Minimizing Latency for Secure Coded Computing Using Secret Sharing via Staircase Codes

Rawad Bitar, *Member, IEEE*, Parimal Parag, *Member, IEEE*, and Salim El Rouayheb, *Member, IEEE*

***Abstract*— We consider the setting of a Master server, M, who possesses confidential data and wants to run intensive computations on it, as part of a machine learning algorithm for example. The Master wants to distribute these computations to untrusted workers who volunteered to help with this task. However, the data must be kept private in an information theoretic sense. Some of the workers may be stragglers, e.g., slow or busy. We are interested in reducing the delays experienced by the Master. We focus on linear computations as an essential operation in many iterative algorithms. We propose a solution based on new codes, called Staircase codes, introduced previously by two of the authors. Staircase codes allow flexibility in the number of stragglers up to a given maximum, and universally achieve the information theoretic limit on the download cost by the Master, leading to latency reduction. We find upper and lower bounds on the Master's mean waiting time. We derive the distribution of the Master's waiting time, and its mean, for systems with up to two stragglers. We show that Staircase codes always outperform existing solutions based on classical secret sharing codes. We validate our results with extensive implementation on Amazon EC2.**

***Index Terms*— Distributed computing, data privacy, secret sharing, secure coded computing, machine learning.**

I. Introduction

We consider the setting of distributed computing in which a server M, referred to as Master, possesses *confidential* data (e.g., personal, genomic or medical data) and wants to perform intensive computations on it. M wants to divide these computations into smaller computational tasks and distribute them to n *untrusted* worker machines that can perform these smaller tasks in parallel. The workers then return their results to the Master, who can process them to obtain the result of its original task.

In this paper, we are interested in applications in which the worker machines do not belong to the same system or cluster as the Master. Rather, the workers are online computing machines that can be hired or can volunteer to help the Master in its computations, e.g., crowdsourcing platforms like the SETI@home [3] and folding@home [4] projects. The additional constraint, which we worry about here, is that the workers cannot be trusted with the sensitive data, which must remain hidden from them. Privacy could be achieved using fully homomorphic encryption that allows computing over encrypted data. However, homomorphic encryption incurs high computation and storage overheads [5], which may not be feasible in certain applications.

We propose information theoretic security to achieve the privacy requirement. Information theoretic security is typically used to provide privacy with no constraints on the computational power of the adversary (compromised workers). Our main motivation for information theoretic security is the low complexity of the resulting schemes (compared to homomorphic encryption). The assumption that we have to make here is a limit on the number of workers colluding against the Master.

We focus on linear computations (matrix multiplication) since they form a building block of many iterative algorithms, such as principal component analysis, support vector machines and other gradient-descent based algorithms [6], [7]. The workers introduce random delays due to the difference of their workloads or network congestion. This causes the Master to wait for the slowest workers, referred to as stragglers in the distributed computing community [8], [9]. Our goal is to reduce the aggregate delay experienced by the Master.

Privacy can be achieved by encoding the data, with random keys, using linear secret sharing codes [10] as illustrated in Example 1. However, these codes are not specifically designed to minimize latency as we will highlight later.

Example 1: Let the matrix A denote the data set owned by M and let $\mathbf{x}$ be a given vector. M wants to compute $A\mathbf{x}$. Suppose that M gets the help of 3 workers out of which at most 1 may be a straggler. M generates a random matrix R of same dimensions as A with entries drawn over the same alphabet as the entries of A. M encodes A and R into 3 shares $S_1 = R$, $S_2 = R + A$ and $S_3 = R + 2A$ using a secret sharing

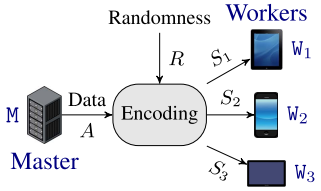
Manuscript received September 9, 2019; revised February 6, 2020; accepted April 6, 2020. Date of publication April 17, 2020; date of current version August 14, 2020. The work of the first and last authors was supported in part by NSF Grants CCF 1817635 and CNS 1801630 and by ARL Grant W911NF-17-1-0032. The work of the second author was supported in part by the Science and Engineering Research Board (SERB) under Grant No. DSTO-1677, the Department of Telecommunications, Government of India, under Grant DOTC-0001, the Robert Bosch Center for Cyber-Physical Systems, the Centre for Networked Intelligence (a Cisco CSR initiative) of the Indian Institute of Science, Bangalore. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies. This article was presented in part at the 2017 IEEE International Symposium on Information Theory (ISIT). The associate editor coordinating the review of this article and approving it for publication was G. Durisi. *(Corresponding author: Rawad Bitar.)*

