

# Field Trace Polynomial Codes for Secure Distributed Matrix Multiplication

Roberto Assis Machado\*, Rafael G. L. D'Oliveira†, Salim El Rouayheb‡ and Daniel Heinlein§

\*SMSS, Clemson University, USA, robertoassismachado@gmail.com

†RLE, Massachusetts Institute of Technology, USA, rafael@mit.edu

‡ECE, Rutgers University, USA, salim.elrouayheb@rutgers.edu

§Department of Communications and Networking, Aalto University, Finland, daniel.heinlein@aalto.fi

**Abstract**—We consider the problem of communication efficient secure distributed matrix multiplication. The previous literature has focused on reducing the number of servers as a proxy for minimizing communication costs. The intuition being that the more servers are used, the higher is the communication cost. We show that this is not the case in general. Our central technique relies on adapting results from the literature on repairing Reed-Solomon codes in which, instead of downloading the whole output of a computing task, a user downloads field traces of it. We present Field Trace Polynomial (FTP) codes, a family of codes, that leverage this technique and characterize regimes for which they outperform existing codes in the literature.

**Index Terms**—security, distributed computation, coding theory

An extended version of this paper, including all proofs, is accessible at: <https://arxiv.org/abs/2108.08798>

## I. INTRODUCTION

We consider the problem of secure distributed matrix multiplication (SDMM) in which a user has two matrices,  $A \in \mathbb{F}_q^{a \times b}$  and  $B \in \mathbb{F}_q^{b \times c}$ , and wishes to compute their product,  $AB \in \mathbb{F}_q^{a \times c}$ , with the assistance of  $N$  servers, without leaking any information about either  $A$  or  $B$  to any server. We assume that all servers are honest but curious (passive), in that they are not malicious and will faithfully follow the pre-agreed upon protocol. However, any  $T$  of them may collude to try to eavesdrop and deduce information about either  $A$  or  $B$ .

We follow the setting proposed in [1], with many follow-up works [2]–[13]. The performance metric initially used was the download cost, i.e., the total amount of data downloaded by the users from the server. Subsequent work has also considered the upload cost [11], the total communication cost [14], and computational costs [9].

Different partitionings of the matrices lead to different trade-offs between upload and download costs. In this paper, we consider the inner product partitioning given by  $A = [A_1 \ \cdots \ A_L]$  and  $B^\top = [B_1^\top \ \cdots \ B_L^\top]$  such that  $AB = A_1B_1 + \cdots + A_LB_L$ , where all products  $A_\ell B_\ell$  are well-defined and of the same size. Under this partitioning, a polynomial code is a polynomial  $h(x) = f(x) \cdot g(x)$ , whose coefficients encode the submatrices  $A_k B_\ell$ . The  $N$  servers compute the evaluations  $h(\alpha_1), \dots, h(\alpha_N)$  for certain  $\alpha_1, \dots, \alpha_N$ . The next step is where the scheme we propose differs from previous works. In previous works, the servers send these evaluations to the user. The polynomial  $h(x)$  is

constructed so that no  $T$ -subset of evaluations reveals any information about  $A$  or  $B$  ( $T$ -security), but so that the user can reconstruct  $AB$  given all  $N$  evaluations (decodability). In order to contrast them with our approach, we refer to these schemes as *traditional* polynomial schemes.

Examples of traditional polynomial schemes for the inner product partitioning are the secure MatDot codes in [6] and the codes in [11]. The main focus in the literature was on minimizing the minimum amount of helping servers  $N$ , also known as the recovery threshold, in order to reduce the communication cost. The intuition being, that the more servers used, the higher the communication cost. We show that this is not generally the case, i.e., in some cases it is possible to reduce the total communication by contacting more servers.

In this paper, we present field trace polynomial (FTP) codes, a non-traditional polynomial scheme inspired by techniques from the repair literature, specifically the trace-based methods for repairing Reed-Solomon codes, first introduced in [15]. In the Reed-Solomon codes repair setting, servers store different evaluations of a polynomial  $h(\alpha_i)$ , some of which may be lost due to node failures. The repair problem consists of finding schemes to recover any lost evaluation while minimizing the download bandwidth (referred to as the repair bandwidth in that setting) [16]. The key tool utilized is the field trace. If  $\mathbb{E}$  is a finite algebraic extension of a field  $\mathbb{F}$ , then the field trace  $\text{tr}_{\mathbb{E}/\mathbb{F}} : \mathbb{E} \rightarrow \mathbb{F}$  is a linear functional over the extension field  $\mathbb{E}$  when seen as a vector space over  $\mathbb{F}$ . Instead of sending full evaluations of  $h(\alpha_i) \in \mathbb{E}$ , the servers can repair an evaluation by sending traces  $\text{tr}_{\mathbb{E}/\mathbb{F}}(h(\alpha_i)) \in \mathbb{F}$ . Since  $\mathbb{F}$  is a sub-field of  $\mathbb{E}$ , this results in a smaller download cost. Indeed, the download cost decreases as more servers are used.

FTP codes follow the same idea. By increasing the number of servers and transmitting traces of the evaluations, instead of the whole evaluations, FTP codes can obtain better download costs. However, since the SDMM setting is a computation-offload setting and not just an information storage one, we must also account for the upload cost. Thus, care must be taken so that the decrease in the download cost, obtained from the use of the field trace, is not outgrown by the increase in the upload cost, from having to contact more servers. In Theorem 1, we characterize the total communication rate achieved by FTP codes.

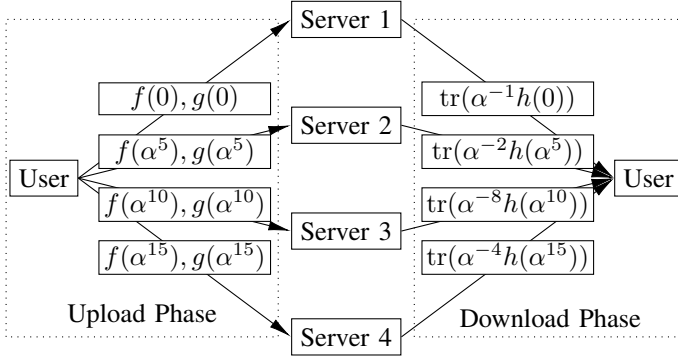


Fig. 1: An example of an FTP code detailed in Section III. The user computes carefully chosen evaluations of the polynomials  $f(x)$  and  $g(x)$  and uploads them to the servers. Each server then computes the product of their received evaluations, which is itself an evaluation of the polynomial  $h(x) = f(x) \cdot g(x)$ . In a traditional polynomial scheme, the servers would then transmit these evaluations to the user. Utilizing an FTP code, the servers compute the trace of these evaluations and send them instead. Thus, decreasing the download cost.

