\}$ needs to have at least L_w bits of room for the desired source W_2, i.e., $(3L_v - 2L_w) + (2L_v - L_w) \geq L_w$ so that $R = L_w/L_v \leq 5/4 < 4/3$.

Interestingly, if we exclude the graphs in $\mathcal{G}^{\text{Thm 6}}_{C<3/4}$, i.e., those for which rate $4/3$ cannot be achieved, then the sufficient condition in Theorem 2 generalizes immediately to more than

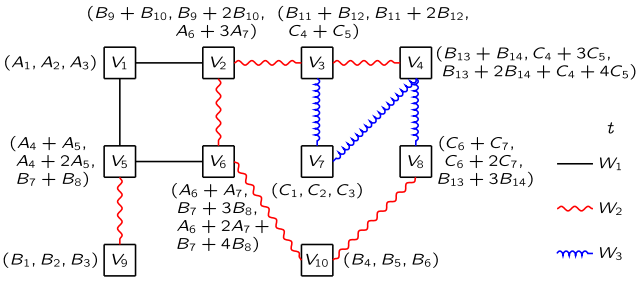

Fig. 11. An example graph $G \in \mathcal{G}_{C\geq 4/3}$ with $K = 3$ source symbols and code construction for rate $4/3$. $W_1 = (a_1, \cdots, a_4)$, $W_2 = (b_1, \cdots, b_4), W_3 = (c_1, \cdots, c_4)$ and each $A_i \backslash B_j \backslash C_m$ is a generic linear combination of $(a_1, \cdots, a_4) \backslash (b_1, \cdots, b_4) \backslash (c_1, \cdots, c_4)$.

$K = 2$ source symbols. This result is stated in the following theorem.

Theorem 7 (Sufficient Condition of $\mathcal{G}_{C=4/3}$): A graph $G \in \mathcal{G}_{C\geq 4/3}$ if $G \notin \mathcal{G}^{\text{Thm 6}}_{C<3/4}$ and G contains no internal edge or for any internal edge, its residing path contains a 1-color node.

The code construction of Theorem 7 is almost identical to that of Theorem 2 and it turns out to work as long as the structures in Theorem 6 are avoided. The detailed proof is deferred to Section IV-G and an example is shown in Fig. 11 to illustrate the idea. The assignment is still interference based, i.e., for each source symbol W_k, decompose all nodes connected to W_k-edges according to $W_{k'}$-components, where $k' \neq k$ (each node will belong only to one such component) and assign the same W_k symbols within the same $W_{k'}$-components. After this operation, the interference dimension is controlled; the absence of internal edges (after removing 1-color nodes) will guarantee the independence of desired symbols. The condition $G \notin \mathcal{G}^{\text{Thm 6}}_{C<3/4}$ helps to guarantee that if a pair of connecting 2-color nodes are connected to edges associated with more than 2 source symbols (e.g., in Fig. 11, $\{V_2, V_3\}$ are associated with 3 sources), the pair of nodes must be special 2-color and the interference caused by the two interfering source symbols is still limited to $L_w/2$ dimensions (e.g., in Fig. 11, V_2, V_3 are special 2-color and for W_2-edge $\{V_2, V_3\}$, the interference is one A_i combination in V_2 and one C_m combination in V_3, i.e., two dimensions in total). Other edges are the same as those in Theorem 2 and decoding constraints (2) hold (see Fig. 11).

IV. Proofs

A. Proof of Theorem 1: $\mathcal{G}_{C=2}, \mathcal{G}_{C=3/2}$

In this section, we provide the full characterization of $\mathcal{G}_{C=2}$ and $\mathcal{G}_{C=3/2}$. From the proof, we can obtain that the three highest storage code capacity values are $2, 3/2, 4/3$.

1) If and Only If Condition of $\mathcal{G}_{C=2}$: We show that $G(\mathcal{V}, \mathcal{E}, t) \in \mathcal{G}_{C=2}$ if and only if every node $V \in \mathcal{V}$ is 1-color. We prove the if part and the only if part sequentially.

If Part: If every node $V \in \mathcal{V}$ is 1-color, then we prove that the capacity is 2. First, we show that $R \leq 2$. Note that we assume G has no isolated nodes, so G must contain one edge, say W_k-edge $\{V_i, V_j\}$. From the decoding constraint (2), we have

$$L_w \overset{(1)}{=} H(W_k) \overset{(2)}{=} I(V_i, V_j; W_k) \leq H(V_i, V_j) \leq 2L_v \quad (4)$$

$$\Rightarrow R \overset{(3)}{=} L_w/L_v \leq 2 \quad (5)$$

where the last inequality in (4) follows from the fact that each coded symbol V_i contains at most L_v bits.

Second, we show that symbol rate $R = 2$ is achievable, by an MDS code. Note that each node is 1-color; suppose there are M_k nodes that are only connected to W_k-edges, $k \in [K]$ and denote this set of nodes by $\mathcal{V}_k$. Choose the field size p to be a prime that is no smaller than $\max_{k\in[K]} M_k$. Set $L_w = 2\log_2 p, L_v = \log_2 p$ so that each source\coded symbol is comprised of $2\backslash 1$ symbols\symbol from $\mathbb{F}_p$ and the rate achieved is 2. Generate MDS coded symbols as follows.

$$W_k = (W_k(1); W_k(2)) \in \mathbb{F}_p^{2\times 1} \quad (6)$$
$$X_k = (X_k(1); \cdots; X_k(M_k)) \triangleq \mathbf{V}_k W_k \in \mathbb{F}_p^{M_k\times 1} \quad (7)$$

where $\mathbf{V}_k \in \mathbb{F}_p^{M_k\times 2}$ is a full rank Vandermonde matrix so that from any two elements of X_k, we can recover W_k (i.e., MDS). Finally, we assign each node in $\mathcal{V}_k$ a distinct element of X_k so that from any W_k-edge, we can decode W_k.

Only if Part: We show that if there exists an M-color node, $M \geq 2$, then $R < 2$ so that the capacity must be strictly smaller than 2 and further, the capacity will drop to $3/2$ at least. Suppose we have an M-color node V_i that is connected to W_{k_1}-edge $\{V_i, V_{j_1}\}, \cdots, W_{k_M}$-edge $\{V_i, V_{j_M}\}$, then

$$\begin{aligned} ML_w &\overset{(1)}{=} H(W_{k_1}, \cdots, W_{k_M}) && (8)\\ &\overset{(2)}{=} I(V_i, V_{j_1}, \cdots, V_{j_M}; W_{k_1}, \cdots, W_{k_M}) && (9)\\ &\leq H(V_i, V_{j_1}, \cdots, V_{j_M}) && (10)\\ &\leq (M+1)L_v && (11)\\ \Rightarrow R &\overset{(3)}{=} L_w/L_v \leq (M+1)/M \leq 3/2 < 2. && (12)\end{aligned}$$

2) If and Only If Condition of $\mathcal{G}_{C=3/2}$: We show that $G(\mathcal{V}, \mathcal{E}, t) \in \mathcal{G}_{C=3/2}$ if and only if every node $V \in \mathcal{V}$ is 2-color or 1-color (and there exists a 2-color node) and there are no connected 2-color nodes. We prove the if part and the only if part sequentially.

