Variable Packet-Error Coding

Xiaoqing Fan, Oliver Kosut, and Aaron B. Wagner

Abstract—We consider a problem in which a source is encoded into $N$ packets, an unknown number of which are subject to adversarial errors en route to the decoder. We seek code designs for which the decoder is guaranteed to be able to reproduce the source subject to a certain distortion constraint when there are no packets errors, subject to a less stringent distortion constraint when there is one error, etc. Focusing on the special case of the erasure distortion measure, we introduce a code design based on the polytope codes of Kosut, Tong, and Tse. The resulting designs are also applied to a separate problem in distributed storage.

I. INTRODUCTION

Consider a communication scenario in which a source sends information to a destination over several nonintersecting paths in a network. These paths could be used to increase the data rate beyond what would be achievable with a single path, or they could be used to provide redundancy to allow the decoder to recover from errors introduced by the network. It is also possible to simultaneously achieve both goals, subject to a tradeoff between the two, which is the topic of this paper. In particular, we shall assume that some number of paths are subject to adversarial errors, and we shall seek codes that achieve high data rates while still ensuring that the encoder can reconstruct the original message reasonably well in the face of those errors.

While coding for adversarial errors is a classical subject [24] [3], prior work in coding theory seeks to optimize only the worst-case performance of the code, that is, how well it performs when the number of errors introduced by the network is the maximum. For many real systems, however, this approach is overly pessimistic. Indeed, if the errors are due to an attack by an adversarial jammer, then the system may experience no errors at all in the typical case, since the network may only come under attack occasionally. We therefore desire a system that achieves some performance objective when the maximum number of errors are present while guaranteeing that a higher level of performance is achieved when there are fewer, or no, errors. This is not provided by the conventional approach to the problem, which is to use maximum distance separable (MDS) codes with a minimum distance that exceeds twice the maximum number of possible errors. For such codes the decoder can fully recover the source when the maximum number of errors occurs, but should no errors occur then the decoder is no better off than if they did.

We seek designs whose performance improves as the number of errors decreases. Since prior work has shown that source-channel separation is not optimal for this problem [1], it is properly formulated using rate-distortion theory. We assume that a source sequence in encoded into $N$ packets (or messages) at a given rate $R$, at most $T$ of which may be adversarially altered by the network. The decoder receives $N$ packets without knowing which packets were altered or how many have been altered (except that it knows that the total number of altered packets does not exceed $T$). The decoder then outputs a reconstruction of the source. We are given a distortion measure between the source and reproduction, and we seek codes that guarantee a certain level of distortion when there are $T$ errors, a lower level of distortion when there are $T-1$ errors, and so on.

In this paper we shall focus exclusively on the erasure distortion measure: the per-letter distortion is zero if the source and reconstruction symbols agree, one if the reconstruction symbol is a special “erasure” symbol, and infinity otherwise. Thus there is an infinite penalty for guessing a source symbol incorrectly, and the decoder should output the erasure symbol for any source symbol about which it is unsure. Assuming there are no errors in the reconstruction, the distortion of a string is then the fraction of erasures in the reconstruction. The erasure distortion measure is reasonable for a wide array of physical sources. For audio and video, it is typically possible to interpolate over unknown samples, pixels, or frames at the receiver. Similarly, humans can often recover a natural language source when some of the characters have been erased [4]. Even executable computer code, which is typically viewed as being unamenable to lossy compression, is suitable to compression under the erasure distortion measure: execution of the program at the decoder could simply pause whenever it reached an erasure and wait for further information, without ever executing incorrect instructions. Focusing on the erasure distortion measure is also a useful simplifying assumption when considering new problems, akin to the way that the binary erasure channel is a good starting point in the study of modern coding theory [20].

For this problem we provide a code construction that is inspired by the polytope codes introduced by Kosut, Tong, and Tse [15] in the context of network coding with adversarial nodes. Polytope codes are similar to linear maximum distance separable (MDS) codes but with an added feature: for a certain number of errors, which exceeds the decoding radius of the code, it is possible to always decode some of the codeword symbols even though it is not possible to decode all of them. This is to be contrasted with conventional MDS codes, for which in general none of the coded symbols can be decoded unless they all can. This “partial decodability” property will be crucial in our use of polytope codes. Our construction of polytope codes departs significantly from that of Kosut, Tong, and Tse, and is arguably more transparent. Nonetheless,
we shall still call them polytope codes to emphasize their connection to this earlier work.

The problem studied here can be viewed as an instance of a "large-alphabet" channel. In classical studies of channel capacity, the channel law is held fixed and the blocklength is permitted to grow without bound (e.g. [5]). In the case of discrete memoryless channels with finite alphabet, this model well captures the practical regime in which the blocklength is much bigger than the number of channel inputs or outputs. While this model has proven to be very successful, the asymptotic that it considers is not always the right one. For the problem in which a sender sends data over several independent paths in a network, some of which may alter the data adversarially en route, the "blocklength" is naturally viewed as the number of distinct paths, which is generally small, while the "alphabet" is the number of distinct messages that can be sent on one path, which is generally very large. Thus the appropriate model is in some sense dual to the classical one: the blocklength is fixed while the input and output alphabet sizes are permitted to grow without bound, as is done in this paper. Such channels have arisen in network coding [12], although many fundamental Shannon-theoretic questions about them are not well understood. One notable exception is that, as alluded to earlier, source-channel separation is known to be optimal for such channels if the source is Gaussian and the distortion measure is quadratic or if the source is binary and the distortion measure is erasure distortion [2]. Thus we already know that such channels behave differently from conventional ones. We call communication over such channels packet-error coding (or path-error coding) (PEC).

In this paper, we are interested in packet-error coding in which the number of packet errors is variable and a single code simultaneously provides different performance guarantees depending on the number of packet errors. We call this variable packet-error coding (VPEC). VPEC is closely related to the multiple descriptions (MD) problem [11] in network information theory. The difference is that in the MD problem each message is either received correctly or not received at all; the network does not introduce errors. The MD problem has received considerable attention [10], [11], [18] since it was introduced, including the special case in which the distortion measure is erasure [2]. Allowing the adversary to introduce errors instead of erasures seems to significantly alter the problem, however. In particular, although techniques from coding theory have been successfully applied to the MD problem [18], the polytope codes that shall prove so effective here do not appear to be useful for the MD problem.

Having developed the polytope code constructions for the VPEC problem, we subsequently apply essentially the same codes to the distributed storage system (DSS) problem in the presence of an active adversary. In a DSS, a file is stored across multiple storage nodes in a redundant fashion so as to recover from node failures. Beginning with Dimakis et al. [7], there has been considerable recent interest in applying techniques from network coding to the DSS problem. The problem has also been studied when several of the storage nodes are controlled by a malicious adversary [6], [17], [19], [16], [21].

Unlike the network coding problem originally studied for polytope codes [15], in which the network topologies can be arbitrary, the DSS problem yields highly constrained network topologies that are in fact similar to the one-hop network of the VPEC problem. That is, one is confronted with many data packets, some of which may be adversarially corrupted, and trustworthy packets must be identified. This similarity allows the use of the same polytope code constructions, and the partial decodability property will again be critical.

The rest of the paper is organized as follows. Section II describes the VPEC problem in detail and states the main theorem. Polytope codes are then defined in Section III and used to prove the main theorem in Section IV. We prove a partial optimality result for polytope codes in Section V. The DSS problem is described and our result stated in Section VI and our main theorem for the DSS problem is proved in Section VII.

II. PROBLEM FORMULATION AND RESULTS

A. Problem Formulation

Let $N$ be a positive integer and define $[N] = \{1, 2, \ldots, N\}$. Let $x^n$ denote the source message in $\mathcal{X}^n$, where $\mathcal{X} = [K]$ is the alphabet for the source. We will call $n$ the blocklength of the source. We do not assume that a probability distribution over $\mathcal{X}^n$ is given; all of our results will be worst-case over this space. Given the source sequence $x^n$, the encoder creates $N$ packets (or messages, or codewords) via the functions

$$f_\ell : \mathcal{X}^n \rightarrow \mathcal{X}^nR \quad \ell \in \{1, \ldots, N\}.$$ 

Note that we only consider the problem in which all of the packets have the same rate $R$. The encoder sends the packets

$$(f_1(x^n), f_2(x^n), \ldots, f_N(x^n)),$$

which we will often abbreviate as

$$(C_1, C_2, \ldots, C_N).$$

The decoder employs a function

$$g : \prod_{\ell=1}^{N} \mathcal{X}^nR \rightarrow \mathcal{X} \cup e \to \mathcal{X} \cup e^n$$

to reproduce the source given the received packets. The fidelity of the reproduction is measured using the erasure distortion measure [5] p. 338): for $x \in \mathcal{X}$ and $\hat{x} \in \mathcal{X} \cup e$, define

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } \hat{x} = e \\ \infty & \text{otherwise} \end{cases} \quad (1)$$

We extend the single-letter distortion measure $d(\cdot, \cdot)$ to strings in the usual way

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i).$$

1When the length of the vector is particularly important, we indicate it using a superscript.
We call the tuple \((f_1, \ldots, f_N, g)\) a code for the problem. We shall consider codes for which the source \(x^n\) can be perfectly reconstructed when all of the packets are received unaltered, i.e.,

\[
\max_{x^n \in X^n} d(x^n, g(C_\ell, \ell \in [N])) = 0.
\]

We call such codes feasible. For feasible codes, we shall consider how well the decoder can reproduce the source when at most \(T\) of the packets are received in error

\[
D_T(f_1, \ldots, f_N, g) := \max \left( \max_{x^n \in X^n} \max_{A \subseteq [N]: |A| \leq T} \max_{C_A} d(x^n, g(C_A, \tilde{C}_A)) \right).
\]

Here \(g(C_A, \tilde{C}_A)\) denotes the decoder’s output when its input is \(C_\ell = f_\ell(x^n)\) for all \(\ell \in A^c\) and \(\tilde{C}_\ell\) for all \(\ell \in A\).

Definition 1: The rate-distortion pair (R-D pair) \((R, D)\) is achievable if for all \(\epsilon > 0\), there exists a feasible code \((f_1, \ldots, f_N, g)\) for some blocklength with rate at most \(R + \epsilon\) such that

\[
D_T(f_1, \ldots, f_N, g) \leq D + \epsilon.
\]

B. Main Result

Our main result is the following.

Theorem 1: Suppose the maximum number of altered packets \(T\) satisfies \(T \geq 1\) and the number of packets \(N\) satisfies \(N \geq T + \left(\frac{1}{2}\right) + 2\).