Rawad Bitar and Salim El Rouayheb are with the ECE Department, Rutgers University, New Jersey, NJ 08854 USA (e-mail: rawad.bitar@rutgers.edu; salim.elrouayheb@rutgers.edu).

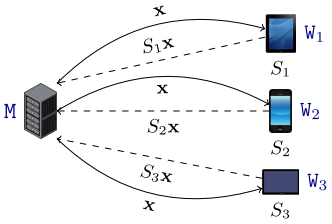
Parimal Parag is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru 560012, India (e-mail: parimal@iisc.ac.in).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2020.2988506



(a) The Master M encodes its data A with a random matrix R into 3 secret shares $S_1,\ S_2,\ S_3$. Any two shares can decode A. For example, $S_1 = R$, $S_2 = A + R$, and $S_3 = A + 2R$. M sends the share S_i to worker $\mathtt{W}_i$. The randomness R is used to ensure privacy.



(b) To compute $A\mathbf{x}$, M sends $\mathbf{x}$ to all the workers. Each worker $\mathtt{W}_i$ computes $S_i\mathbf{x}$ and sends the result to M. M can decode $A\mathbf{x}$ after receiving any two responses, e.g., $A\mathbf{x} = S_2\mathbf{x} - S_1\mathbf{x} = (A + R)\mathbf{x} - R\mathbf{x}$.

Fig. 1. Secure distributed matrix multiplication with 3 workers. The Master encodes its data using a linear secret sharing code, e.g., Shamir's codes (given in the caption) [11], [12] or Staircase codes (given in Table I) [13], [14]. Decoding $A\mathbf{x}$ follows from the linearity of the code.

TABLE I

THE SHARES SENT BY M TO EACH WORKER WHEN USING STAIRCASE CODES. IN THIS EXAMPLE, EACH SHARE IS DIVIDED INTO TWO SUB-SHARES. THE OPERATIONS SHOWN ARE IN $GF(5)$

S_1	S_2	S_3
$A_1 + A_2 + R_1$	$A_1 + 2A_2 + 4R_1$	$A_1 + 3A_2 + 4R_1$
$R_1 + R_2$	$R_1 + 2R_2$	$R_1 + 3R_2$

scheme [11], [12]. M sends share S_i to worker $\mathtt{W}_i$ (Figure 1a) and then sends $\mathbf{x}$ to all the workers. Each worker computes $S_i\mathbf{x}$ and sends it back to M (Figure 1b). M can decode $A\mathbf{x}$ after receiving any 2 responses. For instance, if the first two workers respond, M can obtain $A\mathbf{x} = S_2\mathbf{x} - S_1\mathbf{x}$. No information about A is revealed to the workers, because A is one-time padded by R.

In the previous example, even if there were no stragglers, M still has to wait for the full responses of two workers, and the response of the third one will not be used for decoding. In addition, M always has to decode $R\mathbf{x}$ in order to decode $A\mathbf{x}$. Hence, more delays are incurred by spending communication and computation resources on decoding $R\mathbf{x}$, which is only needed for privacy. We overcome those limitations by using Staircase codes introduced in [13], [14] which do not always require decoding $R\mathbf{x}$. Thus, possibly reducing the computation load at the workers and the communication cost at the Master. In addition, Staircase codes allow more flexibility in the number of responses needed for decoding $A\mathbf{x}$, as explained in the next example.