If Part: We show that $C = 3/2$ if the condition above is satisfied. First, G contains a 2-color node, so from (12) we have that $R \leq 3/2$. Second, we show that $R = 3/2$ is achievable, again by an MDS code. Note that each node is 1-color or 2-color. Consider the nodes that are connected to W_k-edges, among which suppose M_k^1 are 1-color (denote this set by $\mathcal{V}_k^1$) and M_k^2 are 2-color (denote this set by $\mathcal{V}_k^2$). Choose the field size p to be a prime that is no smaller than $\max_k(2M_k^1 + M_k^2)$. Set $L_w = 3\log_2 p, L_v = 2\log_2 p$, i.e., each source\coded symbol is comprised of $3\backslash 2$ symbols from $\mathbb{F}_p$ and the rate achieved is $3/2$. Generate MDS coded symbols as follows.

$$W_k = (W_k(1); W_k(2); W_k(3)) \in \mathbb{F}_p^{3\times 1} \quad (13)$$
$$\begin{aligned} X_k &= (X_k(1); \cdots; X_k(2M_k^1 + M_k^2))\\ &\triangleq \mathbf{V}_k W_k \in \mathbb{F}_p^{(2M_k^1+M_k^2)\times 1} && (14)\end{aligned}$$

where $\mathbf{V}_k \in \mathbb{F}_p^{(2M_k^1+M_k^2)\times 3}$ is a full rank Vandermonde matrix so that from any three elements of X_k, we can recover W_k (i.e., MDS). The existence of such a full rank Vandermonde matrix is guaranteed due to our field size choice. Finally, we assign each node in $\mathcal{V}_k^1\backslash\mathcal{V}_k^2$ two\one distinct elements\element of X_k. Note that any node $V \in \mathcal{V}$ will be assigned two $\mathbb{F}_p$ symbols. To verify that the decoding constraint (2) holds, consider any W_k-edge $\{V_i, V_j\}$, where V_i, V_j cannot both be 2-color because from our condition of $\mathcal{G}_{C=3/2}$, 2-color nodes do not connect. As a 2-color node contains one element of X_k and a 1-color node contains two elements of X_k, (V_i, V_j) will contain at least three elements of X_k, from which we can recover W_k.

Only if Part: We show that $C \neq 3/2$ if the condition of $\mathcal{G}_{C=3/2}$ is violated, i.e., if 1) there only exist 1-color nodes, 2) there is an M-color node, where $M \geq 3$, or 3) if there are connected 2-color nodes, say V_i, V_j. For Case 1), $G \in \mathcal{G}_{C=1}$; for Case 2), from (12) we have $R \leq 4/3 < 3/2$; for Case 3), we next show that $R \leq 4/3$ so that the proof is complete and when capacity 2 and $3/2$ cannot be achieved, it drops to $4/3$ at least (and rate $4/3$ is achievable for some graph, e.g., Theorem 2, so the third-highest capacity value is $4/3$).

Suppose $\{V_i, V_j\}$ is a W_k-edge. As V_i is 2-color, V_i must be connected to some node $V_{i_1}, i_1 \neq i$ with a W_{k_1}-edge, where $k_1 \neq k$. Consider the W_{k_1}-edge $\{V_i, V_{i_1}\}$ and we have

$$\begin{aligned} L_w &\overset{(1)}{=} H(W_{k_1}) && (15)\\ &\overset{(2)}{=} I(V_i, V_{i_1}; W_{k_1}) && (16)\\ &= H(V_i, V_{i_1}) - H(V_i, V_{i_1}|W_{k_1}) && (17)\\ &\leq 2L_v - H(V_i, V_{i_1}|W_{k_1}) && (18)\\ \Rightarrow H(V_i|W_{k_1}) &\leq H(V_i, V_{i_1}|W_{k_1}) \leq 2L_v - L_w. && (19)\end{aligned}$$

Symmetrically, V_j is 2-color so that V_j must be connected to $V_{j_1}, j_1 \neq j$ with a W_{k_2}-edge, where $k_2 \neq k$. Note that j_1 may be the same as i_1 and k_2 may be the same as k_1. The same proof will work under all circumstances. Consider the W_{k_2}-edge $\{V_j, V_{j_1}\}$. Following the derivation of (19), we have

$$H(V_j|W_{k_2}) \leq 2L_v - L_w. \quad (20)$$

Finally consider the W_k-edge $\{V_i, V_j\}$ and we have

$$\begin{aligned} L_w &\overset{(1)}{=} H(W_k|W_{k_1}, W_{k_2}) && (21)\\ &\leq H(W_k, V_i, V_j|W_{k_1}, W_{k_2}) && (22)\\ &\overset{(2)}{=} H(V_i, V_j|W_{k_1}, W_{k_2}) && (23)\\ &\leq H(V_i|W_{k_1}) + H(V_j|W_{k_2}) && (24)\\ &\overset{(19)(20)}{\leq} 2L_v - L_w + 2L_v - L_w && (25)\\ \Rightarrow R &\overset{(3)}{=} L_w/L_v \leq 4/3. && (26)\end{aligned}$$

B. Proof of Theorem 2: Sufficient Condition of $\mathcal{G}_{C=4/3}$ With $K = 2$

We show that if a graph[2] $G(\mathcal{V}, \mathcal{E})$ contains no internal edge or each residing path contains one 1-color node, then $R = 4/3$ is achievable. We first present the code construction and then prove it satisfies the decoding constraint (2).

[2]For simplicity, the edge association mapping t is omitted from the graph notation G in this section.

1) Code Construction: Choose the field size p to be a prime that is greater than $4|\mathcal{E}|$. Set $L_w = 4\log_2\ p, L_v = 3\log_2 p$ so that each source\coded symbol is comprised of $4\backslash 3$ symbols from $\mathbb{F}_p$ and the rate achieved is $4/3$.

Consider the set of nodes that are connected to W_k-edges, $k \in \{1,2\}$ and denote this set by $\mathcal{V}_k$. Consider the subgraph of $G(\mathcal{V},\mathcal{E})$ whose node set is $\mathcal{V}_k$ and edge set is comprised of all W_{3-k}-edges that are connected to some node in $\mathcal{V}_k$, denoted by $\mathcal{E}_{3-k}$ and denote this subgraph by $G_k(\mathcal{V}_k,\mathcal{E}_{3-k})$. Decompose $G_k(\mathcal{V}_k,\mathcal{E}_{3-k})$ into W_{3-k}-components and suppose we have M_k such components. Among these M_k W_{3-k}-components, suppose M_k^1 components are comprised of 1-color nodes (each such component is an isolated node) and label them as $P_k^{[1]},\cdots,P_k^{[M_k^1]}$; the remaining $M_k^2 = M_k - M_k^1$ components are comprised of 2-color nodes and label them as $Q_k^{[1]},\cdots,Q_k^{[M_k^2]}$. For an example of subgraph $G_k(\mathcal{V}_k,\mathcal{E}_{3-k})$ and its decomposition, refer to Fig. 6.

Generate generic linear combinations of the source symbols as follows.

$$\begin{aligned} W_k &= (W_k(1);\cdots;W_k(4)) \in \mathbb{F}_p^{4\times 1}, k\in\{1,2\} \quad (27)\\ X_k &= (X_k(1);\cdots;X_k(3M_k^1+2M_k^2))\\ &\triangleq \mathbf{H}_k W_k \in \mathbb{F}_p^{(3M_k^1+2M_k^2)\times 1} \quad (28)\end{aligned}$$

where $\mathbf{H}_k$ is a $(3M_k^1+2M_k^2)\times 4$ matrix over the field $\mathbb{F}_p$ and each element of $\mathbf{H}_k$ is chosen uniformly and independently from $\mathbb{F}_p$. Thus our construction is randomized and we will show that the probability that all decoding constraints (2) are satisfied is strictly larger than 0 so that one feasible code construction exists.

Consider $P_k^{[m]}, m \in [M_k^1]$, denote its node by V, and set

$$V = \big(X_k(3m-2), X_k(3m-1), X_k(3m)\big). \quad (29)$$

This step completes the assignment for all 1-color nodes, each of which must reside in one $P_k^{[m]}$.