**Theorem 1.** Let  $L$  and  $T$  be positive integers,  $p_1, \dots, p_L$  be prime numbers, listed in increasing order,  $q_0$  be a prime power, and set  $q = q_0^{p_1 p_2 \dots p_L}$ . Let  $A \in \mathbb{F}_q^{a \times b}$ ,  $B \in \mathbb{F}_q^{b \times c}$  be two matrices, and  $N_i = p_i + 2L + 2T - 2$ , for every  $i \in [L]$ . Then, there exists an FTP code with partitioning parameter  $L$  and security parameter  $T$ , which securely computes  $AB$  utilizing  $N_L$  servers with a total communication rate of

$$\mathcal{R} = \left( \frac{N_L b}{L} \left( \frac{1}{a} + \frac{1}{c} \right) + \sum_{i=1}^L \frac{N_i}{p_i} \right)^{-1}. \quad (1)$$

In Theorem 2, we show that FTP codes outperform any traditional polynomial scheme when the number of columns in  $A$  (or rows in  $B$ ) is sufficiently smaller than both the number of rows in  $A$  and columns in  $B$ .

**Theorem 2.** For every traditional polynomial scheme with partitioning parameter  $L$  and security parameter  $T$ , there exists a constant  $K > 0$ , which depends on  $L$  and  $T$ , and an FTP code with communication rate higher than that of the traditional polynomial scheme, whenever the matrix dimensions  $a$ ,  $b$ , and  $c$  are such that  $b \left( \frac{1}{a} + \frac{1}{c} \right) < K$ .

The expression for the constant  $K$  is given in (5). Theorem 2 is proved by constructing a particular FTP code and comparing it to the trivial lower bound on the recovery threshold of a traditional polynomial scheme, namely that  $N > L$ . To provide some context, the current state of the art for inner product partitioning is given by the traditional polynomial scheme in [11] which has a recovery threshold of  $N = L + 2T$  (or  $N = L + T$  with precomputations).

## A. Related Work

For distributed computations, polynomial codes were originally introduced in [17] in order to mitigate stragglers in distributed matrix multiplication. This was followed by a series of works, [18]–[21].

The literature on SDMM has also studied different variations on the model we focus on here. For instance, in [11], [22]–[24] the encoder and decoder are considered to be separate, in [22] servers are allowed to cooperate. In [25] the authors consider a hybrid setup between SDMM and private information retrieval where the user has a matrix  $A$  and wants to privately multiply it with a matrix  $B$  belonging to some public list. FTP codes can be readily used or adapted to many of these settings as has been done with other polynomial schemes (e.g., [12], [26]).

There is now a vast literature on the repair problem for distributed storage systems (e.g., [16], [27]–[29]). The field trace method relevant to us was developed in [15] and later extended in [30]. Methods from repair were used in [31]–[33] to construct communication-efficient secret sharing.

## B. Main Contributions

Our main contributions are summarized below.

- We show a connection between SDMM and the repair problem of Reed-Solomon codes for distributed storage. This is done by treating the user as a repair node wishing to restore a polynomial evaluation. The essential difference between these settings is that in SDMM the upload cost is important. This occurs because the SDMM setting is a computation-offload setting and not just an information storage one, in which the storage upload can be amortized. Other differences include more flexibility in choosing the amount of servers, in designing the code, and in choosing the evaluation points.
- By adapting the techniques used in the repair of Reed-Solomon codes, we present FTP codes for SDMM. Contrary to traditional polynomial codes, FTP codes achieve higher download rates by communicating with more servers. We show that they are secure, decodable, and present their total communication rate in Theorem 1.
- In Theorem 2, we show that FTP codes outperform any traditional polynomial scheme when the number of columns in  $A$  (or rows in  $B$ ) is sufficiently smaller than both the number of rows in  $A$  and columns in  $B$ .

## II. PRELIMINARIES

In this section, we introduce some notation and concepts needed for the rest of the paper. For example, we define  $[M, N] = \{M, M + 1, \dots, N\}$  and  $[M] = [1, M]$ .

**Definition 1.** Let  $\mathcal{C}$  be a linear code of length  $n$  over a finite field  $\mathbb{F}_q$ . The dual code of  $\mathcal{C}$  is the linear subspace

$$\mathcal{C}^\perp = \left\{ d \in \mathbb{F}_q^n : \sum_{i=1}^n d_i c_i = 0 \quad \forall c \in \mathcal{C} \right\}.$$

**Definition 2.** Let  $V = (v_1, v_2, \dots, v_n) \in \mathbb{F}_q^n$  be a vector with non-zero entries and  $\Omega = \{\alpha_1, \alpha_2, \dots, \alpha_n\} \subset \mathbb{F}_q$  a set of distinct elements. The *Generalized Reed-Solomon Code* with parameters  $n, k, \Omega$ , and  $V$  is given by  $GRS_{\mathbb{F}_q}(n, k, \Omega, V) = \{(v_1 f(\alpha_1), v_2 f(\alpha_2), \dots, v_n f(\alpha_n)) : f(x) \in \mathbb{F}_q[x] \text{ and } \deg(f) < k\}$ .

If  $V = (1, 1, \dots, 1)$ , then  $GRS_{\mathbb{F}_q}(n, k, \Omega, V)$  is the Reed-Solomon Code  $RS_{\mathbb{F}_q}(n, k, \Omega)$ . In addition, the dual code of an  $RS_{\mathbb{F}_q}(n, k, \Omega)$  is  $GRS_{\mathbb{F}_q}(n, n-k, \Omega, V')$ , where  $V' = (v'_1, \dots, v'_n)$  and  $(v'_i)^{-1} = \prod_{j \neq i} (\alpha_i - \alpha_j)$ .