1) If \(0 \leq R < \frac{1}{N-T}\), then there is no finite \(D\) for which \((R, D)\) is achievable.

2) Let \(F(T)\) denote \(T + \left(\frac{1}{4}\right) + 1\). Then for any \(\frac{1}{N-T} \leq R \leq \frac{1}{N-2T}\), the rate-distortion pair

\[
\left( R, \frac{F(T)(N-T)(1-(N-2T)R)}{NT} \right)
\]

is achievable.

The performance in part 2) is achieved using polytope codes and should be compared against what can be obtained using conventional MDS codes. Suppose we map \(N-2T\) source symbols to \(N\) coded symbols using an \((N, N-2T)\) MDS code (we can, if necessary, group several source symbols together to ensure that the source alphabet is large enough to guarantee the existence of such a code). Let each coded packet consist of exactly one of the coded symbols. The rate per packet is then \(R = 1/(N-2T)\), and since the minimum distance of the code is \(2T+1\), the decoder can always recover the source sequence exactly, even when there are \(T\) errors. Thus this scheme achieves the rate-distortion pair \((1/(N-2T), 0)\).

On the other hand, if we use an \((N, N-T)\) MDS code, then the decoder can reconstruct the source when there are no errors, and since the minimum distance is \(T + 1\), it can always detect when there are \(T\) or fewer errors and output the all-erasure string in response. Hence this code can achieve the rate-distortion pair \((1/(N-T), 1)\). A simple time-sharing argument shows that the line connecting these points

\[
\left( R, \frac{N-T-(N-T)(N-2T)R}{T} \right)
\]

is achievable. This is shown in Fig. 1 for \(N = 3\) and \(T = 1\) and in Fig. 2 for \(N = 5\) and \(T = 2\), along with the achievable rate-distortion pairs from Theorem 1. We see that Theorem 1 does strictly better.

When \(N = 3\) and \(T = 1\), there is actually a simple design that is not dominated by the above schemes. When \(R = \frac{2}{3}\), let the blocklength of the source message be three and write the source as \((x_1, x_2, x_3)\). We transmit

\[
(x_1, x_2) (x_2, x_3) (x_3, x_1)
\]

as the three packets. The decoder can check whether the copy of \(x_i\) is the same between the two packets in which it appears for each \(i\). If the two packets have the same value of \(x_i\), then this common value must be correct. Since the channel can alter at most one packet, there can be at most two components of \((x_1, x_2, x_3)\) on which there is disagreement. If there is disagreement about two source components, however, then the decoder can identify which packet was altered, exclude it, and then determine all of the source components from the remaining packets. Thus the maximum number of components about which the decoder can be uncertain is one. It follows that the R-D pair \((2/3, 1/3)\) is achievable. This point lies outside the region achieved by polytope codes, as shown in Fig. 1.

Since the rate-distortion pair \((1/(N-2T), 0)\) is achievable, and the set of achievable pairs is convex, to show part 2) of Theorem 1 it suffices to show that

\[
\left( \frac{1}{N-T} \frac{F(T)}{N} \right)
\]

is achievable. In the next section, we will show how polytope codes can be used toward this end. Note that, per the statement of Theorem 1, the resulting scheme can only be applied when \(N \geq F(T) + 1\). In particular, the blocklength must grow with the square of the number of errors. This is undesirable; one would prefer to have linear scaling. In Section 7 we show that this quadratic scaling cannot be improved by changing the decoder—it is intrinsic to the code itself. Of course, since \(N\) represents the number of independent paths in the network between the encoder and the decoder, we are generally interested in small values of \(N\) and \(T\), so that the scaling behavior is not paramount.

III. Polytope Codes

Polytope codes were introduced by Kosut, Tong, and Tse in the context of network coding with adversarial nodes. Polytope codes are akin to linear MDS codes, except that the arithmetic operations are performed over the reals and extra low rate “check” information is included in the transmission. Our construction is somewhat simpler than the one given in [15]. To understand this construction it is helpful to begin with the special case in which there are \(N = 3\) packets subject to at most \(T = 1\) error.
A. $N = 3$, $T = 1$ case

One trivial design for this case is to simply send the true source sequence in all three packets. Since there is at most one error, the decoder can always recover the source sequence by using a majority rule. That is, it can recover the source exactly when there are no errors but also when there is one. As such, this scheme achieves the rate-distortion pair $(1, 0)$.

This scheme is unsatisfactory, however, since it is wasteful when there no errors.

One may consider using a $(3, 2)$ MDS code instead. For instance, we could choose the blocklength $n = 2$ and encode two source symbols $x_1$ and $x_2$ into three packets as

$$x_1, x_2, x_1 \oplus x_2,$$  

(3)

where $\oplus$ denotes modulo arithmetic. The decoder can determine whether a single error has been introduced by verifying whether the received packets satisfy the linear relation in (3). If so, then there are no errors, and the decoder can reproduce the source exactly. Thus it is feasible. If not, then the decoder knows that one error is present, but it has no way of identifying which packet is in error. Since there is an infinite penalty for guessing a source symbol incorrectly, it must output the all-erasure string, achieving the rate-distortion pair $(1/2, 1)$. The striking thing about this example is that the decoder always receives at least one of the two source symbols correctly; the problem is that it does not know which of the two is correct.

Now suppose that the source is viewed as a pair of vectors of positive integers of length $N_0$, $x_1^{N_0}$ and $x_2^{N_0}$, and the three transmitted packets consist of

$$x_1^{N_0}, x_2^{N_0}, x_1^{N_0} + x_2^{N_0},$$

(4)

where now the addition is performed over the reals. We also send the quantities

$$\langle x_i^{N_0}, x_j^{N_0} \rangle$$

(5)

for all $i$ and $j$ as part of each packet. As before, the decoder can always detect whether an error has been introduced. If it detects no error, it can output the source sequence correctly. But now if it detects an error, it can always identify at least one of the three packets as correct, by the following reasoning. Since the inner products in (5) are included in all three packets, they can always be recovered correctly. Let

$$\tilde{x}_i^{N_0}, \tilde{x}_2^{N_0}, \tilde{x}_3^{N_0},$$

(6)

denote the vectors in the three received packets, and assume that exactly one of them has been altered. If for any $i$ we have

$$||\tilde{x}_i^{N_0}||^2 \neq ||x_i^{N_0}||^2,$$

then we know that the $i$th packet is in error and the other two must be correct. So we shall assume that

$$||\tilde{x}_i^{N_0}||^2 = ||x_i^{N_0}||^2,$$

for all $i$.

Now construct a graph with nodes $\tilde{x}_1^{N_0}, \tilde{x}_2^{N_0},$ and $\tilde{x}_3^{N_0}$ and an edge between $\tilde{x}_i^{N_0}$ and $\tilde{x}_j^{N_0}$ (for $i \neq j$) if

$$\langle \tilde{x}_i^{N_0}, \tilde{x}_j^{N_0} \rangle = \langle x_i^{N_0}, x_j^{N_0} \rangle.$$

We call this the syndrome graph. Consider the number of edges in the syndrome graph. If the syndrome graph is fully connected, then for some collection of constants $a_{ij}$ we must have

$$||\tilde{x}_3^{N_0} - \tilde{x}_1^{N_0} - \tilde{x}_2^{N_0}||^2 = \sum_{i,j} a_{ij} \langle \tilde{x}_i^{N_0}, \tilde{x}_j^{N_0} \rangle$$

(7)

$$= \sum_{i,j} a_{ij} \langle x_i^{N_0}, x_j^{N_0} \rangle$$

(8)

$$= ||x_3^{N_0} - x_1^{N_0} - x_2^{N_0}||^2$$

(9)

$$= 0.$$  

(10)

Thus

$$\tilde{x}_3^{N_0} = \tilde{x}_1^{N_0} + \tilde{x}_2^{N_0},$$

which contradicts the assumption that one of the these vectors was altered.

Thus the graph must be missing at least one edge. Since only one packet can be received in error, the graph cannot be missing all three edges, however. Thus it must have either one
edge or two. If it has exactly one edge, then the vector with no edges must be the one in error, so the other two vectors can be identified as correct. If the graph has two edges, then the vector with two edges must be correct. In the end, then, the decoder can always recover at least one of the transmitted packets correctly. This is of course not the same as recovering one of the source vectors—if the decoder recovers \( x_{3N}^{N} \) then it cannot reproduce any of the source symbols with certainty. But using a “layering” argument one can transform this code into one for which decoding any of the three transmitted packets correctly allows one to recover some positive fraction of the source symbols correctly (see Section [IV]).

The property that the decoder can always correctly recover a transmitted packet even when the number of errors is outside the decoding radius of the code we call 

**guaranteed partial decodability.** This property comes at slight cost in rate compared with conventional MDS codes; one must send the norms and inner products in \( \hat{r} \) in addition to the vectors, and \( x_{3N}^{N} \) can take larger values than either \( x_{1N}^{N} \) or \( x_{2N}^{N} \) because the addition in \( \hat{r} \) is done over the reals. But in the limit of a large source blocklength, this penalty can be made arbitrarily small, and the rate can be made arbitrarily close to 1/2.

We next describe how to extend this idea to general \( N \) and \( T \). The resulting construction is then used to prove Theorem [I]. See [9] for a slightly different decoding algorithm that yields the same performance.

### B. General \((N, T)\): Source

Consider a source message \( x^n \ (x^n \in X^n) \) with length \( n = (N - T)N_0K_0 \) for some large natural numbers \( N_0 \) and \( K_0 \). Divide the message into \((N - T)N_0 \) subvectors, each having \( K_0 \) symbols. We can use a \( K_0 \)-length vector (each entry taken from \( [K] \)) to represent \( K_0 \)-length \([1, ..., K] \); here we use \((0, ..., 0)\) to represent \( K_0 \). Thus, the original source message can also be viewed as an integer vector with length \((N - T)N_0 \). Moreover, \( x^n \) can be viewed as a concatenation of \(N - T\) vectors, each having \( N_0 \) entries in \([1, ..., K_0] \). In what follows, we will view the source vector in this way and write:

\[ x^n = (x_{1N_0}^{N_0}, ..., x_{N(T-1)N_0}^{N_0}, x_{N(T-1)N_0+1}^{N_0}) \]

### C. Encoding Functions

The encoding is performed with the aid of an eligible generator matrix.