Consider $Q_k^{[m]}, m \in [M_k^2]$. Suppose $Q_k^{[m]}$ contains J nodes $V_{i_1},\cdots,V_{i_J}$ (each must be 2-color). Consider node $V_{i_j}, j\in[J]$.

$$\begin{aligned} &\text{If } V_{i_j} \text{ is } W_k\text{-special, set} V_{i_j}^{[k]} \triangleq\\ &X_k\big(3M_k^1+2m-1\big)+(2j-1)X_k\big(3M_k^1+2m\big); \quad (30)\\ &\text{otherwise, set } V_{i_j}^{[k]} = (V_{i_j}^{[k]}(1), V_{i_j}^{[k]}(2))\\ &\triangleq \Big(X_k\big(3M_k^1+2m-1\big)+(2j-1)X_k\big(3M_k^1+2m\big),\\ &X_k\big(3M_k^1+2m-1\big)+2jX_k\big(3M_k^1+2m\big)\Big). \quad (31)\end{aligned}$$

Finally, after setting $V^{[1]}, V^{[2]}$ for each 2-color node V, we are ready to set V.

$$\begin{aligned} &\text{If } V \text{ is normal, then set}\\ &V = (V^{[1]}(1), V^{[2]}(1), V^{[1]}(2)+V^{[2]}(2)); \quad (32)\\ &\text{otherwise, set } V = (V^{[1]}, V^{[2]}). \quad (33)\end{aligned}$$

Note that for a special 2-color node, at least one of $V^{[1]}, V^{[2]}$ will be one symbol so that V will contain no more than three symbols (when a node V is simultaneously W_1-special and W_2-special, V will have only two symbols and we may zero-pad to make its length three). This completes the assignment for all 2-color nodes and the code construction is complete.

2) Proof of Correctness: We show that the decoding constraint (2) is satisfied. Consider any edge $\{V_i,V_j\} \in \mathcal{E}$ and suppose it is a W_k-edge.

When V_i, V_j contain one 1-color code, say V_i, then V_i contains three elements of X_k (say, $X_k(m_1), X_k(m_1+1), X_k(m_1+2)$; refer to (29)) and V_j contains at least one generic combination of distinct two elements of X_k or one distinct element of X_k (say, $X_k(m_2)+jX_k(m_2+1)$ or $X_k(m_2)$; refer to (29) - (33)). These 4 symbols in X_k can be written as a multiplication of a 4×4 matrix, denoted by $\mathbf{T}_{ij}$ and the source symbol vector W_k. View the determinant of $\mathbf{T}_{ij}$ as a polynomial $T_{ij}(\mathbf{H}_1,\mathbf{H}_2)$, whose variables are the elements of $\mathbf{H}_1,\mathbf{H}_2$ (refer to (69)). $T_{ij}(\mathbf{H}_1,\mathbf{H}_2)$ is not a zero-polynomial as we may set $X_k(m_1)=W_k(1), X_k(m_1+1) = W_k(2), X_k(m_1+2) = W_k(3), X_k(m_2) = W_k(4), X_k(m_2+1)=0$ so that $\mathbf{T}_{ij}$ is an identity matrix and $T_{ij}(\mathbf{H}_1,\mathbf{H}_2)=1$.

We are left with cases where V_i, V_j are both 2-color. Note that V_i, V_j cannot be W_k-special. We have three cases.

1) V_i, V_j are both W_{3-k}-special. Then each of V_i, V_j contains two generic combinations of two distinct elements of X_k (refer to (31) and (33)), say

$$\begin{aligned} &X_k(m_1)+(2j_1-1)X_k(m_1+1),\\ &X_k(m_1)+2j_1X_k(m_1+1) \text{ from } V_i \text{ and}\\ &X_k(m_2)+(2j_2-1)X_k(m_2+1),\\ &X_k(m_2)+2j_2X_k(m_2+1) \text{ from } V_j \quad (34)\end{aligned}$$

where the elements of X_k are all distinct because V_i, V_j belong to different W_{3-k}-components in the decomposition of $G_k(\mathcal{V}_k,\mathcal{E}_{3-k})$. Otherwise, $\{V_i,V_j\}$ is an internal edge after removing 1-color nodes, which contradicts the condition of Theorem 2. From the four symbols in (34), we can recover four distinct elements of X_k, i.e., $(X_k(m_1); X_k(m_1+1); X_k(m_2); X_k(m_2+1))$, which can be similarly written as $\mathbf{T}_{ij}^{4\times 4}W_k$. View $\det(\mathbf{T}_{ij})$ as a polynomial $T_{ij}(\mathbf{H}_1,\mathbf{H}_2)$, which is not the zero-polynomial.

2) One of V_i, V_j is W_{3-k}-special, say V_i and the other is normal, say V_j. From (31) - (33), we have

$$\begin{aligned} V_i &= \big(X_k(m_1)+(2j_1-1)X_k(m_1+1),\\ &\quad X_k(m_1)+2j_1X_k(m_1+1),\\ &\quad X_{3-k}(m^*)+(2i_1-1)X_{3-k}(m^*+1)\big)\\ V_j &= \big(X_k(m_2)+(2j_2-1)X_k(m_2+1),\\ &\quad X_{3-k}(m^*)+(2i_2-1)X_{3-k}(m^*+1),\\ &\quad X_k(m_2)+2j_2X_k(m_2+1)\\ &\quad + X_{3-k}(m^*)+2i_2X_{3-k}(m^*+1)\big) \quad (35)\end{aligned}$$

where the elements of X_k are all distinct due to the same reason as above; the elements of X_{3-k} must be the same, i.e., m^* appears in both V_i, V_j because $\{V_i,V_j\}$ is a W_k-edge so that V_i, V_j belong to the same W_k-component in the decomposition of $G_{3-k}(\mathcal{V}_{3-k},\mathcal{E}_k)$.

Further, V_i, V_j are distinct so $i_1 \neq i_2$ in (35) (refer to (31)). Thus from (V_i, V_j), we can first decode and remove $X_{3-k}(m^*), X_{3-k}(m^*+1)$, leaving us with four distinct elements of X_k, i.e., $(X_k(m_1); X_k(m_1+1); X_k(m_2); X_k(m_2+1)) = \mathbf{T}_{ij}^{4\times 4} W_k$. View $\det(\mathbf{T}_{ij})$ as a polynomial $T_{ij}(\mathbf{H}_1, \mathbf{H}_2)$, which is non-zero.

3) V_i, V_j are both normal. From (31) - (33), we have

$$\begin{aligned} V_i &= \big(X_k(m_1) + (2j_1-1)X_k(m_1+1), \\ &\quad X_{3-k}(m^*) + (2i_1-1)X_{3-k}(m^*+1), \\ &\quad X_k(m_1) + 2j_1 X_k(m_1+1) \\ &\quad + X_{3-k}(m^*) + 2i_1 X_{3-k}(m^*+1) \\ V_j &= \big(X_k(m_2) + (2j_2-1)X_k(m_2+1), \\ &\quad X_{3-k}(m^*) + (2i_2-1)X_{3-k}(m^*+1), \\ &\quad X_k(m_2) + 2j_2 X_k(m_2+1) \\ &\quad + X_{3-k}(m^*) + 2i_2 X_{3-k}(m^*+1)\big) \end{aligned} \quad (36)$$

where the elements of X_k are all distinct, the elements of X_{3-k} must be the same, and $i_1 \neq i_2$. Thus from (V_i, V_j), we can first decode and remove $X_{3-k}(m^*), X_{3-k}(m^*+1)$, leaving us with $(X_k(m_1); X_k(m_1+1); X_k(m_2); X_k(m_2+1)) = \mathbf{T}_{ij}^{4\times 4} W_k$ and $\det(\mathbf{T}_{ij})$ is a non-zero polynomial $T_{ij}(\mathbf{H}_1, \mathbf{H}_2)$.

Finally, consider all edges of $G(\mathcal{V}, \mathcal{E})$ and consider $\prod_{i,j:\{V_i,V_j\}\in\mathcal{E}} T_{ij}(\mathbf{H}_1, \mathbf{H}_2)$, which is a polynomial with degree at most $4|\mathcal{E}|$. Now each element of $\mathbf{H}_1, \mathbf{H}_2$ is selected independently and uniformly from $\mathbb{F}_p$, where $p > 4|\mathcal{E}|$. By the Schwartz–Zippel lemma [28], [29], [30], we have

$$\Pr\left(\prod_{i,j:\{V_i,V_j\}\in\mathcal{E}} T_{ij}(\mathbf{H}_1, \mathbf{H}_2) = 0\right) \leq \frac{4|\mathcal{E}|}{p} < 1. \quad (37)$$

Therefore there exists a realization of $\mathbf{H}_1, \mathbf{H}_2$ so that each $T_{ij}(\mathbf{H}_1, \mathbf{H}_2) \neq 0$ and each $\mathbf{T}_{ij}$ has full rank, i.e., W_k can be recovered from $\{V_i, V_j\}$ and all decoding constraints (2) are satisfied.