Let  $\mathbb{E}$  be a finite algebraic extension of a field  $\mathbb{F}$ . The degree  $s = [\mathbb{E} : \mathbb{F}]$  of a field extension  $\mathbb{E}/\mathbb{F}$  is the dimension of  $\mathbb{E}$  as a vector space over  $\mathbb{F}$ . Thus, any element  $v \in \mathbb{E}$  can be expressed as a vector  $(v_0, v_1, \dots, v_{s-1}) \in \mathbb{F}^s$ .

**Definition 3.** Let  $\mathbb{E}$  be a finite algebraic extension of a field  $\mathbb{F}$ . The field trace  $\text{tr}_{\mathbb{E}/\mathbb{F}} : \mathbb{E} \rightarrow \mathbb{F}$  is the  $\mathbb{F}$ -linear functional  $\text{tr}_{\mathbb{E}/\mathbb{F}}(x) = x + x^{|\mathbb{F}|} + x^{|\mathbb{F}|^2} + \dots + x^{|\mathbb{F}|^{s-1}}$ .

Note that the codomain of the field trace is the subfield  $\mathbb{F}$ . It is this fact that allows savings in the download cost. We also note that, given an  $\mathbb{F}$ -basis  $\{\lambda_0, \lambda_1, \dots, \lambda_{s-1}\}$  of  $\mathbb{E}$ , there exists a *trace-dual*  $\mathbb{F}$ -basis  $\{\mu_0, \mu_1, \dots, \mu_{s-1}\}$  of  $\mathbb{E}$ , i.e., such that  $\text{tr}_{\mathbb{E}/\mathbb{F}}(\lambda_i \mu_j)$  equals 1 if  $i = j$ , and equals 0 otherwise.

The next proposition plays a crucial role in proving the decodability of FTP codes in Lemma 1.

**Proposition 1** (Page 759 in [34]). *Let  $\{\lambda_0, \lambda_1, \dots, \lambda_{s-1}\}$  and  $\{\mu_0, \mu_1, \dots, \mu_{s-1}\}$  be trace-dual  $\mathbb{F}$ -bases of  $\mathbb{E}$ . Then, given an element  $\beta \in \mathbb{E}$ , the coefficients of its expansion in the basis  $\{\mu_0, \mu_1, \dots, \mu_{s-1}\}$  are given by  $\text{tr}_{\mathbb{E}/\mathbb{F}}(\lambda_i \beta)$ , so that*

$$\beta = \sum_{i=0}^{s-1} \text{tr}_{\mathbb{E}/\mathbb{F}}(\lambda_i \beta) \mu_i. \quad (2)$$

### III. A MOTIVATING EXAMPLE: $L = T = 1$

We begin our description of FTP codes with the following example, which we present in as much detail as possible in order to showcase the essential ingredients of the scheme. We compare the traditional polynomial scheme using  $N' = 3$  servers with an FTP code using  $N = 4$  servers.

In this example, a user wishes to multiply two matrices  $A \in \mathbb{F}_{16}^{a \times b}$  and  $B \in \mathbb{F}_{16}^{b \times c}$  with the assistance of non-colluding helper servers. The solution to this via traditional polynomial schemes utilizes  $N' = 3$  servers and involves picking two random matrices  $R \in \mathbb{F}_{16}^{a \times b}$  and  $S \in \mathbb{F}_{16}^{b \times c}$  and constructing the polynomials  $f'(x) = A + Rx$  and  $g'(x) = B + Sx$ . The user then selects three distinct non-zero elements  $\beta_1, \beta_2, \beta_3 \in \mathbb{F}_{16}$  and uploads both  $f'(\beta_i)$  and  $g'(\beta_i)$  to Server  $i$ . Each server then computes the product  $f'(\beta_i) \cdot g'(\beta_i)$ . This is equivalent to computing an evaluation  $h'(\beta_i)$  of the polynomial  $h'(x) = AB + (AS + RB)x + RSx^2$ . The user then downloads each  $h'(\beta_i)$ , obtaining three evaluations of a polynomial of degree two. Therefore, the user can retrieve the polynomial  $h(x)$  and compute  $h(0) = AB$  as desired.

Security of the traditional scheme follows from the fact that  $I(f'(\beta_i), g'(\beta_i); A, B) = 0$ . As for the communication

costs, first the user uploads  $f'(\beta_i)$  and  $g'(\beta_i)$ , which cost  $ab$  and  $bc$ , symbols respectively, three times. Thus, the upload cost is  $3(ab + bc)$  symbols of  $\mathbb{F}_{16}$ . Then, the user downloads  $h(\beta_i)$ , which costs  $ac$  symbols of  $\mathbb{F}_{16}$ , three times, obtaining a download cost of  $3ac$  symbols of  $\mathbb{F}_{16}$ . Since the user retrieves  $AB \in \mathbb{F}_{16}^{a \times c}$ , which consists of  $ac$  symbols of  $\mathbb{F}_{16}$ , the total communication rate is given by  $\mathcal{R}' = \frac{ac}{3ab+3bc+3ac}$ .

Based on techniques from the literature on repairing Reed-Solomon codes for distributed storage, we present an FTP code which obtains a lower download cost by utilizing  $N = 4$  servers. Let  $\alpha \in \mathbb{F}_{16}$  be an algebraic element of degree 4 such that  $\alpha^4 + \alpha + 1 = 0$ . Then, the finite field  $\mathbb{F}_2(\alpha)$ , constructed by extending the binary field with the algebraic element  $\alpha$ , is identified by  $\mathbb{F}_{16}$ . Let  $f(x) = A + R(x - \alpha)$  and  $g(x) = B + S(x - \alpha)$ , where  $R$  and  $S$  are the same random matrices used in the traditional polynomial scheme above. Then,  $h(x) = f(x) \cdot g(x)$  is such that  $h(\alpha) = AB$ .<sup>1</sup>