**Definition 2:** A is an eligible \((N, N - T)\)-generator matrix if its entries are nonnegative integers and

1) \( A \) is an \( N \times (N - T) \) matrix of the following form:

\[
A = \begin{bmatrix}
    1 & 0 & \cdots & 0 \\
    0 & 1 & \ddots & \vdots \\
    \vdots & \ddots & \ddots & 0 \\
    0 & \cdots & 0 & 1 \\
    a_{1,1} & a_{1,2} & \cdots & a_{1,N-T} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{T,1} & a_{T,2} & \cdots & a_{T,N-T}
\end{bmatrix},
\]

2) Every \((N - T) \times (N - T)\) submatrix of \( A \) is nonsingular. The existence of such matrix is guaranteed by the following lemma.

**Lemma 1:** For any \( T \geq 1 \) and \( N \geq T \) there exists an eligible \((N, N - T)\)-generator matrix of the form

\[
A = \begin{bmatrix}
    1 & 0 & \cdots & 0 \\
    0 & 1 & \ddots & \vdots \\
    \vdots & \ddots & \ddots & 0 \\
    0 & \cdots & 0 & 1 \\
    \alpha_1 & \alpha_2 & \cdots & \alpha_{N-T} \\
    \alpha_2 & \alpha_3 & \cdots & \alpha_{N-T} \\
    \vdots & \vdots & \ddots & \vdots \\
    \alpha_{N-T-1} & \alpha_{N-T} & \cdots & \alpha_{N-T}
\end{bmatrix}, \tag{11}
\]

where \( \alpha_1, \ldots, \alpha_T \) are distinct positive integers. We call such a matrix a \( V \)-matrix, since its lower portion has a Vandermonde structure.

**Proof:** We find the required \( \alpha_1, \ldots, \alpha_T \) by induction. Clearly there exists a positive integer \( \alpha_1 \) such that

\[
A_1 = \begin{bmatrix}
    1 & 0 & \cdots & 0 \\
    0 & 1 & \ddots & \vdots \\
    \vdots & \ddots & \ddots & 0 \\
    0 & \cdots & 0 & 1 \\
    \alpha_1 & \alpha_2 & \cdots & \alpha_{N-T} \\
    \alpha_2 & \alpha_3 & \cdots & \alpha_{N-T} \\
    \vdots & \vdots & \ddots & \vdots \\
    \alpha_{N-T-1} & \alpha_{N-T} & \cdots & \alpha_{N-T}
\end{bmatrix}
\]

is such that every \((N - T) \times (N - T)\) submatrix of \( A_1 \) is nonsingular. Indeed, taking \( \alpha_1 = 1 \) suffices. Now suppose we have positive integers \( \alpha_1, \ldots, \alpha_{t-1} \) such that every \((N - T) \times (N - T)\) submatrix of

\[
A_{t-1} = \begin{bmatrix}
    1 & 0 & \cdots & 0 \\
    0 & 1 & \ddots & \vdots \\
    \vdots & \ddots & \ddots & 0 \\
    0 & \cdots & 0 & 1 \\
    \alpha_1 & \alpha_2 & \cdots & \alpha_{N-T} \\
    \alpha_2 & \alpha_3 & \cdots & \alpha_{N-T} \\
    \vdots & \vdots & \ddots & \vdots \\
    \alpha_{N-T-1} & \alpha_{N-T} & \cdots & \alpha_{N-T}
\end{bmatrix}
\]

is nonsingular. Consider the matrix

\[
A_t = \begin{bmatrix}
    1 & 0 & \cdots & 0 \\
    0 & 1 & \ddots & \vdots \\
    \vdots & \ddots & \ddots & 0 \\
    0 & \cdots & 0 & 1 \\
    \alpha_1 & \alpha_2 & \cdots & \alpha_{N-T} \\
    \alpha_2 & \alpha_3 & \cdots & \alpha_{N-T} \\
    \vdots & \vdots & \ddots & \vdots \\
    \alpha_{N-T-1} & \alpha_{N-T} & \cdots & \alpha_{N-T}
\end{bmatrix}, \tag{12}
\]

viewed as a function of the variable \( \alpha_t \). For any given \((N - T) \times (N - T)\) submatrix of \( A_t \) of the form

\[
A_t = \begin{bmatrix}
    \alpha_1 & \alpha_2 & \cdots & \alpha_{N-T} \\
    \vdots & \vdots & \ddots & \vdots \\
    \alpha_{N-T-1} & \alpha_{N-T} & \cdots & \alpha_{N-T}
\end{bmatrix},
\]

there must exist a natural number \( \alpha_t \) such that this particular \((N - T) \times (N - T)\) matrix is nonsingular, by the following
reasoning. The rows of \( \tilde{A} \) are linearly independent by the induction hypothesis. Let \( [v_1 \ v_2 \ \cdots \ v_{N-T}] \) be a nonzero row vector such that
\[
\begin{bmatrix}
v_1 & v_2 & \cdots & v_{N-T}
\end{bmatrix},
\]
is full rank. Then let \( [\tilde{v}_1 \ \tilde{v}_2 \ \cdots \ \tilde{v}_{N-T}] \) denote the component of \( [v_1 \ v_2 \ \cdots \ v_{N-T}] \) that is orthogonal to the row space of \( \tilde{A} \) and note that \( [\tilde{v}_1 \ \tilde{v}_2 \ \cdots \ \tilde{v}_{N-T}] \) must be nonzero. Then we can find a natural number \( \alpha_t \) so that
\[
\sum_{i=1}^{N-T} \tilde{v}_i \alpha_t^i \neq 0.
\]
This follows from the fact that the left-hand side is a nonzero \((N-T)\)-degree polynomial in \( \alpha_t \), so that there must be a positive integer that is not a root. We conclude that the determinant of the \((N-T) \times (N-T)\) matrix in (12), which is evidently an \((N-T)\)-degree polynomial in \( \alpha_t \), is not identically zero.

Next we show that there is one choice of \( \alpha_t \) that ensures that every \((N-T) \times (N-T)\) submatrix of \( A_t \) is nonsingular. The determinant of any given \((N-T) \times (N-T)\) submatrix is a nonzero \((N-T)\)-degree polynomial in \( \alpha_t \), as noted earlier. Thus it has at most \((N-T)\) roots according to fundamental theorem of algebra. Thus all of the submatrices together have at most \((N-T)(N-T-1)\) \((N-T)\) roots. Since this is finite, there must exist a natural number \( \alpha_t \) that is not a root of any of these polynomials.

The encoding functions are then as follows:

1) We generate \( N \) vectors, \( y_1^{N_0} \ldots y_N^{N_0} \) via the linear transformation
\[
\begin{bmatrix}
y_1^{N_0} \\
\vdots \\
y_N^{N_0}
\end{bmatrix} = A
\begin{bmatrix}
x_1^{N_0} \\
\vdots \\
x_N^{N_0}
\end{bmatrix},
\]
where \( A \) is an eligible \((N,N-T)\)-generator matrix provided by Lemma 1. In particular, we have
\[
y_i^{N_0} = x_i^{N_0}
\]
for all \( 1 \leq i \leq N - T \). We assume that each vector is encoded using \((K_0 + \lceil \log_K (\alpha(N - T)) \rceil)N_0 \) symbols, where \( \alpha = \max_{i,j} \alpha_{i,j} \).

2) We also transmit \((N-T) + \frac{(N-T)(N-T-1)}{2} \) norms/inner products:
\[
F_{ij} = \langle x_i^{N_0}, x_j^{N_0} \rangle, \quad \forall 1 \leq i < j \leq N - T
\]
in all \( N \) packets. This requires that \( \lceil 2K_0 + \log_K N_0 \rceil (N-T) + \frac{(N-T)(N-T-1)}{2} \) extra symbols to be included in each packet.

D. General \((N,T)\): Decoding Functions

The decoder receives \( \tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0} \) and the norms/inner products between \( \{x_1^{N_0}, \ldots, x_N^{N_0}\} \). The decoder will identify a subset of the components of \( y_1^{N_0}, \ldots, y_N^{N_0} \) that is sure have been unaltered\(^a\). We first note that the norms and inner products can always be recovered without error.

\(^a\)Later we will show how to use this identification to prove Theorem 1.

Lemma 2: The decoder can correctly recover \( F_{ij} \) for \( i, j \in \{1, \ldots, N\} \) when \( N \geq 2T + 1 \). Since \( y_1^{N_0}, \ldots, y_N^{N_0} \) are linear combinations of \( x_1^{N_0}, \ldots, x_N^{N_0} \). This means that we can correctly recover \( F_{ij} = \langle y_i^{N_0}, y_j^{N_0} \rangle \) for \( i, j \in \{1, \ldots, N\} \).

The proof of this lemma is straightforward and omitted.

As in the case of \( N = 3, T = 1 \), we call this the syndrome graph. The decoder then performs the following operations:

1) Delete all vertices with no loops and their incident edges in the syndrome graph. Let \( G = (V, E) \) denote the new graph.

2) Let \( V' \) be the set of vertices \( v_i \in V \) such that \( v_i \) is contained in a clique of size at least \( N - T \) in \( G \).

3) Let \( V^* \) be the set of vertices \( v_i \in V' \) such that \( (v_i, v_j) \in E \) for all \( v_j \in V' \).

4) Output the codewords corresponding to the vertices in \( V^* \) as correct.

We shall show that the rate of this code can be made arbitrarily close to \( 1/(N - T) \). We shall then prove that the codewords \( \tilde{y}_i^{N_0} \) on channels corresponding to the vertices \( v_i \in V^* \) are correct.

E. General \((N,T)\): Coding Rate

Proposition 1: For any \( \epsilon > 0 \), there exists natural numbers \( K_0 \) and \( N_0 \) such that the rate of each packet does not exceed \( 1/(N - T) + \epsilon \).

Proof: The rate of each packet is upper bounded by
\[
\frac{(K_0 + \lceil \log_K (\alpha(N - T)) \rceil)N_0}{K_0N_0(N - T)} + \frac{[2K_0 + \log_K N_0]((N - T) + \frac{(N-T)(N-T-1)}{2})}{K_0N_0(N - T)},
\]
where we recall that \( \alpha = \max_{i,j} \alpha_{i,j} \). If we let \( N_0 = K_0 \) and send both to infinity, the second term tends to zero while the first term tends to \( 1/(N - T) \).

F. General \((N,T)\): Partial Decodability of Polytope Codes

We are interested in polytope codes because of the following property.

Theorem 2: Given \( T \), when \( N \geq T + \lceil \frac{T^2}{4} \rceil + 2 \), the decoder can identify least \( N - T - \lceil \frac{T^2}{4} \rceil - 1 \) of the transmitted packets as being received correctly.