C. Proof of Theorem 3: Necessary Condition of $\mathcal{G}_{C=4/3}$ With $K = 2$

We show that $R = 4/3$ cannot be achieved if a graph G contains an internal W_k-edge $\{V_{i_1}, V_{i_P}\}$ and its residing W_{3-k}-path $(\{V_{i_1}, V_{i_2}\}, \cdots, \{V_{i_{P-1}}, V_{i_P}\})$ contains only 2-color nodes, $V_{i_1}, \cdots, V_{i_P}$, among which at most one is special (if there exists, suppose it is $V_{i_p}, 1 < p < P$).

To set up the proof by contradiction, let us assume that $R = \lim_{L_w\to\infty} L_w/L_v = 4/3$ is asymptotically achievable (the same proof works for the exact achievable case by replacing $o(L_w)$ with zero), i.e.,

$$L_v = (3L_w)/4 + o(L_w). \quad (38)$$

We show that a 2-color node V must contain at least $L_w/4$ bits (minimum amount of desired information) and at most $L_w/2$ bits (maximum amount of interference) of information about each of W_1 and W_2. This result is stated in the following lemma.

Lemma 1 (2-color Node): When $R = 4/3$, for any 2-color node V, we have

$$\begin{aligned} L_w/4 + o(L_w) &\leq H(V|W_k) \\ &\leq L_w/2 + o(L_w), \ \forall k \in \{1,2\}. \end{aligned} \quad (39)$$

Proof: As V is 2-color, we have a W_1-edge $\{V, V_{j_1}\}$ and a W_2-edge $\{V, V_{j_2}\}$. We prove (39) when $k = 1$ and the proof when $k = 2$ follows from symmetry.

Consider the W_1-edge $\{V, V_{j_1}\}$. Following the steps in (15) to (19), we have

$$H(V|W_1) \leq 2L_v - L_w \overset{(38)}{=} L_w/2 + o(L_w). \quad (40)$$

Consider the W_2-edge $\{V, V_{j_2}\}$. We have

$$\begin{aligned} L_w &\overset{(1)}{=} H(W_2|W_1) && (41)\\ &\leq H(W_2, V, V_{j_2}|W_1) && (42)\\ &\overset{(2)}{=} H(V, V_{j_2}|W_1) && (43)\\ &\leq H(V|W_1) + H(V_{j_2}) && (44)\\ &\leq H(V|W_1) + L_v && (45)\\ \Rightarrow \quad H(V|W_1) &\geq L_w - L_v \overset{(38)}{=} L_w/4 + o(L_w). && (46) \end{aligned}$$

■

Next, we tighten the result in Lemma 1 when the 2-color node is further normal. Specifically, a normal 2-color node V must contain exactly $L_w/2$ bits of information about each of W_1 and W_2. This result is stated in the following lemma.

Lemma 2 (Normal 2-Color Node): When $R = 4/3$, for any normal 2-color node V, we have

$$H(V|W_k) = L_w/2 + o(L_w), \ \forall k \in \{1,2\}. \quad (47)$$

Proof: As V is normal 2-color, it must be connected to V_{j_1} through a W_1-edge and V_{j_2} through a W_2-edge, and further V_{j_1}, V_{j_2} are 2-color. The '$\leq$' direction of (47) has been proved in (39), so we only need to prove the '$\geq$' direction, which is considered in the following when $k = 2$ and the proof when $k = 1$ follows from symmetry.

Consider the W_1-edge $\{V, V_{j_1}\}$. We have

$$\begin{aligned} L_w &\overset{(1)}{=} H(W_1|W_2) && (48)\\ &\leq H(W_1, V, V_{j_1}|W_2) && (49)\\ &\overset{(2)}{=} H(V, V_{j_1}|W_2) && (50)\\ &\leq H(V|W_2) + H(V_{j_1}|W_2) && (51)\\ &\overset{(39)}{\leq} H(V|W_2) + L_w/2 + o(L_w) && (52)\\ \Rightarrow \quad H(V|W_2) &\geq L_w/2 + o(L_w) && (53) \end{aligned}$$

where (52) holds because V_{j_1} is a 2-color node so that we may apply (39) of Lemma 1.

■

After establishing the properties on the nodes, we proceed to consider the edges. We show that for any two connected 2-color nodes, the interference contained in them is $L_w/2$ bits if the two nodes contain one normal 2-color node.

Lemma 3 (W_{3-k}-edge): When $R = 4/3$, for any W_{3-k}-edge $\{V_i, V_j\}$ where V_i, V_j are 2-color and at least one of V_i, V_j is normal, we have

$$H(V_i, V_j|W_{3-k}) = L_w/2 + o(L_w). \tag{54}$$

Proof: Suppose V_i is normal, then on the one hand, we have

$$H(V_i, V_j|W_{3-k}) \geq H(V_i|W_{3-k}) \overset{(47)}{=} L_w/2 + o(L_w). \tag{55}$$

On the other hand, we have

$$H(V_i, V_j|W_{3-k}) \overset{(19)}{\leq} 2L_v - L_w \overset{(38)}{=} L_w/2 + o(L_w) \tag{56}$$

so that the proof is complete. ∎

We now go from the properties of edges in Lemma 3 to those of paths that were made up of such edges. We show that the interference contained in a sequence of such edges, i.e., a path, is $L_w/2$ bits, in the following lemma.

Lemma 4 (W_{3-k}-path): When $R = 4/3$, for a W_{3-k}-path $(\{V_{i_1}, V_{i_2}\}, \cdots, \{V_{i_{P-1}}, V_{i_P}\})$ where $V_{i_1}, \cdots, V_{i_{p-1}}, V_{i_{p+1}}, \cdots, V_{i_P}, 1 < p < P$ are normal and V_{i_p} is either normal or special, we have

$$H(V_{i_1}, \cdots, V_{i_p}|W_{3-k}) \leq L_w/2 + o(L_w) \tag{57}$$
$$H(V_{i_p}, \cdots, V_{i_P}|W_{3-k}) \leq L_w/2 + o(L_w). \tag{58}$$

Proof: We prove (57) and (58) follows similarly. The proof is based on a straightforward application of the submodular property on the entropy function to (54) in Lemma 3 (note that each edge in the path contains at most one special 2-color node).

$$\begin{aligned} & H(V_{i_1}, V_{i_2}|W_{3-k}) + \cdots + H(V_{i_{p-1}}, V_{i_p}|W_{3-k}) \\ \geq\ & H(V_{i_1}, \cdots, V_{i_p}|W_{3-k}) + H(V_{i_2}|W_{3-k}) \\ & + \cdots + H(V_{i_{p-1}}|W_{3-k}) \qquad (59) \\ \overset{(47)(54)}{\Longrightarrow}\ & (p-1)L_w/2 \geq H(V_{i_1}, \cdots, V_{i_p}|W_{3-k}) \\ & + (p-2)L_w/2 + o(L_w) \qquad (60) \\ \Longrightarrow\ & H(V_{i_1}, \cdots, V_{i_p}|W_{3-k}) \leq L_w/2 + o(L_w). \qquad (61) \end{aligned}$$

∎

Equipped with the above lemmas, we are ready to demonstrate a contradiction as follows.

$$\begin{aligned} L_w &= L_w/2 + L_w/2 \qquad (62) \\ &\overset{(57)(58)}{\geq} H(V_{i_1}, \cdots, V_{i_p}|W_{3-k}) \\ &\quad + H(V_{i_p}, \cdots, V_{i_P}|W_{3-k}) + o(L_w) \qquad (63) \\ &\geq H(V_{i_1}, \cdots, V_{i_P}|W_{3-k}) \\ &\quad + H(V_{i_p}|W_{3-k}) + o(L_w) \qquad (64) \\ &\overset{(2)(39)}{\geq} H(V_{i_1}, V_{i_P}, W_k|W_{3-k}) \\ &\quad + L_w/4 + o(L_w) \qquad (65) \\ &\geq H(W_k|W_{3-k}) + L_w/4 + o(L_w) \qquad (66) \\ &\overset{(1)}{=} 5L_w/4 + o(L_w) \qquad (67) \\ \Rightarrow\ 1 &\geq 5/4 \ \text{(contradiction)} \qquad (68) \end{aligned}$$

where (64) follows from submodularity; the first term of (65) follows from the decoding constraint (2) of the W_k-edge $\{V_{i_1}, V_{i_P}\}$ and the second term of (65) follows by applying Lemma 1 to the 2-color node V_{i_p}; the last step follows by dividing by L_w on both hand sides and letting $L_w \to \infty$.