Example 2 (Staircase codes): Consider the same setting as Example 1. Instead of using a classical secret sharing code, M now encodes A and R using the Staircase code given in Table I. The Staircase code requires M to divide the matrices A and R into $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}^T$ and $R = \begin{bmatrix} R_1 & R_2 \end{bmatrix}^T$. In this setting, M sends two sub-shares to each worker, hence each task consists of 2 sub-tasks. The Master sends $\mathbf{x}$ to all the workers. Each worker multiplies the sub-shares by $\mathbf{x}$ (going from top to bottom) and sends each multiplication back to M independently. Now, M has two possibilities for decoding: 1) M receives the first sub-task from all the workers, i.e., receives $(A_1 + A_2 + R_1)\mathbf{x}$, $(A_1 + 2A_2 + 4R_1)\mathbf{x}$ and $(A_1 + 3A_2 + 4R_1)\mathbf{x}$ and decodes $A\mathbf{x}$ which is the concatenation of $A_1\mathbf{x}$ and $A_2\mathbf{x}$. Note that here M decodes only $R_1\mathbf{x}$ and does not need to decode $R_2\mathbf{x}$. 2) M receives all the sub-tasks from any 2 workers and decodes $A\mathbf{x}$. Here M has to decode $R_1\mathbf{x}$ and $R_2\mathbf{x}$. Note that if M uses the first three sub-shares, it only decodes half of $R\mathbf{x}$, i.e., $R_1\mathbf{x}$, and does not need to decode $R_2\mathbf{x}$. This allows the master to save on communication cost. After receiving enough sub-tasks, the Master sends a message to the workers instructing them to stop computing the remaining sub-tasks. One can check that no information about A is revealed to the workers, because each sub-share is padded by a random matrix.

A. Contributions

To the extent of our knowledge, this paper is the first work to analyze latency for private distributed coded computing under the presence of stragglers. We consider the distributed computing setting described above in which we require the workers to learn no information (in an information theoretic sense) about the Master's data. We study the waiting time of the Master caused by delays of the workers. We follow the literature, e.g., [6], [15], and model the service time at the workers as a shifted exponential random variable. This service time includes upload time, computation time and download time, i.e., computation and network latency. Finding codes that minimize the delay at the Master is still an open problem in general. In this work, we take the download communication cost as a proxy for delay when designing the coding schemes. More precisely, we study the performance of the recently introduced Staircase codes [13], [14] that achieve the information theoretic limit on download cost [16] and compare them to classical secret sharing codes. The encoding and decoding at the Master add delays that are proportional to inverting a $k \times k$ matrix and multiplying this inverse by a vector of length k. The encoding and decoding complexities are of order $\mathcal{O}(k \log k)$ when the generator matrix is a Vandermonde matrix. However, since the codes mentioned in this paper have same encoding and decoding complexities, we do not account for those delays in our latency analysis. Therefore, our delay analysis (and comparison to Staircase codes) is the same for codes that requires a threshold on the number of stragglers, such as [17], [18], and for classical secret sharing schemes.

Before we state our contributions, we introduce some necessary notations. We denote by n the number of workers available to help the Master, k denote the minimum number of non stragglers and z the maximum number of colluding workers. We refer to such secure distributed computing system by an (n, k, z) system. We make the following contributions:

1) *General bounds for systems with any number of stragglers:* We derive an upper and a lower bound on the Master's mean waiting time when using Staircase codes (Theorem 1). Moreover, we derive the exact distribution of the Master's waiting time when using Staircase codes, in an integral form (Theorem 4). Using the upper bound, we compare the performance of Staircase codes to classical secret sharing codes and characterize the savings obtained by Staircase codes. We show that Staircase codes always outperform classical secret sharing codes.
2) *Exact characterization for systems with up to* 2 *stragglers:* We use the integral expression of Theorem 4 to find the exact distribution of the Master's waiting time for systems with up to $n - k = 1$ and up to $n - k = 2$ stragglers (Corollary 5). Moreover, we derive the exact expressions of the Master's mean waiting time for these systems (Theorem 2) and use these expressions to show the tightness of our upper bound.
3) *Simulations and validation:* We ran extensive MATLAB simulations for different system parameters. Our main observation is that the upper bound, based on Jensen's inequality, is a good approximation of the mean waiting time. Furthermore, we validate our results with extensive implementation on Amazon EC2 clusters. The savings obtained on EC2 clusters are within the range of the values predicted by the theoretical model. To give an example, for $n = 4$ workers, large data and high traffic regime, our implementation shows 59% (Figure 4a) savings in the mean waiting time while the theoretical model predicts 66% savings (Figure 3a).