Our scheme, illustrated in Figure 1, works as follows. First, the user uploads the evaluations  $f(y_i)$  and  $g(y_i)$  to each Server  $i$ . Then, each Server  $i$  computes  $\text{tr}(\alpha^{-j_i} h(y_i))$ , where  $(j_1, j_2, j_3, j_4) = (1, 2, 8, 4)$  and  $(y_1, y_2, y_3, y_4) = (0, \alpha^5, \alpha^{10}, \alpha^{15})$ , and sends it back to the user.<sup>2</sup> To show that the user is able to decode  $AB$ , we denote the answer of each Server  $i$  by  $S_i$ . Then,

$$\begin{aligned} & \alpha^4(S_1 + S_2 + S_3 + S_4) + \alpha^5 S_2 + \alpha^{10} S_3 + \alpha^{15} S_4 \\ &= \alpha^4 \text{tr}(\alpha^{-1} h(0)) + \alpha^{-2} h(\alpha^5) + \alpha^{-8} h(\alpha^{10}) + \alpha^{-4} h(\alpha^{15}) \\ & \quad + \text{tr}(\alpha^3 h(\alpha^5) + \alpha^2 h(\alpha^{10}) + \alpha^{11} h(\alpha^{15})) \\ &= \alpha^4 \text{tr}(h(\alpha)) + \text{tr}(\alpha h(\alpha)) = h(\alpha) = AB \end{aligned}$$

The first equality follows from the  $\mathbb{F}_4$ -linearity of the field trace together with the fact that  $\alpha^5, \alpha^{10}, \alpha^{15} \in \mathbb{F}_4$ . The second equality follows from utilizing the equation  $h(x) = h_0 + h_1 x + h_2 x^2$  and the facts that  $\alpha^{15} = 1$  and  $\alpha^4 + \alpha + 1 = 0$ . The third equality follows from utilizing the equation  $\text{tr}(x) = x + x^4$  and the fact that  $\alpha^4 + \alpha + 1 = 0$ .

Security follows by showing that  $I(f(y_i), g(y_i); A, B) = 0$ , as is done in Lemma 2. As for the communication costs, first the user uploads  $f(y_i)$  and  $g(y_i)$ , which cost  $2ab$  and  $2bc$ , symbols respectively, four times. Thus, the upload cost is  $4(2ab + 2bc)$  symbols of  $\mathbb{F}_4$ . Then, the user downloads  $\text{tr}(\alpha^{-j_i} h(y_i))$ , which costs  $ac$  symbols of  $\mathbb{F}_{16}$ , four times, obtaining a download cost of  $4ac$  symbols of  $\mathbb{F}_4$ . Since the user retrieves  $AB \in \mathbb{F}_{16}^{a \times c}$ , which consists of  $2ac$  symbols of  $\mathbb{F}_4$ , the total communication rate is given by  $\mathcal{R} = \frac{ac}{4ab+4bc+2ac}$ .

We note that the download cost of the FTP code is lower than that of the traditional polynomial code, and that the opposite is true for the upload cost. In terms of total communication, the FTP code outperforms the traditional one whenever  $\mathcal{R} > \mathcal{R}'$ . This occurs whenever the matrix dimensions satisfy the inequality  $b \left( \frac{1}{a} + \frac{1}{c} \right) < 1$ .

<sup>1</sup>Unlike in the repairing Reed-Solomon codes setting, in the SDMM setting, we have more freedom to design the code and choose the evaluation points.

<sup>2</sup>Here, we denote the field trace by  $\text{tr} := \text{tr}_{\mathbb{F}_{16}/\mathbb{F}_4} : \mathbb{F}_{16} \rightarrow \mathbb{F}_4$  given by  $\text{tr}(x) = x + x^4$ . We apply the trace function element-wise on matrices which is equivalent to an element-wise exponentiation.

We note that this FTP code could also outperform any traditional polynomial scheme with  $N' = 2$  servers, if such a scheme existed. Indeed, the only way for a traditional polynomial scheme to outperform the FTP code for all matrix dimensions would be for it to use  $N' = 1$  server, which is not possible because of the 1-security constraint.

#### IV. FIELD TRACE POLYNOMIAL CODES

In this section, we present the general construction for FTP codes. The main idea is to perform the same technique as in Section III,  $L$  times, each one retrieving  $A_i B_i$ , but doing so utilizing a single polynomial  $h(x) = f(x) \cdot g(x)$ .

**Choosing the Field:** We begin by choosing the field over which we operate. Let  $L \in \mathbb{Z}^+$ ,  $\{p_1, \dots, p_L\}$  be a set of distinct prime numbers in increasing order,  $q_0$  a prime power, and set  $q = q_0^{p_1 p_2 \dots p_L}$ . We then operate over  $\mathbb{F}_q$ . For  $i \in [L]$ , we let  $\alpha_i \in \mathbb{F}_{q_0}$  be such that  $\mathbb{F}_{q_0}(\alpha_i)$  is a field extension of  $\mathbb{F}_{q_0}$  of order  $p_i$ . And thus,  $\mathbb{F}_q = \mathbb{F}_{q_0}(\alpha_1, \alpha_2, \dots, \alpha_L)$ . We also define  $F_i = \mathbb{F}_{q_0}(\alpha_j : 1 \leq j \leq L \text{ and } j \neq i)$ .

**Choosing the Polynomials:** As described in the introduction, we consider the setting where the user partitions the matrices  $A \in \mathbb{F}_q^{a \times b}$  and  $B \in \mathbb{F}_q^{b \times c}$  as  $A = [A_1 \dots A_L]$  and as  $B^\top = [B_1^\top \dots B_L^\top]$  such that  $AB = A_1 B_1 + \dots + A_L B_L$ , where each  $A_i \in \mathbb{F}_q^{a \times \frac{b}{L}}$  and  $B_i \in \mathbb{F}_q^{\frac{b}{L} \times c}$ . In order to obtain  $T$ -security  $R_1, \dots, R_T \in \mathbb{F}_q^{a \times \frac{b}{L}}$  and  $S_1, \dots, S_T \in \mathbb{F}_q^{\frac{b}{L} \times c}$  are chosen independently and uniformly at random. We then define  $f, g \in \mathbb{F}_q[x]$  as the polynomials of degree  $L + T - 1$  such that,  $f(\alpha_i) = A_i$ ,  $g(\alpha_i) = B_i$ , for every  $i \in [L]$ , and  $f(\alpha_{L+i}) = R_j$ ,  $g(\alpha_{L+i}) = S_j$ , for every  $j \in [T]$ .