We shall prove Theorem 2 via a sequence of lemmas. The first two establish that the codewords associated with nodes in \( V^* \) were received correctly.
Lemma 3: Suppose the k packets $i_1, \ldots, i_k$ are unaltered, and let $i_{k+1}$ be some other packet for which there exists $l_1, \ldots, l_k$ such that
\[ y_{i_{k+1}}^N = \sum_{j=1}^{k} l_j y_{i_j}^N. \]  
(15)

If there is a self-loop on $v_{i_{k+1}}$ in $G$, and $(v_{i_{k+1}}, v_{i_j}) \in E$ for all $j \in \{1, \ldots, k\}$, then the codeword $\hat{y}_{i_{k+1}}^N$ in packet $i_{k+1}$ is also unaltered.

**Proof:** We may rewrite (15) as
\[ \left\| y_{i_{k+1}}^N - \sum_{j=1}^{k} l_j y_{i_j}^N \right\|^2 = 0. \]  
(16)

Since there is a self-loop on $v_{i_{k+1}}$,
\[ \langle \hat{y}_{i_{k+1}}^N, y_{i_{k+1}}^N \rangle = \langle y_{i_{k+1}}^N, y_{i_{k+1}}^N \rangle. \]

Moreover, since there is an edge $(v_{i_{k+1}}, v_{i_j})$ for all $j \in \{1, \ldots, k\}$,
\[ \langle \hat{y}_{i_{k+1}}^N, y_{i_j}^N \rangle = \langle y_{i_{k+1}}^N, y_{i_j}^N \rangle. \]

By expanding the left-hand side of (16) in terms of inner products, as in (7.9)–(10), we have that
\[ 0 = \left\| y_{i_{k+1}}^N - \sum_{j=1}^{k} l_j y_{i_j}^N \right\|^2 \]
\[ = \left\| \hat{y}_{i_{k+1}}^N - \sum_{j=1}^{k} l_j y_{i_j}^N \right\|^2 \]
\[ = \left\| \hat{y}_{i_{k+1}}^N - y_{i_{k+1}}^N \right\|^2 \]
where we have used the assumption that packets $i_1, \ldots, i_k$ are unaltered, and (15). This proves that packet $i_{k+1}$ is unaltered.

Lemma 4: For any $i \in V^*$, we have $\hat{y}_{i}^N = y_{i}^N$.

**Proof:** There must exist $N$ packets that are unaltered. Suppose they are packets $i_1, \ldots, i_{N-T}$. Then $v_{i_1}, \ldots, v_{i_{N-T}}$ must form a clique in the syndrome graph $\hat{G}$. From the definition of $V^*$, for any vertex $i \in V^*$, there is a self-loop on $i$ and $(i, v_j) \in E$ for all $j \in \{1, \ldots, N-T\}$. By construction, every $(N-T) \times (N-T)$ submatrix of generator matrix $A$ is nonsingular. This implies that the vector $y_{i}^N$ can be represented as a linear combination of the other $N-T$ vectors
\[ y_{i_{k+1}}^N = \sum_{j=1}^{N-T} l_j y_{i_j}^N \]
for some linear coefficients $l_j$. By Lemma 3, the codeword $y_{i}^N$ in packet $i$ is unaltered.

The final lemma lower bounds the size of $V^*$. It is a purely graph-theoretic assertion that may have independent uses.

Lemma 5: Consider an undirected graph $G = (V, E)$ with at least $N-T$ nodes in which every node has a self-loop. Let $V'$ denote the set of nodes that are contained in a clique of size at least $N-T$, and suppose that $V'$ is not empty. Let $V^* = \{ v \in V' : (v, \hat{v}) \in E \ \forall \hat{v} \in V' \}$.

Then we have $|V^*| \geq N - F(T)$, where $F(T)$ is defined in Theorem 2.

**Proof:** For any set of edges $E_0$, let
\[ \mathcal{N}(v, E_0) := \{ v_j \in V' \setminus \{ v \} : (v, v_j) \notin E_0 \}. \]

We construct a set of edges $E' \supset E$ as follows. Begin by setting $E' = E$. If there is a pair $v_i, v_j \in V'$ such that $(v_i, v_j) \notin E'$ and
\[ |\mathcal{N}(v_j, E')| > 1, |\mathcal{N}(v_i, E')| > 1, \]
then add $(v_i, v_j)$ to $E'$. Repeat until there is no such pair $v_i, v_j$. Note that for the resulting $E'$, for $v_i \in V'$, $\mathcal{N}(v_i, E') = 0$ if and only if $\mathcal{N}(v_i, E) = 0$. Thus
\[ V^* = \{ v_i \in V' : \mathcal{N}(v_i, E') = 0 \}. \]

Moreover, for any pair $(v_i, v_j) \in V'$ with $(v_i, v_j) \notin E'$, either $|\mathcal{N}(v_i, E')| = 1$ or $|\mathcal{N}(v_j, E')| = 1$. For convenience, we write $\mathcal{N}(v_i) := \mathcal{N}(v_i, E')$ from now on.

Let $v_0$ be an element of $V'$ maximizing $|\mathcal{N}(v)|$, and let $l_0 := |\mathcal{N}(v_0)|$.

Each element $v_k \in V'$ is contained in a clique of $C(v_k)$ of size exactly $N - F(T)$. Since $E' \supset E$, $C(v_k)$ is also a clique on the graph with edges $E'$. Let $C_0 = C(v_0) \setminus \{ v_0 \}$. Fix $v_k \in C_0$, and suppose $(v_i, v_j) \notin E'$ for $v_i \in V'$. We claim that $v_i$ cannot be in $\mathcal{N}(v_k)$. If it were, then $\mathcal{N}(v_k) \geq 2$, in which case $l_0 \geq 2$, which would imply that $|\mathcal{N}(v_k)| \geq 2$. But $(v_0, v_k) \notin E'$, which contradicts the construction of $E'$. Moreover, $v_k$ cannot be in $C(v_k)$ by definition. Hence, if $(v_i, v_k) \notin E'$, then $v_i \in D$, where $D := V' \setminus \mathcal{N}(v_0) \setminus C(v_k)$.

In particular, if $v_j \in C_0 \cap V' \setminus v_0$, then $(v_j, v_k) \notin E'$ for some $v_k \in D$; i.e. $v_j \notin \mathcal{N}(v_k)$. Thus
\[ V' \setminus V^* \subset (V' \setminus C_0 \setminus V^*) \cup (V' \cap C_0 \setminus V^*) \]
\[ \subset \{ v_k \} \cup \{ v_j \setminus C(v_k) \} \cup \{ v_k \setminus D \cup \{ v_j \setminus \mathcal{N}(v_k) \cap C(v_k) \} \}

Hence,
\[ |V'| - |V^*| \leq 1 + |\mathcal{N}(v_k)| + |D| + \sum_{v \in D} |\mathcal{N}(v)| \leq (|D| + 1)(l_0 + 1), \]  
(17)

where we have used the fact that $|\mathcal{N}(v)| \leq l_0$ for all $v \in V'$. Since $\mathcal{N}(v_k), C(v_k) \subset V'$ and $\mathcal{N}(v_k) \cap C(v_k) = \emptyset$:
\[ |D| = |V'| - |\mathcal{N}(v_k)| - |C(v_k)| = |V'| - l_0 + T - N. \]

Substituting this into (17) gives
\[ |V^*| \geq |V'| - (|V'| + 1)(l_0 + 1) \]
\[ = |V'| - (T - l_0 + |V'| - N + 1)(l_0 + 1) \]
\[ \geq N - (T - l_0 + 1)(l_0 + 1) \]
\[ \geq N - F(T). \]

**Proof of Theorem 2:** For each $i \in V^*$, we have $\hat{y}_{i}^N = y_{i}^N$ by Lemma 4 and $|V^*| \geq N - F(T)$ by Lemma 5.

There may be several such cliques, in which case $C(v_k)$ can be chosen to be any one of them.
IV. PROOF OF THEOREM 1

We next show how to use polytope codes to create a code for our original problem. The main difficulty is that, in a polytope code, some of the packets contain only parities, and even if the decoder can determine such packets with certainty, it cannot necessarily recover any of the original source symbols. We circumvent this issue with a layered construction. First we prove the impossibility result in part 1).

A. Proof of Theorem 1 Part 1)

Fix $0 < R < \frac{1}{N - T}$ and $\epsilon > 0$ such that $R + \epsilon < \frac{1}{N - T}$. If there does not exist a feasible code with rate at most $R + \epsilon$ then the conclusion is immediate. Otherwise, consider any feasible code with rate at most $R + \epsilon$, and let $n$ denote the length of the source string that it encodes.

Consider endowing the space $X^n$ with an i.i.d. uniform probability distribution. Since the code is feasible, the source string must be a function of the messages, i.e.,

$$H(x^n|C_1, \ldots, C_N) = 0.$$ 

Since $C_1, \ldots, C_N$ are also deterministic functions of the source string, we must have

$$H(C_1, \ldots, C_N) = H(x^n) = n \log K.$$ 

Therefore

$$H(C_1, \ldots, C_T|C_{T+1}, \ldots, C_N)$$

$$\geq H(C_1, \ldots, C_N) - H(C_{T+1}, \ldots, C_N)$$

$$\geq H(C_1, \ldots, C_N) - \sum_{i=T+1}^{N} H(C_i)$$

$$= n \log K - \sum_{i=T+1}^{N} H(C_i)$$

$$\geq n \log K - (N - T)n(R + \epsilon) \log K > 0.$$ 

Thus $(C_1, \ldots, C_T)$ is not a deterministic function of $(C_{T+1}, \ldots, C_N)$. It follows that there must exist two source sequences $x^n_1$ and $x^n_2$ such that $x^n_1 \neq x^n_2$,

$$f_i(x^n_1) \neq f_i(x^n_2)$$

for some $1 \leq i \leq T$

and

$$f_j(x^n_1) = f_j(x^n_2)$$

for all $T + 1 \leq j \leq N$.

Since the code is feasible, when the decoder receives the message

$$(f_1(x^n_1), f_2(x^n_1), \ldots, f_N(x^n_1)),$$

it must output string $x^n_1$. But then the decoder will also output $x^n_2$ if the true source sequence is $x^n_2$ and the adversary alters the first $T$ packets so that

$$f_1(x^n_1), f_T(x^n_1), f_{T+1}(x^n_2), \ldots, f_N(x^n_2))$$

$$= (f_1(x^n_1), f_T(x^n_1), f_{T+1}(x^n_2), \ldots, f_N(x^n_2))$$

is received. Since $x^n_1$ and $x^n_2$ are different, the distortion of the code is infinite.