D. Proof of Theorem 4: One Internal Edge

We show that $R = 4/3$ is achievable if a graph $G(\mathcal{V}, \mathcal{E})$ contains only one internal edge, say W_k-edge $\{V_i, V_j\}$ and each residing path has at least two special 2-color nodes. We first present the code construction and then prove it satisfies the decoding constraint (2).

1) Code Construction: The first part of the code construction is the same as that in Section IV-B.1. The second part is presented now, where we need to make the following updates. Generate two more generic linear combinations of W_k symbols.

$$\overline{X}_k = (\overline{X}_k(1); \overline{X}_k(2)) \triangleq \overline{\mathbf{H}}_k W_k \in \mathbb{F}_p^{2\times 1} \tag{69}$$

where each element of $\overline{\mathbf{H}}_k \in \mathbb{F}_p^{2\times 4}$ is independent and uniform over $\mathbb{F}_p$.

Consider the internal W_k-edge $\{V_i, V_j\}$ and find its all residing W_{3-k}-paths whose nodes are all 2-color (i.e., no 1-color nodes). Suppose there are M such paths, denoted by $P_1, \cdots, P_M$. Start from $V_i \backslash V_j$ and visit each path $P_m, m \in [M]$ along the W_{3-k}-edges until we see a special 2-color node, denoted by $V_{i_m} \backslash V_{j_m}$. Denote the set of $V_{i_m} \backslash V_{j_m}$ nodes as $\mathcal{V}_i \backslash \mathcal{V}_j$. Note that every node in $\mathcal{V}_i, \mathcal{V}_j$ is W_k-special and $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ (as each residing path has at least two special 2-color nodes).

V_i, V_j are normal and suppose they are currently set as (by the construction in Section IV-B.1)

$$\begin{aligned} V_i = \big(& X_k(m^*) + (2j_1 - 1)X_k(m^* + 1), \\ & X_{3-k}(m_1) + (2i_1 - 1)X_{3-k}(m_1 + 1), \\ & X_k(m^*) + 2j_1 X_k(m^* + 1) \\ & + X_{3-k}(m_1) + 2i_1 X_{3-k}(m_1 + 1) \\ V_j = \big(& X_k(m^*) + (2j_2 - 1)X_k(m^* + 1), \\ & X_{3-k}(m_1) + (2i_2 - 1)X_{3-k}(m_1 + 1), \\ & X_k(m^*) + 2j_2 X_k(m^* + 1) \\ & + X_{3-k}(m_1) + 2i_2 X_{3-k}(m_1 + 1)\big) \end{aligned} \tag{70}$$

where because the W_k-edge $\{V_i, V_j\}$ is internal, the desired symbols are limited to $X_k(m^*), X_k(m^* + 1)$. Then each W_k-special 2-color node in $\mathcal{V}_i, \mathcal{V}_j$ is currently set as

$$\begin{aligned} V \in \mathcal{V}_i \cup \mathcal{V}_j: \\ V = \big(& X_k(m^*) + (2j_3 - 1)X_k(m^* + 1), \\ & X_{3-k}(m_2) + (2i_3 - 1)X_{3-k}(m_2 + 1), \\ & X_{3-k}(m_2) + 2i_3 X_{3-k}(m_2 + 1)\big) \end{aligned} \tag{71}$$

and update it to

$$\begin{aligned} \text{if } V \in \mathcal{V}_i: V = \big(& X_k(m^*), \\ & X_{3-k}(m_2) + (2i_3 - 1)X_{3-k}(m_2 + 1), \\ & X_{3-k}(m_2) + 2i_3 X_{3-k}(m_2 + 1)\big) \end{aligned} \tag{72}$$

$$\begin{aligned} \text{if } V \in \mathcal{V}_j: V = \big(& X_k(m^* + 1), \\ & X_{3-k}(m_2) + (2i_3 - 1)X_{3-k}(m_2 + 1), \\ & X_{3-k}(m_2) + 2i_3 X_{3-k}(m_2 + 1)\big). \end{aligned} \tag{73}$$

For every normal 2-color node V in the segment of residing path $P_m, m \in [M]$ from V_i to the node before V_{i_m}. Update V as follows.

$$\begin{aligned} V &= \big(X_k(m^*) + (2j_4 - 1)X_k(m^* + 1), \\ &\quad X_{3-k}(m_3) + (2i_4 - 1)X_{3-k}(m_3 + 1), \\ &\quad X_k(m^*) + 2j_4 X_k(m^* + 1) \\ &\quad + X_{3-k}(m_3) + 2i_4 X_{3-k}(m_3 + 1)\big) \qquad (74) \\ \rightarrow\ V &= \big(X_k(m^*) + (2j_4 - 1)\overline{X}_k(1), \\ &\quad X_{3-k}(m_3) + (2i_4 - 1)X_{3-k}(m_3 + 1), \\ &\quad X_k(m^*) + 2j_4 \overline{X}_k(1) \\ &\quad + X_{3-k}(m_3) + 2i_4 X_{3-k}(m_3 + 1)\big). \qquad (75) \end{aligned}$$

Similarly replace $X_k(m^* + 1)$ by $\overline{X}_k(1)$ for all nodes (except V_{i_m}) that are connected to the above V through a W_{3-k}-path.

For every normal 2-color node V in the segment of residing path $P_m, m \in [M]$ from V_j to the node before V_{j_m}. Update V as follows.

$$\begin{aligned} V &= \big(X_k(m^*) + (2j_5 - 1)X_k(m^* + 1), \\ &\quad X_{3-k}(m_4) + (2i_5 - 1)X_{3-k}(m_4 + 1), \\ &\quad X_k(m^*) + 2j_5 X_k(m^* + 1) \\ &\quad + X_{3-k}(m_4) + 2i_5 X_{3-k}(m_4 + 1)\big) \qquad (76) \\ \rightarrow\ V &= \big(\overline{X}_k(2) + (2j_5 - 1)X_k(m^* + 1), \\ &\quad X_{3-k}(m_4) + (2i_5 - 1)X_{3-k}(m_4 + 1), \\ &\quad \overline{X}_k(2) + 2j_5 X_k(m^* + 1) \\ &\quad + X_{3-k}(m_4) + 2i_5 X_{3-k}(m_4 + 1)\big). \qquad (77) \end{aligned}$$

Similarly replace $X_k(m^*)$ by $\overline{X}_k(2)$ for all nodes (except V_{j_m}) that are connected to the above V through a W_{3-k}-path. The description of the code construction is complete.

2) Proof of Correctness: The proof of correctness is similar to that in Section IV-B.2, where we wish to show that for each edge $\{V_i, V_j\}$, the interference dimension is limited to two so that interference can be decoded and removed and the linear mapping from the four linear combinations of desired source symbols to the four desired source symbols, described by a 4×4 matrix $\mathbf{T}_{ij}$ may have full rank and then the existence of a feasible code construction (i.e., a choice of $\mathbf{H}_1, \mathbf{H}_2, \overline{\mathbf{H}}_k$) is guaranteed by the Schwartz–Zippel lemma [28], [29], [30] (refer to (37)).