**Choosing the Evaluation Points:** For each  $i \in [L]$ , denote  $N_i = p_i + 2L + 2T - 2$  and  $n = N_L + L = p_L + 3L + 2T - 2$ . Consider the set  $\{\alpha_1, \alpha_2, \dots, \alpha_L\}$  of primitive elements defined above. Let  $\alpha_{L+1}, \dots, \alpha_n \in \mathbb{F}_{q_0}$  be distinct elements which are also distinct from  $\{\alpha_1, \alpha_2, \dots, \alpha_L\}$ . We then define  $\Omega = \{\alpha_1, \dots, \alpha_n\}$ . The evaluation points the user sends are those in the set  $\{\alpha_{L+1}, \dots, \alpha_n\}$  of size  $N_L$ .<sup>3</sup>

**Upload Phase:** The FTP code uses  $N_L$  servers. The user uploads  $f(\alpha_{L+i})$  and  $g(\alpha_{L+i})$  to each Server  $i$ .

**Download Phase:** Let  $v_j = \prod_{\substack{1 \leq i \leq n \\ i \neq j}} (\alpha_j - \alpha_i)^{-1}$  and  $k_i(x)$  be the annihilator polynomial for  $\{\alpha_j : j \in [n] \setminus ([L+1 : N_i + L] \cup \{i\})\}$ . Then, for each  $i \in [L]$ , Server  $j$  computes  $\text{tr}_{\mathbb{F}_q/F_i}(v_{L+j} k_i(\alpha_{L+j}) h(\alpha_{L+j}))$  and sends these  $F_i$ -values to the user.

**User Decoding:** In Lemma 1, we show that the user is able to retrieve  $h(\alpha_i) = A_i B_i$  from  $\{\text{tr}_{\mathbb{F}_q/F_i}(v_j k_i(\alpha_j) h(\alpha_j)) : j \in [L+1 : N_i + L]\}$ . Combining these, the user can decode  $AB = A_1 B_1 + \dots + A_L B_L$ .

#### V. PROOF OF THEOREM 1

We break the proof into different Lemmas. We show that FTP codes are decodable, in Lemma 1,  $T$ -secure, in Lemma 2, and characterize their performance, in Lemma 3. These statements combined prove Theorem 1.

<sup>3</sup>Thus,  $q_0 \geq N_L$  is required.

**Lemma 1.** *Given positive integers  $L$  and  $T$ , let  $p_1, p_2, \dots, p_L$  be distinct prime numbers in the ascending order and  $q_0$  be a prime power with  $q_0 \geq N_L$ . Let  $A = [A_1 \dots A_L] \in \mathbb{F}_q^{a \times b}$  and  $B^\top = [B_1^\top \dots B_L^\top] \in \mathbb{F}_q^{c \times b}$ , where  $q = q_0^{p_1 p_2 \dots p_L}$ . Then,  $h(\alpha_i)$  can be decoded using  $N_i$  servers, for  $i \in [L]$ .*

*Proof.* Let  $f(x), g(x) \in \mathbb{F}_q[x]$  be polynomials such that

$$\begin{aligned} f(\alpha_i) &= A_i, & g(\alpha_i) &= B_i & \text{for } i \in [L] \text{ and} \\ f(\alpha_{L+i}) &= R_i, & g(\alpha_{L+i}) &= S_i & \text{for } i \in [T], \end{aligned}$$

using the inner product partitioning  $A = [A_1 \dots A_L]$  and  $B^\top = [B_1^\top \dots B_L^\top]$  and uniformly distributed random  $\mathbb{F}_q$ -matrices  $R_i, S_i$ . Therefore,  $h(x) = f(x)g(x)$  is a polynomial of degree  $2L + 2T - 2$  such that  $h(\alpha_i) = A_i B_i$ , for  $i \in [L]$ .

Let  $\mathcal{C} = RS_{\mathbb{F}_q}(n, 2L + 2T - 1, \Omega)$ . Since the degree of  $h$  is smaller than  $2L + 2T - 1$ , the vector  $(h(\alpha_1), h(\alpha_2), \dots, h(\alpha_n))$  is in  $\mathcal{C}$ .

Define, for  $i \in [L]$ ,  $U_i = \{L + 1, \dots, L + N_i\}$ , which has  $N_i$  elements. For each  $i \in [L]$ , let

$$k_i(x) = \prod_{j \in [n] \setminus (U_i \cup \{i\})} (x - \alpha_j).$$

Note that  $\deg(k_i(x)) = p_L + L - 1 - p_i$ ,

and so  $\deg(k_i(x)x^s) < p_L + L - 1$  for  $s = 0, \dots, p_i - 1$ .

Using  $V$  as in Definition 2, it follows that the Generalized Reed-Solomon with parameter  $n$ ,  $p_L + L - 1$ ,  $\Omega$  and  $V$  contains the element  $(v_1 k_i(\alpha_1) \alpha_1^s, \dots, v_n k_i(\alpha_n) \alpha_n^s)$ , for  $s = 0, \dots, p_i - 1$ , i.e.,

$$\begin{aligned} & (v_1 k_i(\alpha_1) \alpha_1^s, \dots, v_n k_i(\alpha_n) \alpha_n^s) \\ & \in \mathcal{C}^\perp = GRS_{\mathbb{F}_q}(n, p_L + L - 1, \Omega, V). \end{aligned}$$

The dual-code property implies that

$$\sum_{j=1}^n v_j k_i(\alpha_j) \alpha_j^s h(\alpha_j) = 0.$$

For each  $i \in [L]$ , we have  $k_i(\alpha_j) = 0$  if  $j \notin U_i \cup \{i\}$ , hence

$$v_i k_i(\alpha_i) \alpha_i^s h(\alpha_i) = - \sum_{j \in U_i} v_j k_i(\alpha_j) \alpha_j^s h(\alpha_j).$$

Applying  $\text{tr}_i := \text{tr}_{\mathbb{F}_q/F_i}$  to both sides yields, using the  $F_i$ -linearity and  $\alpha_j \in F_i$  for  $j \in U_i$ ,

$$\text{tr}_i(v_i k_i(\alpha_i) \alpha_i^s h(\alpha_i)) = - \sum_{j \in U_i} \alpha_j^s \text{tr}_i(v_j k_i(\alpha_j) h(\alpha_j)).$$