B. Proof of Theorem 1 Part 2)

As noted earlier it suffices to show that the R-D pair $(\frac{1}{N - T}, \frac{F(T)}{N})$ is achievable. To show this we use a “layered” construction in which we use $N$ polytope codes whose transformation matrices are row rotations of each other. Divide the source into $N$ equal-sized parts. The first part is encoded into packets using a polytope code with transformation matrix

$$A = \begin{bmatrix}
\alpha^1_1 & \alpha^1_2 & \cdots & \alpha^1_{N-T} \\
0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 1 \\
\alpha^1_T & \alpha^2_T & \cdots & \alpha^{N-T}_T
\end{bmatrix}.$$ 

The second part is encoded using the transformation matrix

$$A = \begin{bmatrix}
\alpha^2_1 & \alpha^2_2 & \cdots & \alpha^{N-T}_T \\
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 1 \\
\alpha^1_T & \alpha^2_T & \cdots & \alpha^{N-T}_T
\end{bmatrix},$$ 

i.e., the first downward row rotation. The other parts of the source are encoded similarly.

The rate of this code can be made arbitrarily close to $1/(N - T)$. At the decoder, we form a syndrome graph in which there is an edge between packets $i$ and $j$ (allowing for $j = i$) if there is an edge between $i$ and $j$ in the syndrome graphs of all of the layers. For this syndrome graph, delete all nodes without self-loops, along with their edges. The resulting graph must have at least one clique of size at least $N - T$, due to the presence of at least $N - T$ unaltered packets. Thus Lemma 5 implies that there are at least $N - F(T)$ nodes that are connected to all nodes contained in a clique of size at least $N - T$. In particular, these $N - F(T)$ nodes must be connected to an unaltered set of nodes of size $N - T$. By Lemma 3 the codewords in all of these $N - F(T)$ packets were received correctly. For each packet, $N - T$ of its layers correspond to systematic rows of the matrix and $T$ layers correspond to parities. Thus the decoder can reconstruct a fraction

$$\frac{(N - T)(N - F(T))}{N(N - T)} = \frac{N - F(T)}{N}$$ 

of the source symbols.

V. AN IMPOSSIBILITY RESULT

By definition, a polytope code

$$(f_1, \ldots, f_N, g)$$

is characterized by $(N, T, A, N_0, K_0)$, where $N$ is the number of packets, $T$ is the maximum number of packets that can be
altered, $A$ is an eligible $(N, N - T)$-generator matrix, and $N_0$ and $K_0$ are encoding parameters (see Section III). From Theorem 1 we know that for

$$N \geq F(T) + 1 \text{ and } \frac{1}{N - T} \leq R \leq \frac{1}{N - 2T},$$

the R-D pair

$$(R, F(T)(N - T)(1 - (N - 2TR))NT)$$

is achievable using polytope codes. However, when $N \leq F(T)$, the decoder in Section III-D no longer works.

This raises the question of whether our design can be improved when $N \leq F(T)$, especially since $F(T)$ grows superlinearly with $T$. We next show the following impossibility result. When $N = F(T)$, for all sufficiently large $N_0$ and $K_0$, our existing polytope code construction lacks the partial decodability property; there exists a set of received packets for which there is no single packet that can be determined to be correct with certainty. Thus, at least as far as partial decodability is concerned, neither the decoder nor the analysis can be improved to relax the $N \geq F(T) + 1$ condition; the code itself would need to change. Recall that, for polytope codes, in order to drive the rate to $1/(N - T)$, we send both $N_0$ and $K_0$ to infinity; see (14).

To state and prove this result, we use the concept of possible transmitted codewords.

**Definition 3:** Fix $N_0$, $K_0$ and $K$. Given a set of received codewords $\{\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}\}$ and recovered $\{F_{j_1j_2}\}$ for $j_1, j_2 \in [N]$ (see Lemma 2), if a set of codewords $\{\tilde{x}_1^{N_0}, \ldots, \tilde{x}_N^{N_0}\}$ satisfies:

1. $F_{j_1j_2} = (\tilde{x}_1^{N_0}, \tilde{x}_N^{N_0})$ for all $j_1, j_2 \in [N]$;
2. The identity $\tilde{x}_N^{N_0} = \tilde{y}_N^{N_0}$ holds for at least $N - T$ values of $j$ out of $j \in [N]$;
3. $\tilde{x}_{N-j}^{N_0} = \sum_{j=1}^{N_0} a_{i,j} \tilde{x}_j^{N_0}$ for all $i \in [T]$,

then this set of codewords is called a Possible Transmitted Codeword (PTC) for $\{\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}\}$ and $\{F_{j_1j_2}\}$. Further, let

\[
\text{PTC}(\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}; F_{j_1j_2}) = \\{\tilde{x}_1^{N_0}, \ldots, \tilde{x}_N^{N_0}\}
\]

denote the set of all possible transmitted codewords for $\{\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}\}$ and $\{F_{j_1j_2}\}$.

**Definition 4:** Fix $N_0$, $K_0$ and $K$ and then fix a set of received packets $\{\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}\}$ and recovered $\{F_{j_1j_2}\}$ for $j_1, j_2 \in [N]$. We call $\{\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}; F_{j_1j_2}\}$ totally undecodable if PTC($\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}; F_{j_1j_2}$) has the following property: for any $i \in [N]$, there exists $\{\tilde{x}_1^{N_0}, \ldots, \tilde{x}_i^{N_0}\}$ and $\{\tilde{x}_{i+1}, \ldots, \tilde{x}_{N_0}\}$ in PTC($\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}; F_{j_1j_2}$) such that $\tilde{x}_i^{N_0} \neq \tilde{x}_{i+1}^{N_0}$.

**Theorem 3:** Fix $T > 1$, $N = F(T)$ and let $A$ be an $(N, N - T)$ $V$-matrix. Then for all sufficiently large $N_0$ and $K_0$ there exists a set of received packets $\{\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}\}$ along with $\{F_{j_1j_2}\}$ such that $\{\tilde{y}_1^{N_0}, \ldots, \tilde{y}_N^{N_0}; F_{j_1j_2}\}$ is totally undecodable.

**Proof:** We begin by showing the conclusion for some $N_0$ and for all sufficiently large $K_0$.

Write the $V$-matrix as:

$$A = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & 1 \\
a_{1,1} & a_{1,2} & \cdots & a_{1,N-T} \\
a_{2,1} & a_{2,2} & \cdots & a_{2,N-T} \\
\vdots & \vdots & \ddots & \vdots \\
a_{T,1} & a_{T,2} & \cdots & a_{T,N-T}
\end{bmatrix}.$$
modified version of $X$, $X_i$, obtained by replacing the $i$th row block in $X$ with
\[
\begin{bmatrix}
c_i^+ + c_i^-(-H_{i+1,1}) & \cdots & c_i^+ + c_i^-(-H_{i+1,L}) \\
\end{bmatrix}
\]
Note that this has the effect of replacing $\begin{bmatrix} \nu_i & \nu_i + \mu_i & \nu_i \end{bmatrix}$ with $\begin{bmatrix} \nu_i + \mu_i & \nu_i \end{bmatrix}$ and vice versa. We view $X$ and the various $X_i$ as different source realizations with blocklength $(N-T)N_0K_0$ where $N_0 = 2L$ and $K_0$ is any integer satisfying
\[
\log_K K_0 \geq \max_{i,j} \mu_{i,j} + \nu_{i,j}.
\]
Since $H$ is Hadamard, the inner product between any two rows of $X$ must equal the inner product between the corresponding rows of $X_i$ for all $i$. Thus, all of these source realizations will result in the same norms and inner products being sent as part of the polytope code. Let $\{F_{j_1,j_2}\}$ denote these norms and inner products.

Next we construct codewords from these source realizations. Let
\[
\bar{X} = AX
\]
and for $i \in \{0, \ldots, \lceil \frac{T}{2} \rceil \}$, let
\[
X_i = AX_i.
\]
Observe that since $\mu_i$ is in the null space of the matrix in \[18\], rows
\[
\left\{N - T + 1, \ldots, N - T + \left\lfloor \frac{T}{2} \right\rfloor - 1\right\}
\]
of $\bar{X}$ and $\bar{X}_i$ will be the same for all $i \in \{0, \ldots, \lceil \frac{T}{2} \rceil - 1\}$.

Finally, construct a set of received packets as follows. Packets 1 through $N - T$ are the first $N - T$ rows of $\bar{X}$, respectively. Packets
\[
\left\{N - T + 1, \ldots, N - T + \left\lfloor \frac{T}{2} \right\rfloor - 1\right\}
\]
are set to be rows $\{N - T + 1, \ldots, N - T + \left\lfloor \frac{T}{2} \right\rfloor - 1\}$ of any of the $X_i$, $i \in \{0, \ldots, \lceil \frac{T}{2} \rceil - 1\}$ (recall that these rows coincide across $\bar{X}$ and these $\bar{X}_i$). For
\[
i \in \left\{N - T + \left\lceil \frac{T}{2} \right\rceil, \ldots, N\right\},
\]
received packet $i$ is set to the corresponding row of $\bar{X}_{i-(N-T+\lceil \frac{T}{2} \rceil)}$. Define the matrix $Y$ to be the set of received packets, one per row, starting with the first.

Now the number of packets that differ between $Y$ and $\bar{X}_i$ is at most
\[
\left\lceil \frac{T}{2} \right\rceil + \left\lceil \frac{T}{2} \right\rceil = T
\]
if $i \in \{0, \ldots, \left\lfloor \frac{T}{2} \right\rfloor - 1\}$. Likewise, codeword $\bar{X}_{\lceil \frac{T}{2} \rceil}$ differs from $Y$ in at most
\[
1 + \left(\left\lceil \frac{T}{2} \right\rceil - 1 \right) + \left\lceil \frac{T}{2} \right\rceil = T.
\]
Thus, $\bar{X}_i$, $i \in \{0, \ldots, \left\lceil \frac{T}{2} \right\rceil \}$ is in PTC$(Y, \{F_{j_1,j_2}\})$. For each $i \in \{1, \ldots, N - T\}$, there exists $i_1$ and $i_2$ s.t. row $i$ in $\bar{X}_{i_1}$ and $\bar{X}_{i_2}$ disagree. Moreover, we can pick $\mu_{\lceil \frac{T}{2} \rceil}$ such that for each $i \in \{N - T + 1, \ldots, N\}$, row $i$ in $\bar{X}_{i_1}$ and $\bar{X}_{\lceil \frac{T}{2} \rceil}$ disagree. This is because for each $i \in \{N - T + 1, \ldots, N\}$, there is at most one value for $\mu_{\lceil \frac{T}{2} \rceil}$ such that row $i$ in $\bar{X}_{i_1}$ and $\bar{X}_{\lceil \frac{T}{2} \rceil}$ are the same. Thus the set of integers for which $\mu_{\lceil \frac{T}{2} \rceil}$ does not satisfy the desired condition has at most $T$ elements, and we can choose $\mu_{\lceil \frac{T}{2} \rceil}$ to be any positive integer not in this set.