We now consider each edge of G. The unchanged edges are the same as before and the proof in Section IV-B.2 applies. We are left with the edges that have been updated. First, for the only internal W_k-edge $\{V_i, V_j\}$, the interference is unchanged, i.e., limited to $X_{3-k}(m_1), X_{3-k}(m_1 + 1)$ and the desired symbols are $X_k(m^*), X_k(m^* + 1), \overline{X}_k(1), \overline{X}_k(2)$ so that $\det(\mathbf{T}_{ij})$ is not the zero polynomial. Second, for every W_{3-k}-edge in the segment of residing path $P_m, m \in [M]$ from V_i to V_{i_m}, the desired X_{3-k} symbols are unchanged and the interference from W_k is limited to $X_k(m^*), \overline{X}_k(1)$, i.e., two dimensions (refer to (72), (75)) so that interference can be decoded and removed. Third, for every W_{3-k}-edge in the segment of residing path $P_m, m \in [M]$ from V_j to V_{j_m}, the desired X_{3-k} symbols are unchanged and the interference from W_k is limited to $X_k(m^*+1), \overline{X}_k(2)$, i.e., two dimensions (refer to (73), (77)). Finally, for all other edges that involve a node that has been updated, no matter it is a W_3-edge or a W_{3-k}-edge, we may verify that interference has dimension two and desired symbols have dimension four. The proof of correctness is thus complete.

E. Proof of Theorem 5: Graph G in Fig. 9

We show that $R < 4/3$ for the graph G in Fig. 9. To set up the proof by contradiction, let us assume that $R = \lim_{L_w \to \infty} L_w/L_v = 4/3$ is (asymptotically) achievable, i.e., $L_v = 3L_w/4 + o(L_w)$.

Let us start with a useful inequality, stated in the following lemma.

Lemma 5: When $R = 4/3$, for the graph G in Fig. 9, we have

$$H(V_5, V_6|W_1) \geq L_w/2 + o(L_w). \qquad (78)$$

Proof:

$$\begin{aligned} H(V_5, V_6|W_1) &= H(V_1, V_2, V_5, V_6|W_1) \\ &\quad - H(V_1, V_2|V_5, V_6, W_1) \qquad (79) \\ &\overset{(2)}{\geq} H(V_1, V_2, W_2|W_1) - H(V_1|V_5, W_1) \\ &\quad - H(V_2|V_6, W_1) \qquad (80) \\ &\overset{(1)}{\geq} L_w - H(V_1, V_5|W_1) + H(V_5|W_1) \\ &\quad - H(V_2, V_6|W_1) + H(V_6|W_1) \qquad (81) \\ &\overset{(39)(54)}{\geq} L_w - L_w/2 + L_w/4 \\ &\quad - L_w/2 + L_w/4 + o(L_w) \qquad (82) \\ &= L_w/2 + o(L_w) \qquad (83) \end{aligned}$$

where the first term of (80) follows from the decoding constraint (2) of the W_2-edge $\{V_1, V_2\}$; (82) follows by applying Lemma 3 to edges $\{V_1, V_5\}, \{V_2, V_6\}$ and applying Lemma 1 to nodes V_5, V_6. ■

Next, applying Lemma 3 to edges $\{V_3, V_5\}$, $\{V_3, V_6\}$, $\{V_8, V_5\}$, $\{V_8, V_6\}, \{V_8, V_7\}, \{V_7, V_4\}$ and submodularity repeatedly, we have

$$\begin{aligned} & 3L_w + o(L_w) \\ \overset{(54)}{=}\ & H(V_3, V_5|W_1) + H(V_3, V_6|W_1) \\ & + H(V_8, V_5|W_1) + H(V_8, V_6|W_1) \\ & + H(V_8, V_7|W_1) + H(V_7, V_4|W_1) \qquad (84) \\ \geq\ & H(V_3, V_5, V_6|W_1) + H(V_3|W_1) \\ & + H(V_8, V_5, V_6|W_1) + H(V_8|W_1) \\ & + H(V_8, V_7, V_4|W_1) + H(W_7|W_1) \qquad (85) \\ \overset{(39)(47)}{\geq}\ & H(V_3, V_4, V_5, V_6, V_7, V_8|W_1) + H(V_5, V_6|W_1) \\ & + H(V_8|W_1) + 5L_w/4 \qquad (86) \\ \overset{(2)(47)(78)}{\geq}\ & H(V_3, V_4, W_2|W_1) \\ & + L_w/2 + L_w/2 + 5L_w/4 \qquad (87) \\ \overset{(1)}{\geq}\ & 13L_w/4 \qquad (88) \\ \Rightarrow\ & 3 \geq 13/4 \text{ (contradiction)} \qquad (89) \end{aligned}$$

where (86) follows by applying Lemma 1 to 2-color node V_7 and applying Lemma 2 to normal 2-color nodes V_3, V_8 and the first term of (87) follows from the decoding constraint (2) of the W_2-edge $\{V_3, V_4\}$. We have arrived at a contradiction and the proof is complete.

F. Proof of Theorem 6: Necessary Condition of $\mathcal{G}_{C=4/3}$ With $K > 2$

We show that $R < 4/3$ if a graph G contains any one of the three structures in Theorem 6. Let us consider the three structures sequentially.

The first structure is that G contains an M-color node, where $M \geq 4$. From (12), we have $R \leq (M+1)/M \leq 5/4 < 4/3$.

The second structure is that G contains a 3-color node V that is connected to an M-color node V_{i_1}, where $M \geq 2$. Suppose $\{V, V_{i_1}\}$ is a W_{k_1}-edge. As V is 3-color, we have a W_{k_2}-edge $\{V, V_{i_2}\}$ and a W_{k_3}-edge $\{V, V_{i_3}\}$, where k_1, k_2, k_3 are distinct and i_1, i_2, i_3 are distinct. As V_{i_1} is M-color, $M \geq 2$, we have a W_k-edge $\{V_{i_1}, V_j\}$ where j might be i_2 or i_3 (but $j \neq i_1$) and k might be k_2 or k_3 (but $k \neq k_1$). The following proof will work under all circumstances.

Consider W_{k_2}-edge $\{V, V_{i_2}\}$ and W_{k_3}-edge $\{V, V_{i_3}\}$. From the decoding constraint (2), we have

$$\begin{aligned} 2L_w &\overset{(1)(2)}{=} I(V, V_{i_2}, V_{i_3}; W_{k_2}, W_{k_3}) && (90)\\ &= H(V, V_{i_2}, V_{i_3}) - H(V, V_{i_2}, V_{i_3}|W_{k_2}, W_{k_3}) && (91)\\ &\leq 3L_v - H(V|W_{k_2}, W_{k_3}) && (92)\\ \Rightarrow\ & H(V|W_{k_2}, W_{k_3}) \leq 3L_v - 2L_w. && (93)\end{aligned}$$

Consider W_k-edge $\{V_{i_1}, V_j\}$. From (19), we have

$$H(V_{i_1}|W_k) \leq 2L_v - L_w. \tag{94}$$

Adding (93) and (94), we have

$$\begin{aligned} 5L_v - 3L_w &\geq H(V|W_{k_2}, W_{k_3}) + H(V_{i_1}|W_k) && (95)\\ &\geq H(V, V_{i_1}|W_{k_2}, W_{k_3}, W_k) && (96)\\ &\overset{(2)}{\geq} H(W_{k_1}|W_{k_2}, W_{k_3}, W_k) && (97)\\ &\overset{(1)}{=} L_w && (98)\\ \Rightarrow\ R = L_w/L_v &\leq 5/4 < 4/3. && (99)\end{aligned}$$

The third structure is that G contains a normal 2-color node V that is connected to a 2-color node V_i and V, V_i are connected to different types of edges. Suppose $\{V, V_i\}$ is a W_k-edge. As V, V_i are 2-color (the two colors are different) and V is normal, we have a W_{k_1}-edge $\{V_i, V_{i_1}\}$, a W_{k_2}-edge $\{V, V_{j_1}\}$, and a W_{k_3}-edge $\{V_{j_1}, V_{j_2}\}$ where k, k_1, k_2 are distinct and $k_3 \neq k_2$.