Let  $\{\lambda_{s,i} = v_i k_i(\alpha_i) \alpha_i^s : 0 \leq s < p_i\}$ . Since  $v_i k_i(\alpha_i) \neq 0$  and  $\alpha_i$  is a primitive element, the set  $\{\lambda_{s,i}\}$  is an  $F_i$ -basis of  $\mathbb{E}$ . Further, there exists a set  $\{\mu_{s,i} : 0 \leq s < p_i\}$  which is the trace-dual  $\mathbb{F}_{q_0}$ -basis to  $\{\lambda_{s,i}\}$  of  $\mathbb{F}_{q_0}(\alpha_i)$ . Thus,

$$\sum_{s=0}^{p_i-1} \text{tr}_i(\lambda_{s,i} h(\alpha_i)) \mu_{s,i} = h(\alpha_i) = A_i B_i.$$

□

Decodability is then obtained by repeating the process given in Lemma 1 and summing it over  $i \in [L]$ ,

$$\sum_{i=1}^L \sum_{s=0}^{p_i-1} \text{tr}_i(v_i k_i(\alpha_i) \alpha_i^s h(\alpha_i)) \mu_{s,i} = AB.$$

Next, we show that FTP codes are  $T$ -secure.

**Lemma 2.** *FTP codes are  $T$ -secure.*

*Proof.* In extended version. Follows from standard arguments.  $\square$

We now characterize the total communication.

**Lemma 3.** *FTP codes have total communication rate*

$$\mathcal{R} = \left( \frac{N_L b}{L} \left( \frac{1}{a} + \frac{1}{c} \right) + \sum_{i=1}^L \frac{N_i}{p_i} \right)^{-1}.$$

*Proof.* In extended version. Follows from a direct calculation.  $\square$

## VI. EXAMPLE: $L = 3$ AND $T = 2$

We present an example of an FTP code for  $L = 3$  and  $T = 2$  and compare it to the current state of the art [11].

**Choosing the field:** We begin by choosing a set of three prime numbers  $\{5, 7, 11\}$ ,  $q_0 = 27$  a prime power, and set  $27^{5 \cdot 7 \cdot 11} = 27^{385}$ . Thus, we operate over  $\mathbb{F}_{27^{385}}$ . Let  $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{F}_{27}$  be such that  $\mathbb{F}_{27}(\alpha_1)$ ,  $\mathbb{F}_{27}(\alpha_2)$ , and  $\mathbb{F}_{27}(\alpha_3)$  are field extensions of  $\mathbb{F}_{27}$  of degrees 5, 7, and 11, respectively. Therefore,  $\mathbb{F}_{27^{385}} = \mathbb{F}_{27}(\alpha_1, \alpha_2, \dots, \alpha_L)$ . We also define  $F_1 = \mathbb{F}_{27}(\alpha_2, \alpha_3) = \mathbb{F}_{27}^{77}$ ,  $F_2 = \mathbb{F}_{27}(\alpha_1, \alpha_3) = \mathbb{F}_{27}^{55}$ , and  $F_3 = \mathbb{F}_{27}(\alpha_1, \alpha_2) = \mathbb{F}_{27}^{35}$ .

**Choosing the polynomials:** Since  $L = 3$ , consider the setting where the user partitions the matrices  $A \in \mathbb{F}_{27^{385}}^{a \times b}$  and  $B \in \mathbb{F}_{27^{385}}^{b \times c}$  as  $A = [A_1 \ A_2 \ A_3]$  and  $B^\top = [B_1^\top \ B_2^\top \ B_3^\top]$  such that  $AB = A_1 B_1 + A_2 B_2 + A_3 B_3$ , where each  $A_i \in \mathbb{F}_{27^{385}}^{a \times \frac{b}{3}}$  and  $B_i \in \mathbb{F}_{27^{385}}^{\frac{b}{3} \times c}$ . In order to obtain 2-security we choose  $R_1, R_2 \in \mathbb{F}_{27^{385}}^{a \times \frac{b}{3}}$  and  $S_1, S_2 \in \mathbb{F}_{27^{385}}^{\frac{b}{3} \times c}$  all independently and uniformly at random. We then define  $f, g \in \mathbb{F}_{27^{385}}[x]$  as the polynomials of degree 3 such that,  $f(\alpha_i) = A_i$ ,  $g(\alpha_i) = B_i$ , for every  $i \in [3]$ , and  $f(\alpha_{3+j}) = R_j$ ,  $g(\alpha_{3+j}) = S_j$ , for  $j \in [2]$ . Hence,  $h(x) = f(x) \cdot g(x)$  is a polynomial of degree 8 such that  $h(\alpha_i) = A_i B_i$  for every  $i \in [3]$ .

**Choosing the Evaluation Points:** Let  $N_1 = 13$ ,  $N_2 = 15$ ,  $N_3 = 19$ , and  $n = 22$ . Consider the set  $\{\alpha_1, \alpha_2, \alpha_3\}$  of primitive elements defined above. Let  $\alpha_4, \dots, \alpha_{22} \in \mathbb{F}_{27}$  be distinct elements which are not in  $\{\alpha_1, \alpha_2, \alpha_3\}$ . Define  $\Omega = \{\alpha_1, \dots, \alpha_{22}\}$ . The evaluation points the user sends are those in the set  $\{\alpha_4, \dots, \alpha_{22}\}$  of size 19.

**Upload Phase:** The FTP code uses 19 servers. The user uploads  $f(\alpha_{3+i})$  and  $g(\alpha_{3+i})$  to each Server  $i$ .

**Download Phase:** Let  $v_j = \prod_{\substack{1 \leq i \leq 22 \\ i \neq j}} (\alpha_j - \alpha_i)^{-1}$ . Let  $U_1 = \{4, 5, \dots, 16\}$ ,  $U_2 = \{4, 5, \dots, 18\}$  and  $U_3 = \{4, \dots, 22\}$  be the index sets of cardinality 13, 15, and 19, respectively. For  $i \in [3]$ , define  $k_i(x)$  as the annihilator polynomial for

$\{\alpha_j : j \in [22] \setminus ([4 : N_i + 3] \cup \{i\})\}$ , i.e.,  $k_i(x) = \prod_{j \in [22] \setminus ([4 : N_i + 3] \cup \{i\})} (x - \alpha_j)$ .