This establishes the conclusion for $N_0 = 2L$ and all sufficiently large $K_0$. One can accommodate larger values of $N_0$ by prepending a vector of ones to each of the $X_i$ source realizations.

\section{Distributed Storage Problem Formulation and Results}

\subsection{Distribution Storage System}

A distributed storage system (DSS) is a collection of storage nodes, each holding a portion of a single data file. We assume each node has capacity $\alpha$, meaning it can store an element of $X^{\alpha n}$ for some blocklength $n$, where as before $X = [K]$ is the alphabet set. At any given time, there are $N$ active storage nodes, but individual nodes are unreliable and may fail. When one node fails, a new node is created to replace it. The new node contacts $d$ existing nodes and downloads messages from each one, from which it constructs new storage data. The communication links used to transmit these messages each have capacity $\beta \leq \alpha$, meaning they carry elements of $X^{\alpha n \beta}$. The key property that must be maintained is that at any time in this evolution, a data collector (DC) may contact any $k \leq d$ existing nodes, download their contents, and perfectly reconstruct the original file. The specific evolution of the system, such as which nodes fail, which nodes are contacted when a new node is formed, and when the DC downloads data to reconstruct the file, is arbitrary and unknown \textit{a priori}. We further assume that there is a finite upper limit $L$ of storage nodes over the lifetime of the storage system (i.e. $N$ initial nodes and at most $L - N$ node failures and replacements), where $L$ is known in advance of code design.\footnote{This is a simplifying assumption not always made in the distributed storage literature, but it is necessary for our results to hold.}
B. Adversary Model

We assume the presence of an adversary that may take control of a subset of the storage nodes, and alter any message sent from any of those nodes. This includes messages sent when constructing a new node, as well as data downloaded to a DC. Once a code is fixed, all honest (non-adversarial) nodes behave according to this code, but adversarial nodes may deviate from the code by replacing outgoing transmissions with arbitrary messages. The adversary is omniscient in the sense that it knows the complete stored file, as well as every aspect of the code used by the honest nodes. The adversary may control up to $T$ nodes at any given time. That is, as nodes fail and are replaced, the adversary might continue taking control of new nodes, but at no moment does it control more than $T$ nodes. This is a slightly more pessimistic assumption than in [17], in which the adversary could control a total of $T$ nodes over the entire evolution of the system, whether or not they existed simultaneously.

We say a rate $R$ is achievable for a DSS problem with parameters $(\alpha, \beta, \gamma, \delta, \epsilon, T)$ if for some $n$ there exists a code such that a file $f \in \mathcal{F}^n$ can always be reconstructed without error, no matter the evolution of the system or the adversary actions. The storage capacity $C$ is the supremum of all achievable rates.

C. Bounds on Storage Capacity

Using a combination of a cut-set bound and the Singleton bound, it was shown in [17] Theorem 6] that the storage capacity is upper bounded by

$$C \leq \sum_{i=0}^{k-2T-1} \min\{(d-2T-i)\beta, \alpha\}. \quad (20)$$

When $T = 0$, the above bound reduces to the exact storage capacity for functional repair without an adversary originally found in [7]. In other words, this upper bound states that all adversarial problem with both $d$ and $k$ reduced by $2T$.

Two special points on the storage-bandwidth tradeoff are the so-called Minimum Storage Regenerating (MSR) and Minimum Bandwidth Regenerating (MBR) points. The MSR point is given by

$$\alpha = (d-k+1)\beta, \quad C = (k-2T)\alpha$$

and the MBR point is given by

$$\alpha = (d-2T)\beta, \quad C = \left[\frac{(k-2T)(d-2T)}{2} - \frac{(k-2T)^2}{2}\right] \beta.$$

In [19], achievability with exact repair was proved for the MSR point as long as $d-2T \geq 2(k-2T) - 2$ and for the MBR point for all parameters, using linear matrix-product codes.

The following theorem is our main achievability result for the distributed storage problem. The proof appears in Section VII.

**Theorem 4:** The storage capacity $C$ is lower bounded by

$$C \geq \min \left\{ \frac{k-F(T)-1}{\beta} \sum_{i=0}^{k-2T-1} \min\{(d-F(T)-i)\beta, \alpha\}, \quad (d-T)\beta \right\}. \quad (21)$$

where $F(T)$ is as defined in Theorem 1.

The polytope code used to prove this result, described in detail in Sec. VII, is a decoding procedure to that used for VPEC in Sec. VII that identifies a subset $V^*$ of trustworthy incoming packets. When constructing a new storage node, this procedure identifies at least $d - F(T)$ trustworthy incoming packets, and when decoding the file at a DC, this procedure identifies at least $k - F(T)$ trustworthy nodes. This explains the first term in (21), which corresponds to the capacity of a DSS with no adversary but with $d$ and $k$ each reduced by $F(T)$. The second term in (21), limiting the rate to $(d-T)\beta$, ensures that the file could in principle be decoded from the $d-T$ packets sent to a new storage node from honest nodes; this condition ensures that all adversarial packets are either undetected or detected.

Fig. 3 illustrates the above bounds on the bandwidth-storage tradeoff (i.e., achievable $(\alpha, \beta)$ pairs for $C = 1$) for parameters $k = d = T = 1$. Shown is the outer bound (20) found in [17] and the points achievable with polytope codes by Theorem 4. The matrix-product codes from [19] achieve the MBR point, but not the MSR point for these parameters.

![Fig. 3. Bandwidth-storage tradeoff (i.e., achievable $(\alpha, \beta)$ pairs for $C = 1$) for parameters $k = d = T = 1$. Shown is the outer bound (20) found in [17] and the points achievable with polytope codes by Theorem 4. The matrix-product codes from [19] achieve the MBR point, but not the MSR point for these parameters.](image-url)
that $\alpha$ and $\beta$ are integers; if they are not then they can be scaled up and the blocklength $n$ can be scaled down without changing the problem. Let $r$ be the right-hand side of (21). We show that rate $r$ can be achieved asymptotically. We fix integers $N_0$ and $K_0$, which play the same roles in the polytope code structure as for the VPEC codes described above. The asymptotic rate $r$ is achieved when both $N_0$ and $K_0$ go to infinity. The file $f$ will be composed of $N_0K_0r$ symbols from $\mathcal{X}$. The precise blocklength $n$ and rate $R$ will be determined later. We may reparameterize the file as an integer-valued matrix taking values in $\{1, \ldots, K_0\}^{r \times N_0}$. In particular, we write

$$
\begin{bmatrix}
  x_{1,K_0}^{N_0} \\
  \vdots \\
  x_{r,K_0}^{N_0}
\end{bmatrix}
$$

where $x_{i,K_0}^{N_0}$ is an $N_0$-length vector taking values in $\{1, \ldots, K_0\}$. As before, we form norms/inner products $F_{ij} = \langle x_{i,K_0}^{N_0}, x_{j,K_0}^{N_0} \rangle$, $\forall 1 \leq i \leq j \leq r$

to be included in all packets. We also define for convenience $\mathbf{F}$ to be the vector of all $r + \left(\begin{array}{c}r \\ 2 \end{array}\right)$ norms and inner products.

All packets, both for storage on nodes and for transmissions between nodes, will take the form

$$
(y^{\gamma \times N_0}, \mathbf{F}, A_0)
$$

where $y^{\gamma \times N_0}$ is a $\gamma \times N_0$ integer-valued matrix, and $A_0$ is a $\gamma \times r$ integer-valued matrix indicating that, with no adversarial influence, we would have

$$
y^{\gamma \times N_0} = A_0f.
$$

The parameter $\gamma$ represents the size of the data packet: for a storage packet, $\gamma = \alpha$, and for a transmission packet, $\gamma = \beta$.

Coefficient matrices: Fix an integer parameter $q$, to be determined later; $q$ plays a role akin to the field size in a code over a finite field, in that it governs the size of the coefficient choices. Let $A$ be a matrix in $\{1, \ldots, q\}^{\alpha N \times r}$ such that any $r \times r$ submatrix of $A$ is nonsingular. The existence of such a matrix for sufficiently large $q$ is guaranteed by Lemma [7]. Now we randomly choose the following coefficient matrices, each independent from the others. For all $1 \leq i < j \leq L$, let $B_{i \rightarrow j}$ be a matrix chosen randomly and uniformly from $\{1, \ldots, q\}^{\beta \times \alpha}$. For each $j \in \{n + 1, \ldots, L\}$ and each set $V \subseteq \{1, \ldots, j - 1\}$ of size at least $d - F(T)$, let $C_{V \rightarrow j}$ be a matrix chosen randomly and uniformly from $\{1, \ldots, q\}^{\alpha \times |V|^\beta}$. We will prove that for sufficiently large $q$, with positive probability these coefficient matrices yield a code with the required properties, and hence there is at least one successful code.

We now describe operation of the code.

Data stored on initial nodes: The initial data to be stored on the $N_0$ storage nodes is given by

$$
\begin{bmatrix}
  y_{1,K_0}^{\alpha \times N_0} \\
  \vdots \\
  y_{N,K_0}^{\alpha \times N_0}
\end{bmatrix} = Af
$$

where $y_{i,K_0}^{\alpha \times N_0}$ is an integer-valued matrix of size $\alpha \times N_0$. On the $i$th storage node, we store packet

$$
(y_{i,K_0}^{\alpha \times N_0}, \mathbf{F}, A_i)
$$

where $A_i$ is the $\alpha \times r$ submatrix of $A$ corresponding to node $i$.