Consider W_k-edge $\{V, V_i\}$ and W_{k_1}-edge $\{V_i, V_{i_1}\}$. Following the derivation of (93), we have

$$H(V|W_k, W_{k_1}) \leq 3L_v - 2L_w. \tag{100}$$

Consider W_{k_3}-edge $\{V_{j_1}, V_{j_2}\}$. From (19), we have

$$H(V_{j_1}|W_{k_3}) \leq 2L_v - L_w. \tag{101}$$

Adding (100) and (101), we have

$$\begin{aligned} 5L_v - 3L_w &\geq H(V|W_k, W_{k_1}) + H(V_{j_1}|W_{k_3}) && (102)\\ &\geq H(V, V_{j_1}|W_k, W_{k_1}, W_{k_3}) && (103)\\ &\overset{(2)}{\geq} H(W_{k_2}|W_k, W_{k_1}, W_{k_3}) && (104)\\ &\overset{(1)}{=} L_w && (105)\\ \Rightarrow R &= L_w/L_v \leq 5/4 < 4/3. && (106)\end{aligned}$$

G. Proof of Theorem 7: Sufficient Condition of $\mathcal{G}_{C=4/3}$ With $K > 2$

We show that $R = 4/3$ is achievable if a graph $G(\mathcal{V}, \mathcal{E}) \notin \mathcal{G}^{\text{Thm 6}}_{C<3/4}$ contains no internal edge after removing one 1-color nodes in residing paths. We first present the code construction, which is a minor modification of that in Section IV-B.1 and then prove it satisfies the decoding constraint (2), which is similar to that in Section IV-B.2.

1) Code Construction: Choose the field size p to be a prime that is greater than $4|\mathcal{E}|$. Set $L_w = 4\log_2 p, L_v = 3\log_2 p$ so that each source\coded symbol is comprised of $4\backslash 3$ symbols from $\mathbb{F}_p$ and the rate achieved is $4/3$.

Consider the set of nodes that are connected to W_k-edges, $k \in [K]$ and denote this set by $\mathcal{V}_k$. Consider the subgraph of $G(\mathcal{V}, \mathcal{E})$ whose node set is $\mathcal{V}_k$ and edge set is comprised of all edges that are not W_k-edges and are connected to some node in $\mathcal{V}_k$, denoted by $\mathcal{E}_{k^c}$ and denote this subgraph by $G_k(\mathcal{V}_k, \mathcal{E}_{k^c})$. Decompose $G_k(\mathcal{V}_k, \mathcal{E}_{k^c})$ into $W_{k'}$-components, $k' \neq k$ and suppose we have M_k such components. A trivial component with a single node can be classified as a $W_{k'}$-component for any $k' \neq k$ and we just fix one k' (any choice will work). Among these M_k $W_{k'}$-components, suppose M_k^1 components are comprised of 1-color nodes and label them as $P_k^{[1]}, \cdots, P_k^{[M_k^1]}$; M_k^2 components are comprised of 2-color nodes and label them as $Q_k^{[1]}, \cdots, Q_k^{[M_k^2]}$; the remaining $M_k^3 = M_k - M_k^1 - M_k^2$ components are comprised of 3-color nodes and label them as $S_k^{[1]}, \cdots, S_k^{[M_k^3]}$ (each such component is an isolated node as 3-color nodes are connected to only 1-color nodes when $G \notin \mathcal{G}^{\text{Thm 6}}_{C<3/4}$).

Generate generic linear combinations of the source symbols as follows.

$$\begin{aligned} W_k &= (W_k(1); \cdots; W_k(4)) \in \mathbb{F}_p^{4\times 1}, k \in [K] && (107)\\ X_k &= (X_k(1); \cdots; X_k(3M_k^1 + 2M_k^2 + M_k^3))\\ &\triangleq \mathbf{H}_k W_k \in \mathbb{F}_p^{(3M_k^1+2M_k^2+M_k^3)\times 1} && (108)\end{aligned}$$

where each element of $\mathbf{H}_k \in \mathbb{F}_p^{(3M_k^1+2M_k^2+M_k^3)\times 4}$ is chosen uniformly and independently from $\mathbb{F}_p$.

Consider $P_k^{[m]}, m \in [M_k^1]$, denote its node by V, and set

$$V = \big(X_k(3m-2), X_k(3m-1), X_k(3m)\big). \tag{109}$$

This step completes the assignment for all 1-color nodes.

Consider $Q_k^{[m]}, m \in [M_k^2]$. Suppose $Q_k^{[m]}$ contains J nodes $V_{i_1}, \cdots, V_{i_J}$. Consider $V_{i_j}, j \in [J]$, which is 2-color.

$$\begin{aligned} &\text{If } V_{i_j} \text{ is } W_k\text{-special, set } V_{i_j}^{[k]} \triangleq\\ &X_k\big(3M_k^1 + 2m - 1\big) + (2j-1)X_k\big(3M_k^1 + 2m\big); \end{aligned} \tag{110}$$

otherwise, set $V_{i_j}^{[k]} = (V_{i_j}^{[k]}(1), V_{i_j}^{[k]}(2))$

$$\triangleq \Big(X_k\big(3M_k^1 + 2m - 1\big) + (2j-1)X_k\big(3M_k^1 + 2m\big), \\ X_k\big(3M_k^1 + 2m - 1\big) + 2jX_k\big(3M_k^1 + 2m\big)\Big). \quad (111)$$

For any 2-color node V that is connected to W_{k_1}-edges and W_{k_2}-edges,

if V is normal, then set

$$V = (V^{[k_1]}(1), V^{[k_2]}(1), V^{[k_1]}(2) + V^{[k_2]}(2)); \quad (112)$$

otherwise, set $V = (V^{[k_1]}, V^{[k_2]})$. (113)

This step completes the assignment for all 2-color nodes.

Consider $S_k^{[m]}, m \in [M_k^3]$, denote its node by V, and set

$$V^{[k]} \triangleq X_k(3M_k^1 + 2M_k^2 + m). \quad (114)$$

For any 3-color node V that is connected to W_{k_1}-edges, W_{k_2}-edges, and W_{k_3}-edges, set

$$V = (V^{[k_1]}, V^{[k_2]}, V^{[k_3]}). \quad (115)$$

This step completes the assignment for all 3-color nodes and as $G \notin \mathcal{G}^{\text{Thm 6}}_{C<3/4}$ contains no M-color nodes, where $M \geq 4$, the code construction is complete.

2) Proof of Correctness: Consider any edge $\{V_i, V_j\} \in \mathcal{E}$ and suppose it is a W_k-edge.

When V_i, V_j contain one 1-color code, then our assignment ensures that (V_i, V_j) contains four distinct elements of X_k, which can be written as $\mathbf{T}_{ij}^{4\times 4} W_k$ and $\det(\mathbf{T}_{ij})$ is a non-zero polynomial $T_{ij}(\mathbf{H}_1, \cdots, \mathbf{H}_K)$.

We are left with cases where V_i, V_j are both 2-color (because 3-color nodes in $G \notin \mathcal{G}^{\text{Thm 6}}_{C<3/4}$ are connected only to 1-color nodes). When V_i, V_j contain one normal 2-color code, then $G \notin \mathcal{G}^{\text{Thm 6}}_{C<3/4}$ ensures that V_i, V_j are connected to edges that are associated with the same set of two source symbols, i.e., we are back to the $K = 2$ setting considered in Theorem 2 and following the proof in Case 2 and Case 3 of Section IV-B.2, we have $T_{ij}(\mathbf{H}_1, \cdots, \mathbf{H}_K)$ is non-zero. The only remaining case is that V_i, V_j are both special, say V_i is W_{k_1}-special and V_j is W_{k_2}-special, where $k_1 \neq k, k_2 \neq k$. Then from (111) and (113), we know that (V_i, V_j) each contains two distinct elements of X_k (distinctness is due to the absence of internal edges after removing 1-color nodes) so that $T_{ij}(\mathbf{H}_1, \cdots, \mathbf{H}_K)$ is not the zero-polynomial.