Defining  $V$  as in Definition 2 for the dual code, the Server  $j$ , for each  $i \in [3]$ , computes  $\text{tr}_i(v_{3+j} k_i(\alpha_{3+j}) h(\alpha_{3+j}))$  and sends these  $F_i$ -values to the user.

**User Decoding:** In Lemma 1, we show that the user is able to retrieve  $h(\alpha_i) = A_i B_i$  from  $\{\text{tr}_i(v_j k_i(\alpha_j) h(\alpha_j)) : j \in [4 : N_i + 3]\}$ . Because  $V$  is constructed as in Definition 2 for the dual code, it follows that

$$\text{tr}_i(v_1 k_i(\alpha_i) \alpha_i^s h(\alpha_i)) = - \sum_{j \in U_i} \alpha_j^s \text{tr}_i(v_j k_i(\alpha_j) h(\alpha_j)),$$

where the trace dual is  $\text{tr}_i = \text{tr}_{\mathbb{F}_{27^{385}}/\mathbb{F}_{27}(\alpha_j : j \in [3] \setminus \{i\})}$ . Hence, given  $\{\text{tr}_i(v_j k_i(\alpha_j) h(\alpha_j)) : j \in U_i\}$ , the user can compute  $\text{tr}_i(v_1 k_i(\alpha_i) \alpha_i^s h(\alpha_i))$ . Using trace-dual basis elements  $\mu_{i,s}$ , for  $i \in [3]$ , we have  $h(\alpha_i) = \sum_{s=0}^{N_i-1} \text{tr}_i(v_i k_i(\alpha_i) \alpha_i^s h(\alpha_i)) \mu_{i,s}$ . Summing up the results, the user can then decode the product  $AB = h(\alpha_1) + h(\alpha_2) + h(\alpha_3) = A_1 B_1 + A_2 B_2 + A_3 B_3$ .

**Communication Costs:** Server  $i$  has  $h(\alpha_i)$  as in terms of symbols of  $\mathbb{F}_{27^{385}}$ , but transmits only  $\text{tr}_i(v_j k_i(\alpha_j) h(\alpha_j))$ , symbols of  $F_i$  which is one symbol of  $F_i$ . To retrieve  $h(\alpha_i)$ , only a  $p_i$ -th part of the information in each of  $|U_i|$  servers is needed. Hence, to determine  $h(\alpha_i)$ , it is enough to download an  $x$ -th part of each of the  $y$  servers in  $U_i$  for  $(i, x, y) \in \{(1, 5, 13), (2, 7, 15), (3, 11, 19)\}$ .

In conclusion, the FTP code used to compute the product  $AB$ , with partitioning parameters  $L = 3$  and security parameter  $T = 2$ , uses  $N_3 = 19$  servers and achieves a download rate of  $\frac{1}{6.47}$ . In comparison, the state of the art traditional polynomial code in [11] has a download rate equal to  $\frac{1}{7}$ . The scheme in [11] has a total communication rate equal to  $(\frac{7}{3} \cdot b(\frac{1}{a} + \frac{1}{c}) + 7)^{-1}$ . The total communication rate of the above FTP code is  $(\frac{19}{3} \cdot b(\frac{1}{a} + \frac{1}{c}) + \frac{2491}{385})^{-1}$ . A direct calculation shows that the FTP code outperforms the scheme in [11] for matrices with dimensions such that  $b(\frac{1}{a} + \frac{1}{c}) < \frac{51}{385}$ .

## VII. PROOF OF THEOREM 2

We begin by proving a technical lemma.

**Lemma 4.** *Let  $T, N, N', L, L'$  be positive integers,  $\lambda, \eta \geq 0$  be real numbers,  $p_i \geq 2L(L + T - 1)\eta$  be primes, for every  $i \in [L]$ , such that  $p_i < p_{i+1}$  and  $p_L > LN'/L' - 2L - 2T + 2$ ,  $N_i = p_i + 2L + 2T - 2$ , and  $\lambda < \frac{N' - L - 1/\eta}{N_L/L - N'/L'}$ . Then,*

$$\left( \frac{N_L \lambda}{L} + \sum_{i=1}^L \frac{N_i}{p_i} \right)^{-1} > \left( N' \left( \frac{\lambda}{L'} + 1 \right) \right)^{-1}. \quad (3)$$

*Proof.* In extended version. Follows from a direct calculation.  $\square$

Consider a traditional polynomial scheme with partitioning parameter  $L$  and security parameter  $T$ . Given the trivial lower bound  $N' > L$  on its recovery threshold  $N'$ , it follows from a direct calculation that the total communication rate  $\mathcal{R}'$  of the traditional polynomial scheme is upper bounded by

$$\mathcal{R}' < \left( N' \left( \frac{b}{L} \left( \frac{1}{a} + \frac{1}{c} \right) + 1 \right) \right)^{-1}. \quad (4)$$

Let  $\lambda = b(\frac{1}{a} + \frac{1}{c})$ ,  $\mu > (N' - L)^{-1}$  and select  $L$  primes  $p_1, \dots, p_L$  in increasing order and such that  $p_i \geq 2L(L + T - 1)\eta$  and  $p_L > LN'/L - 2L - 2T + 2$ . Next, define  $N_i = p_i + 2L + 2T - 2$  and set

$$K = \frac{N' - L - 1/\eta}{N_L/L - N'/L}. \quad (5)$$

Then, by Theorem 1, the communication rate  $\mathcal{R}$  of the FTP code with these parameters, given by (1), is equal to the left hand side of (3). For these same parameters, the right hand sides of (3) and (4) are the same. Thus, if  $b(\frac{1}{a} + \frac{1}{c}) < K$ , it follows from Lemma 4 that the communication rate of the FTP code  $\mathcal{R}$  is larger than the right hand side of (3), and is therefore larger than the communication rate of the traditional polynomial scheme  $\mathcal{R}'$ .

#### ACKNOWLEDGMENT

Rafael D'Oliveira was supported by MIT Portugal Program (Project SNOB5G with Nr. 045929 [CENTRO-01-0247-FEDER-045929]), MIT Lincoln Laboratory Purchase Order 7000500173, and the FinTech@CSAIL Research Initiative. Salim El Rouayheb was partially supported by the NSF under grant CNS-1801630. Daniel Heinlein is supported by the Academy of Finland, Grant 331044.