Transmissions to form new node: Assume the packet stored on node $i$ is written as in (24). When node $j > i$ is formed, if it contacts node $i$, the packet transmitted from node $i$ to node $j$ is given by

$$
(B_{i \rightarrow j}y_{i,K_0}^{\alpha \times N_0}, \mathbf{F}, B_{i \rightarrow j}A_i).
$$

Formation of new node: When node $j$ is formed, the packet it stores is formed as follows. Node $j$ first determines $\mathbf{F}$ using majority rule among all its received packets. Then it uses the procedure described in Sec. III-D to find a set $V_j^* \subseteq \{1, \ldots, j - 1\}$ of trustworthy incoming packets. By Lemma 5, $|V_j^*| \geq d - F(T)$. Let $z_{j,K_0}^{V_j^*}$ be the $|V_j^*| \times \alpha \times N_0$ matrix composed of the data stored in these trustworthy packets, and let $A_{n-j}$ be the concatenation of the corresponding coefficient matrices. The packet stored at node $j$ is then given by

$$
(C_{V_j^* \rightarrow j}z_{j,K_0}^{V_j^*}, \mathbf{F}, C_{V_j^* \rightarrow j}A_{n-j}).
$$

Decoding at a data collector: To decode the original message, the DC downloads the packets stored on $k$ nodes. After recovering $\mathbf{F}$ using majority rule, it again uses the procedure in Sec. III-D to find a set $V_{DC}^* \subseteq \{1, \ldots, j - 1\}$ of trustworthy incoming packets, where $|V_{DC}^*| \geq k - F(T)$. Let $z_{K_0}^{V_{DC}^*}$ be the $|V_{DC}^*| \times \alpha \times N_0$ matrix composed of the data stored in these trustworthy packets, and let $A_{n-j}$ be the concatenation of the corresponding coefficient matrices. The DC declares its estimate $\hat{f}$ to be the unique $r \times N_0$ matrix such that

$$
z_{K_0}^{V_{DC}^*} = \hat{f}.
$$

If there is no such value or more than one, declare an error.

Rate analysis: First note that $|F_{ij}| \leq K_0^2 N_0$, so the number of symbols required to store $\mathbf{F}$ is at most

$$
(2K_0 + \log K_0) \left[ r + \left(\begin{array}{c}r \\ 2 \end{array}\right) \right].
$$

Next we bound the coefficient matrices $A_i$. By construction, for $i = 1, \ldots, N$, the each element of $A_i$ is in $\{1, \ldots, q\}$. We prove by induction that, for all $j = N + 1, \ldots, L$, each element of $A_j$ is a positive integer no more than

$$
(q^2 \alpha \beta d)^{j-N} q.
$$

Indeed, assume that for all $i < j$, each element of $A_i$ is at most

$$
(q^2 \alpha \beta d)^{j-N-1} q.
$$

Thus, each element of matrix $B_{i \rightarrow j}A_i$ (and hence each element of $A_{n-j}$) is at most

$$
(qa)(q^2 \alpha \beta d)^{j-N-1} q = (q^2 \alpha \beta d)^{j-N} q.
$$
Therefore, for all nodes $i = 1, \ldots, L$, the elements of $A_i$ are at most
\[(q^2 \alpha \beta d)^{L-N} q.\]
Thus the elements of $y_{i,K_0}$ are at most
\[(q^2 \alpha \beta d)^{L-N} q r K_0.\]
Thus to store $y_{i,K_0}$ requires
\[\alpha N_0 (K_0 + \lfloor \log_K (q^2 \alpha \beta d)^{L-N} q r \rfloor)\]
symbols, and to store $A_i$ requires
\[\alpha r \lfloor \log_K (q^2 \alpha \beta d)^{L-N} q \rfloor\]
symbols. The total number of symbols stored on each node in the packet (24) is therefore
\[(2K_0 + \log_K K_0) \left[ r + \left( \frac{r}{2} \right) + \alpha r \lfloor \log_K (q^2 \alpha \beta d)^{L-N} q \rfloor + \alpha N_0 (K_0 + \lfloor \log_K (q^2 \alpha \beta d)^{L-N} q r \rfloor) \right].\]
Similarly, the total number of symbols transmitted from one node to another in the packet (25) is at most
\[(2K_0 + \log_K K_0) \left[ r + \left( \frac{r}{2} \right) + \beta r \lfloor \log_K (q^2 \alpha \beta d)^{L-N} q \rfloor + \beta N_0 (K_0 + \lfloor \log_K (q^2 \alpha \beta d)^{L-N} q r \rfloor) \right].\]
Since $\beta \leq \alpha$, taking the blocklength to be
\[n = \frac{1}{\beta} (2K_0 + \log_K K_0) \left[ r + \left( \frac{r}{2} \right) + r \lfloor \log_K (q^2 \alpha \beta d)^{L-N} q \rfloor + N_0 (K_0 + \lfloor \log_K (q^2 \alpha \beta d)^{L-N} q r \rfloor) \right]\]
allows us to form the storage packets as $\alpha N_0 r$ symbols and the transmission packets as $n \beta$ symbols. Since the file is given by $N_0 K_0 r$ symbols, the rate achieved by this code is
\[R = \frac{N_0 K_0 r}{n}\]
which may be made arbitrarily close to $r$ for sufficiently large $N_0$ and $K_0$.

**Proof of correctness:** The following lemma is proved below.

**Lemma 6:** For sufficiently large $q$, which positive probability on the choice of coefficient matrices $B_{i \rightarrow j}$ and $C_{V \rightarrow j}$, the following hold:
1) for any DC, the corresponding coefficient matrix $\hat{A}$ has rank $r$.
2) for each node $j$, the matrix $\hat{A}_{\rightarrow j}$, consisting of the rows of $\hat{A}_{\rightarrow j}$ corresponding to the honest nodes, has rank $r$.

We first prove that no honest storage nodes ever store faulty data. That is, (23) always holds for stored packets at honest nodes. By construction, the initial honest nodes store only truthful data. We proceed by induction: assume all existing honest nodes hold truthful data, and we show that when a new node $j$ is formed, all packets sent from nodes in $V_j^*$ hold truthful data, even if sent by an adversarial node. There must be at least $d - T$ honest nodes that transmit packets, which, by the inductive hypothesis, all send truthful packets. Thus these $d - T$ nodes form a clique in the syndrome graph. Thus, for any adversarial node $i \in V_j^*$, the syndrome graph must include a self-loop, as well as an edge from $i$ to each of these $d - T$ honest nodes. Moreover, by Lemma 3, matrix $A_{\rightarrow j}$ has rank $r$; in other words, the entire message can be determined from the packets sent from honest nodes. Thus the unaltered data for any node $i \in V_j^*$ is a linear combination of the data sent from honest nodes. Therefore, by Lemma 3, the packet from $i$ to $j$ is unaltered.

Now we show that the DC always decodes correctly. As we have proved, all honest nodes store only truthful data. Thus, when the DC downloads data from $k$ nodes, at least $k - T$ of them contain only truthful data. By a similar argument as above, any node in $V_{DC}^*$ contains truthful data. Since by Lemma 6 matrix $\hat{A}$ has rank $r$, the only value $\tilde{f}$ satisfying (26) is the true value of the file $f$.

**Proof of Lemma 6:** We make use of the information flow graph developed in (17). The basic insight is that the distributed storage problem can be posed as a multicast network coding problem on the information flow graph, described as follows. The graph, denoted $G_{DSS}$, consists of a source node $S$, for each storage node $i$ a pair of nodes $x_i^r$ and $x_i^s$, and for each DC a node $DC_j$. Each pair of storage nodes are connected by a link $x_i^r \rightarrow x_i^s$ of capacity $r$. For the initial storage nodes $j = 1, \ldots, N$, there is a link $S \rightarrow x_{i_0}^r$ of infinite capacity. For subsequent storage nodes $j > N$, there is a link $x_{i_0}^r \rightarrow x_i^r$ of capacity $\beta$ for each of the $d$ nodes $i$ that transmit a message to node $j$. For each data collector, there is a link $x_i^s \rightarrow DC_j$ of infinite capacity for each of the $k$ nodes $i$ from which the DC downloads data. It is shown in (17) Lemma 2) that for any DC, the min-cut of this graph from the source $S$ to $DC_j$ is lower bounded by
\[\sum_{i=0}^{k-1} \min\{(d - i) \beta, \alpha \}.\]

Consider the subgraph $\hat{G}_{DSS}$ of the information flow graph in which, for each node $j > N$, the links incoming to $x_i^s$ from nodes not in $V_j^*$ are deleted, and similarly links to the DC not in $V_{DC}^*$ are deleted. Note that, on this subgraph, the polytope code behaves essentially like an ordinary linear network code without adversaries, except that linear operations are over the integers rather than a finite field. We further define, for each node $i > N$, a different subgraph $\hat{G}_{DSS}^{(i)}$ of the information flow graph, which is the same as $\hat{G}_{DSS}$ except that all incoming links to $x_{i_0}^r$ from honest nodes are retained.

By standard arguments in linear network coding (see, for example, (13)), which apply equally well for integer operations as for a finite field, for sufficiently large $q$, with probability approaching 1, the rank of a coefficient matrix will be equal to the min-cut of the corresponding information flow graph. Therefore, to prove the lemma it is enough to prove the following two min-cut properties:
1) On $\hat{G}_{DSS}$, the min-cut from $S$ to $DC_j$ for any $j$ is at least $r$.
2) On $\hat{G}_{DSS}^{(i)}$, the min-cut from $S$ to $x_{i_0}^r$ is at least $r$.

The first of these properties is easily proved using existing information flow results. In particular, since $|V_j^*| \geq d - F(T)$
and \( |V_{\text{DC}}| \geq k - F(T) \), we may apply [7, Lemma 2] to find that the min-cut on \( G_{\text{DSS}} \) from \( S \) to \( \text{DC}_j \) is lower bounded by

\[
\sum_{i=0}^{k-F(T)-1} \min \{(d - F(T) - i)\beta, \alpha\} \geq r.
\]

The proof of the second min-cut property requires a slight modification of that of [2, Lemma 2]. Let \((U, \bar{U})\) be any cut on \( G_{\text{DSS}}^{(i)} \) where \( S \in U \) and \( \bar{x}^i \in \bar{U} \). Let \( C \) be the set of edges connecting \( U \) to \( \bar{U} \). Let \( z \) be the number of output nodes in \( \bar{U} \). Let \( x_{\text{DSS}} \) be the first such node in \( \bar{U} \). There are two cases:

- If \( \bar{x}^j \in U \), then the edge \( \bar{x}^j \rightarrow x^j \) is in \( C \).
- If \( \bar{x}^j \in \bar{U} \), then the incoming edges to \( \bar{x}^j \), all of which come from output nodes in \( U \), are in \( C \). There are at least \( d - F(T) \) of these edges.

These edges contribute at least \( \min \{(d - F(T))\beta, \alpha\} \) to the cut capacity.