Finally, consider $\prod_{i,j:\{V_i,V_j\}\in\mathcal{E}} T_{ij}(\mathbf{H}_1, \cdots, \mathbf{H}_K)$, which is a non-zero polynomial with degree at most $4|\mathcal{E}| < p$, the field size. By the Schwartz–Zippel lemma [28], [29], [30], there exists a realization of $\mathbf{H}_1, \cdots, \mathbf{H}_K$ so that each $T_{ij}(\mathbf{H}_1, \cdots, \mathbf{H}_K) \neq 0$ and all decoding constraints (2) are satisfied.

V. Discussion

An extremal rate perspective is taken to study the storage code problem over graphs. For the highest capacity values, we have identified a number of combinatorial structures that have significant impact on the code rate - M-color code (i.e., the number of sources associated with a node), internal edge (which captures a direct conflict between alignment of undesired source symbols and independence of desired source symbols), normal 2-color node\special 2-color node (for rate $4/3$, which keeps the same interference\which could change interference up to the extent of $1/4$ source size). Both the achievability and converse results are guided by a linear dimension counting view. The sufficient and necessary conditions presented are not the largest that our proof technique can lead to, i.e., we can solve more graph instances, but a systematic description is still out of current reach. It is not clear which rates will turn out to have hard capacity instances. Specifically, all extremal graphs with storage code capacity $4/3$ appear to go beyond the techniques of this work. Regarding generalizations, we note that our model is the most elementary, where we have focused on the highest capacity values, i.e., best rate scenarios instead of lowest capacity values, i.e., worst rate scenarios, or other physically meaningful rates; decoding constraints are placed on a pair of nodes in this work instead of an arbitrary set of nodes, i.e., we may have a hypergraph rather than a graph [2]; each edge is associated with only one source symbols instead of multiple source symbols where the decoding structure can be more diverse [1]. Finally, from an extremal rate and network perspective, we may view combinatorial objects using the metric of capacity and study further extremal (largest, densest, most (linearly) independent) graphs, set families, vector spaces etc. along the line of extremal combinatorics [31].

References

[1] Z. Li and H. Sun, "On extremal rates of secure storage over graphs," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 4721–4731, 2023.

[2] S. Sahraei and M. Gastpar, "GDSP: A graphical perspective on the distributed storage systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2218–2222.

[3] R. W. Yeung, *Information Theory and Network Coding*. Cham, Switzerland: Springer, 2008.

[4] C. K. Ngai and R. W. Yeung, "Network coding gain of combination networks," in *Proc. Inf. Theory Workshop*, 2004, pp. 283–287.

[5] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin, "Protecting data privacy in private information retrieval schemes," in *Proc. 13th Annu. ACM Symp. Theory Comput. STOC*, 1998, pp. 151–160.

[6] Z. Li and H. Sun, "Conditional disclosure of secrets: A noise and signal alignment approach," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4052–4062, Jun. 2022.

[7] Z. Li and H. Sun, "On the linear capacity of conditional disclosure of secrets," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 3202–3207.

[8] Z. Wang and S. Ulukus, "Communication cost of two-database symmetric private information retrieval: A conditional disclosure of multiple secrets perspective," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 402–407.

[9] E. F. Brickell and D. M. Davenport, "On the classification of ideal secret sharing schemes," *J. Cryptol.*, vol. 4, no. 2, pp. 123–134, Jan. 1991.

[10] H.-M. Sun and S.-P. Shieh, "Secret sharing in graph-based prohibited structures," in *Proc. INFOCOM*, vol. 2, Apr. 1997, pp. 718–724.

[11] R. Dougherty, C. Freiling, and K. Zeger, "Network coding and matroid theory," *Proc. IEEE*, vol. 99, no. 3, pp. 388–405, Mar. 2011.

[12] S. Kamath, V. Anantharam, D. Tse, and C.-C. Wang, "The two-unicast problem," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3865–3882, May 2018.

[13] H. Sun and S. A. Jafar, "Index coding capacity: How far can one go with only Shannon inequalities?" *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3041–3055, Jun. 2015.

[14] M. Blaum, J. Brady, J. Bruck, and J. Menon, "EVENODD: An efficient scheme for tolerating double disk failures in RAID architectures," *IEEE Trans. Comput.*, vol. 44, no. 2, pp. 192–202, Apr. 1995.

[15] M. Blaum, J. Bruck, and A. Vardy, "MDS array codes with independent parity symbols," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 529–542, Mar. 1996.

[16] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.
[17] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, "A survey on network codes for distributed storage," *Proc. IEEE*, vol. 99, no. 3, pp. 476–489, Mar. 2011.
[18] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6925–6934, Nov. 2012.
[19] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5843–5855, Oct. 2014.
[20] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4661–4676, Aug. 2014.
[21] A. Mazumdar, "Storage capacity of repairable networks," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 5810–5821, Nov. 2015.
[22] A. Patra and A. Barg, "Node repair on connected graphs," *IEEE Trans. Inf. Theory*, vol. 68, no. 5, pp. 3081–3095, May 2022.
[23] A. Barg and G. Zémor, "High-rate storage codes on triangle-free graphs," *IEEE Trans. Inf. Theory*, vol. 68, no. 12, pp. 7787–7797, Dec. 2022.
[24] S. Basu and M. Mukherjee, "Optimal storage codes on graphs with fixed locality," 2023, *arXiv:2307.08680*.
[25] J. Katz and L. Trevisan, "On the efficiency of local decoding procedures for error-correcting codes," in *Proc. 32nd Annu. ACM Symp. Theory Comput.*, 2000, pp. 80–86.
[26] S. Yekhanin, "Locally decodable codes," *Found. Trends Theor. Comput. Sci.*, vol. 6, no. 3, pp. 139–255, 2011, doi: 10.1561/0400000030.
[27] H. Sun and S. A. Jafar, "On the capacity of locally decodable codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6566–6579, Oct. 2020.
[28] R. A. Demillo and R. J. Lipton, "A probabilistic remark on algebraic program testing," *Inf. Process. Lett.*, vol. 7, no. 4, pp. 193–195, Jun. 1978.
[29] J. T. Schwartz, "Fast probabilistic algorithms for verification of polynomial identities," *J. ACM*, vol. 27, no. 4, pp. 701–717, Oct. 1980.
[30] R. Zippel, "Probabilistic algorithms for sparse polynomials," in *Proc. Int. Symp. Symbolic Algebr. Manipulation*. Cham, Switzerland: Springer, 1979, pp. 216–226.
[31] S. Jukna, *Extremal Combinatorics: With Applications in Computer Science*, vol. 571. Cham, Switzerland: Springer, 2011.

Zhou Li (Member, IEEE) received the B.E. degree in communications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014. He is currently pursuing the Ph.D. degree with the University of North Texas. His research interests include information theory, network coding, security, and privacy.

Hua Sun (Member, IEEE) received the B.E. degree in communications engineering from the Beijing University of Posts and Telecommunications, China, in 2011, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California at Irvine, Irvine, CA, USA, in 2013 and 2017, respectively.

He is currently an Associate Professor with the Department of Electrical Engineering, University of North Texas, USA. His research interests include information theory and its applications to communications, privacy, security, and storage. He was a recipient of the NSF CAREER Award in 2021, the UNT College of Engineering Junior Faculty Research Award in 2021, and the UNT College of Engineering Distinguished Faculty Fellowship in 2023. His coauthored papers received the IEEE Jack Keil Wolf ISIT Student Paper Award in 2016, the IEEE GLOBECOM Best Paper Award in 2016, and the 2020–2021 IEEE Data Storage Best Student Paper Award.