#### REFERENCES

- [1] W.-T. Chang and R. Tandon, "On the capacity of secure distributed matrix multiplication," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.
- [2] J. Kakar, S. Ebadifar, and A. Sezgin, "On the capacity and straggler-robustness of distributed secure matrix multiplication," *IEEE Access*, vol. 7, pp. 45 783–45 799, 2019.
- [3] H. Yang and J. Lee, "Secure distributed computing with straggling servers using polynomial codes," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 1, pp. 141–150, 2018.
- [4] R. G. L. D'Oliveira, S. El Rouayheb, and D. Karpuk, "Gasp codes for secure distributed matrix multiplication," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 1107–1111.
- [5] R. G. L. D'Oliveira, S. El Rouayheb, D. Heinlein, and D. Karpuk, "Degree tables for secure distributed matrix multiplication," in *2019 IEEE Information Theory Workshop (ITW)*, 2019.
- [6] M. Aliasgari, O. Simeone, and J. Kliewer, "Distributed and private coded matrix computation with flexible communication load," *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 1092–1096, 2019.
- [7] —, "Private and secure distributed matrix multiplication with flexible communication load," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2722–2734, 2020.
- [8] J. Kakar, A. Khristoforov, S. Ebadifar, and A. Sezgin, "Uplink-downlink tradeoff in secure distributed matrix multiplication," *ArXiv*, vol. abs/1910.13849, 2019.
- [9] R. G. L. D'Oliveira, S. E. Rouayheb, D. Heinlein, and D. Karpuk, "Notes on communication and computation in secure distributed matrix multiplication," in *2020 IEEE Conference on Communications and Network Security (CNS)*, 2020, pp. 1–6.
- [10] Q. Yu and A. S. Avestimehr, "Entangled polynomial codes for secure, private, and batch distributed matrix multiplication: Breaking the "cubic" barrier," *ArXiv*, vol. abs/2001.05101, 2020.
- [11] N. Mital, C. Ling, and D. Gunduz, "Secure distributed matrix computation with discrete fourier transform," *arXiv preprint arXiv:2007.03972*, 2020.
- [12] R. Bitar, M. Xhemrishi, and A. Wachter-Zeh, "Adaptive private distributed matrix multiplication," *arXiv preprint arXiv:2101.05681*, 2021.
- [13] B. Hasircioglu, J. Gomez-Vilardebo, and D. Gunduz, "Speeding up private distributed matrix multiplication via bivariate polynomial codes," *arXiv preprint arXiv:2102.08304*, 2021.
- [14] R. G. L. D'Oliveira, S. El Rouayheb, and D. Karpuk, "Gasp codes for secure distributed matrix multiplication," *IEEE Transactions on Information Theory*, pp. 1–1, 2020.
- [15] V. Guruswami and M. Wootters, "Repairing Reed-Solomon codes," *IEEE Transactions on Information Theory*, vol. 63, no. 9, pp. 5684–5698, 2017.
- [16] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, "A survey on network codes for distributed storage," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 476–489, 2011.
- [17] Q. Yu, M. Maddah-Ali, and A. S. Avestimehr, "Polynomial codes: an optimal design for high-dimensional coded matrix multiplication," in *Advances in Neural Information Processing Systems*, 2017, pp. 4403–4413.
- [18] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "Straggler mitigation in distributed matrix multiplication: Fundamental limits and optimal coding," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2022–2026.
- [19] S. Dutta, M. Fahim, F. Haddadpour, H. Jeong, V. Cadambe, and P. Grover, "On the optimal recovery threshold of coded matrix multiplication," *IEEE Transactions on Information Theory*, 2019.
- [20] U. Sheth, S. Dutta, M. Chaudhari, H. Jeong, Y. Yang, J. Kohonen, T. Roos, and P. Grover, "An application of storage-optimal matdot codes for coded matrix multiplication: Fast k-nearest neighbors estimation," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 1113–1120.
- [21] S. Li, M. A. Maddah-Ali, Q. Yu, and A. S. Avestimehr, "A fundamental tradeoff between computation and communication in distributed computing," *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 109–128, 2017.
- [22] H. A. Nodehi and M. A. Maddah-Ali, "Limited-sharing multi-party computation for massive matrix operations," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 1231–1235.
- [23] Z. Jia and S. A. Jafar, "On the capacity of secure distributed matrix multiplication," *arXiv preprint arXiv:1908.06957*, 2019.
- [24] H. Akbari-Nodehi and M. A. Maddah-Ali, "Secure coded multi-party computation for massive matrix operations," *IEEE Transactions on Information Theory*, vol. 67, no. 4, pp. 2379–2398, 2021.
- [25] M. Kim, H. Yang, and J. Lee, "Private coded matrix multiplication," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1434–1443, 2019.
- [26] J. Zhu, Q. Yan, and X. Tang, "Improved constructions for secure multi-party batch matrix multiplication," *arXiv preprint arXiv:2103.09592*, 2021.
- [27] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE transactions on information theory*, vol. 56, no. 9, pp. 4539–4551, 2010.
- [28] K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal exact-regenerating codes for distributed storage at the msr and mbr points via a product-matrix construction," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5227–5239, 2011.
- [29] V. R. Cadambe, S. A. Jafar, H. Maleki, K. Ramchandran, and C. Suh, "Asymptotic interference alignment for optimal repair of mds codes in distributed storage," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2974–2987, 2013.
- [30] I. Tamo, M. Ye, and A. Barg, "Optimal repair of reed-solomon codes: Achieving the cut-set bound," in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2017, pp. 216–227.
- [31] R. Bitar, P. Parag, and S. El Rouayheb, "Minimizing latency for secure coded computing using secret sharing via staircase codes," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4609–4619, 2020.
- [32] A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath, "Centralized repair of multiple node failures with applications to communication efficient secret sharing," *IEEE Transactions on Information Theory*, vol. 64, no. 12, pp. 7529–7550, 2018.
- [33] W. Huang and J. Bruck, "Secret sharing with optimal decoding and repair bandwidth," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 1813–1817.
- [34] W. C. Huffman, J.-L. Kim, and P. Solé, *Concise Encyclopedia of Coding Theory*. Chapman and Hall/CRC, 2021.