Let \( x_{\text{DSS}}^2 \) be the next output node in \( \bar{U} \). Again there are two cases:

- If \( x_{\text{DSS}}^j \in U \), then the edge \( x_{\text{DSS}}^j \rightarrow x^j \) is in \( C \).
- If \( x_{\text{DSS}}^j \in \bar{U} \), then the incoming edges to \( x_{\text{DSS}}^j \), all of which come from output nodes in \( U \), are in \( C \). There are at least \( d - F(T) \) of these edges.

These edges contribute at least \( \min \{(d - F(T) - 1)i\beta, \alpha\} \) to the cut capacity. Continuing this reasoning, we accumulate a total cut capacity of

\[
\sum_{i=0}^{d-F(T)-1} \min \{(d - F(T) - i)\beta, \alpha\}.\]

In addition, since \( x_{\text{in}}^i \) has at least \( d - T \) incoming edges, if \( z < d - T \) then at least \( d - T - z \) incoming edges to \( x_{\text{in}}^i \) are in \( C \). Thus, the total cut capacity is at least

\[
\min_{z \leq d - F(T)} \sum_{i=0}^{z-1} \min \{(d - F(T) - i)\beta, \alpha\} + (d - T - z)\beta.
\]

If \( z \leq d - F(T) \), then since \( F(T) \geq T \) we have \( z \leq d - T \), so (27) is at least

\[
\sum_{i=0}^{z-1} \min \{(d - F(T) - i)\beta, \alpha\} + (d - T - z)\beta
\]

\[
\geq \sum_{i=0}^{z-1} \beta + (d - T - z)\beta
\]

\[
= (d - T)\beta \geq r.
\]

If \( z > d - F(T) \), then (27) is at least

\[
\sum_{i=0}^{d-F(T)-1} \min \{(d - F(T) - i)\beta, \alpha\}
\]

\[
\geq \sum_{i=0}^{k-F(T)-1} \min \{(d - F(T) - i)\beta, \alpha\}
\]

\[
\geq r.
\]

Therefore, in any case the min-cut from \( S \) to \( x_{\text{in}}^i \) is at least \( r \). ■

### Appendix A

**Supporting Lemmas**

**Lemma 7:** For any integer \( m > 1 \) and \( \Lambda \in \mathbb{Z}^{(m-1) \times m} \), there exists a non-zero vector \( x^m \in \mathbb{Z}^m \) such that \( \Lambda x^m = 0 \). Furthermore, if \( \text{rank}(\Lambda') = m - 1 \) for all \( (m - 1) \times (m - 1) \) submatrices \( \Lambda' \) of \( \Lambda \), then any such an \( x^m \) must be in \( \{0\}^m \).

**Proof:** Let \( \lambda_1^m, \ldots, \lambda_{m-1} \) denote the rows of \( \Lambda \). Using the Gram-Schmidt procedure, we may assume that \( \lambda_1^m, ..., \lambda_{m-1} \) are orthogonal. Since \( \lambda_1^m, ..., \lambda_{m-1} \) cannot span \( \mathbb{R}^m \) but \( \mathbb{N}^m \) does, there must exist a vector \( \lambda^m \in \mathbb{N}^m \) that is not in the span of \( \lambda_1, ..., \lambda_{m-1} \). Then the vector:

\[
\lambda^m = \sum_{i=1}^{m-1} (\lambda_1^m)^T \lambda_i \lambda_i^m,
\]

where the sum excludes those \( i \) for which \( \lambda_i^m \) is the zero vector, is in \( \mathbb{Q}^m \) and is orthogonal to \( \lambda_1, ..., \lambda_{m-1} \). Multiplying \( \lambda^m \) by the least common denominator gives a non-zero integer solution to \( \Lambda x^m = 0 \).

When \( \text{rank}(\Lambda') = m - 1 \) for all \( \Lambda' \), we prove that all the entries of \( x^m \) must be non-zero by contradiction. Without loss of generality, suppose that \( x_1 = 0 \). Then

\[
\begin{bmatrix}
\Lambda_2 & \cdots & \Lambda_m
\end{bmatrix}
\begin{bmatrix}
x_2 \\
\vdots \\
x_m
\end{bmatrix} = 0,
\]

where \( \Lambda_2 \) through \( \Lambda_m \) are the second through last columns of \( \Lambda \). Now \( \begin{bmatrix} \Lambda_2 & \cdots & \Lambda_m \end{bmatrix} \) is a non-singular matrix by hypothesis. The above linear system then has a unique solution, namely the zero vector. This implies that \( x^m \) is the zero vector, which is a contradiction. ■

**Lemma 8:** Let \( \alpha_1, \ldots, \alpha_m \) be distinct natural numbers. Then for any integer \( k \geq 0 \), every \( m \times m \) submatrix of

\[
M = \begin{bmatrix}
\alpha^k_1 & \alpha^{k+1}_1 & \cdots & \alpha^{k+m}_1 \\
\alpha^k_2 & \alpha^{k+1}_2 & \cdots & \alpha^{k+m}_2 \\
\vdots & \vdots & \ddots & \vdots \\
\alpha^k_m & \alpha^{k+1}_m & \cdots & \alpha^{k+m}_m
\end{bmatrix}
\]

is nonsingular.

**Proof:** Let \( a = [a_0 \ a_1 \cdots a_m]^T \) be such that \( Ma = 0 \) and \( a_i = 0 \) for some \( i \). It suffices to show that \( a \) must be the zero vector. Now \( a \) is in the nullspace of

\[
M = \begin{bmatrix}
1 & \alpha_1^m & \cdots & \alpha_m^m \\
1 & \alpha_2^m & \cdots & \alpha_m^m \\
\vdots & \vdots & \ddots & \vdots \\
1 & \alpha_1^m & \cdots & \alpha_m^m
\end{bmatrix}.
\]

Consider the polynomial

\[
P(x) = \sum_{i=0}^{m} a_i x^i.
\]

Evidently \( P \) is a degree-\( m \) polynomial with roots \( \alpha_1, \ldots, \alpha_m \). There is a unique nonzero degree-\( m \) polynomial with these roots, however, namely,

\[
P'(x) = \prod_{i=0}^{m} (x - \alpha_i) = \sum_{i=0}^{m} a_i' x^i.
\]
Since all of the $\alpha_i$ are positive, all of the $\alpha'_i$ must be nonzero. It follows that $P(\cdot) \neq P'(\cdot)$ and so $P(\cdot)$ must be the all-zero polynomial.

ACKNOWLEDGMENT

The authors wish to thank Ebad Ahmed for his contributions to this work during its early stages. The scheme in (2), in particular, is due to him. This research was supported by the Army Research Office under grant W911NF-13-1-0455 and by the National Science Foundation under grants CCF-1117128, CCF-1218578, and CCF-1453718.

REFERENCES

[1] E. Ahmed and A. B. Wagner. Lossy source coding with byzantine adversaries. In Proc. IEEE ITW, pages 462–466, Oct 2011.
[2] E. Ahmed and A. B. Wagner. Erasure multiple descriptions. Information Theory, IEEE Transactions on, 58(3):1328–1344, 2012.
[3] N. Cai and R. W. Yeung. Network error correction, ii: Lower bounds. Communications in Information & Systems, 6(1):37–54, 2006.
[4] P. A. Chou and Z. Miao. Rate-distortion optimized streaming of packetized media. Multimedia, IEEE Transactions on, 8(2):390–404, 2006.
[5] T. M. Cover and J. A. Thomas. Elements of information theory. John Wiley & Sons, 2012.
[6] T. K. Dikaliotis, A. G. Dimakis, and T. Ho. Security in distributed storage systems by communicating a logarithmic number of bits. In Proc. Int. Symp. Information Theory, pages 1948 –1952, June 2010.
[7] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran. Network coding for distributed storage systems. IEEE Trans. Inf. Theory, 56(9):4539 –4551, Sep. 2010.
[8] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh. A survey on network codes for distributed storage. Proceedings of the IEEE, 99(3):476 –489, Mar. 2011.
[9] X. Fan, A. B. Wagner, and E. Ahmed. Polytope codes for large-alphabet channels. In Proc. Annual Allerton Conf. on Comm. Control, and Computing, pages 948–955, 2013.
[10] A. E. Gamal and T. M. Cover. Achievable rates for multiple descriptions. Information Theory, IEEE Transactions on, 28(6):851–857, 1982.
[11] V. K. Goyal. Multiple description coding: Compression meets the network. Signal Processing Magazine, IEEE, 18(5):74–93, 2001.
[12] T. Ho and D. Lun. Network coding: an introduction, volume 6. Cambridge University Press Cambridge, 2008.
[13] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong. A Random Linear Network Coding Approach to Multicast. IEEE Trans. Inf. Theory, 52(10):4413–4430, 2006.
[14] O. Kosut. Polytope codes for distributed storage in the presence of an active omniscient adversary. In Proc. Int. Symp. Information Theory, July 2013.
[15] O. Kosut, L. Tong, and D.N.C. Tse. Polytope codes against adversaries in networks. IEEE Trans. Inf. Theory, 60(6):3308–3344, June 2014.
[16] F. Oggier and A. Datta. Byzantine fault tolerance of regenerating codes. In Peer-to-Peer Computing (P2P), 2011 IEEE International Conference on, pages 112–121, 2011.
[17] S. Pawar, S. El Rouayheb, and K. Ramchandran. Securing dynamic distributed storage systems against eavesdropping and adversarial attacks. IEEE Trans. Inf. Theory, 57(10):6734 –6753, Oct. 2011.
[18] R. Puri and K. Ramchandran. Multiple description source coding using forward error correction codes. In Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference on, volume 1, pages 342–346. IEEE, 1999.
[19] K. V. Rashmi, N. B. Shah, K. Ramchandran, and P. Y. Kumar. Regenerating codes for errors and erasures in distributed storage. In Proc. Int. Symp. Information Theory, pages 1202–1206, 2012.
[20] T. Richardson and R. Urbanke. Modern coding theory. Cambridge University Press, 2008.
[21] N. Silberstein, A. S. Rawat, and S. Vishwanath. Error resilience in distributed storage via rank-metric codes. In Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on, pages 1150–1157, 2012.
[22] R. Singleton. Maximum distance-nary codes. Information Theory, IEEE Transactions on, 10(2):116–118, 1964.
[23] J. J. Sylvester. Lx. thoughts on inverse orthogonal matrices, simultaneous signsuccessions, and tessellated pavements in two or more colours, with applications to newton’s rule, ornamental tile-work, and the theory of numbers. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 34(232):461–475, 1867.
[24] R. W. Yeung and N Cai. Network error correction, i: Basic concepts and upper bounds. Communications in Information & Systems, 6(1):19–35, 2006.