B. Related Work

The problem of stragglers has been identified and studied by the distributed computing community, see e.g., [8], [9], [19], [20]. Recently, there has been a growing research interest in studying codes for straggler mitigation and delay minimization in distributed systems with no secrecy constraints. The early body of work focused on minimizing latency of content download in distributed storage systems, see e.g., [15], [21]–[23] and later the focus has shifted to using codes for straggler mitigation in distributed computing, see e.g., [6], [7], [24]–[29].

Secure multiparty computation [30] can be used in this setting to provide privacy. However, the methods there are generic and not tailored to matrix multiplication and therefore do not have efficient communication cost and flexible straggler mitigation. The work that is closest to ours is [10] that studies the problem of distributively multiplying two private matrices under information theoretic privacy constraints using classical secret sharing codes. Subsequently to our initial result that have appeared in [1], several other works have studied different variants of this problem. In particular, [17], [18], [31]–[36] studied the problem of distributively multiplying two private matrices and [37]–[39] studied the problem of running private distributed machine learning algorithm in the presence of stragglers. In terms of delay analysis, all these works use schemes that assume a threshold on the number of stragglers and have similar delays as classical secret sharing schemes. However, the schemes used to multiply two private matrices require the master to use codes for both matrices which results in more tradeoffs between encoding complexity, straggler tolerance, upload cost and download cost see [18] for example.

In general, privacy in distributed computing is studied separately, mostly in the computer science community. Our work can also be related to the work on privacy-preserving algorithms, e.g., [40]–[43]. However, the privacy constraint in this line of work is computational privacy, and the proposed algorithms are not designed for straggler mitigation.

C. Organization

The paper is organized as follows. We formalize the problem and define the model in Section II. In Section III, we present and discuss our main results. We describe the construction of Staircase codes in Section IV. In Sections V and VI, we study the probability distribution of the Master's waiting time and derive bounds on the mean waiting time. We show, in Section VII, that the (random) number of workers that minimizes the waiting time is concentrated around its average. We evaluate the performance of Staircase codes via simulation in Section VIII. In Section IX, we give a representative sample of our implementation on Amazon EC2 clusters and compare them to our theoretical findings. We conclude the paper in Section X.

II. System Model

We consider a Master server $\mathtt{M}$ which wants to perform intensive computations on confidential data represented by an $m \times \ell$ matrix A (typically $m \gg \ell$). $\mathtt{M}$ divides these computations into smaller computational tasks and assigns them to n workers $\mathtt{W}_i$, $i = 1, \ldots, n$, that can perform these tasks in parallel. The division is horizontal, i.e., each worker gets a given number of rows of A with all their corresponding columns.

a) Computations model: We focus on linear computations. The motivation is that a building block in several iterative machine learning algorithms, such as gradient descent, is the multiplication of A by a sequence of $\ell \times 1$ attribute vectors $\mathbf{x}^1, \mathbf{x}^2, \ldots$. In the sequel, we focus on the multiplication $A\mathbf{x}$ with one attribute vector $\mathbf{x}$.

b) Workers model: The workers have the following properties: 1) The workers incur random delays while executing the task assigned to them by $\mathtt{M}$ resulting in what is known as the straggler problem [6], [8], [9]. 2) Up to z, $z < k$, workers can collude, i.e., at most z workers can share with each other the data they receive from $\mathtt{M}$. The threshold z could be thought of as a desired level of security. This has implications on the privacy constraint described later.

c) General scheme: $\mathtt{M}$ encodes A, using randomness, into n shares S_i sent to worker $\mathtt{W}_i$, $i = 1, \ldots, n$. Any k or more shares can decode A, and any collection of z workers obtain zero information about A. For any set $\mathcal{B} \subseteq \{1, \ldots, n\}$, let $S_{\mathcal{B}} = \{S_i, i \in \mathcal{B}\}$ denote the collection of shares given to worker $\mathtt{W}_i$ for all $i \in \mathcal{B}$. The previous requirements can be expressed as,

$$H(A|S_{\mathcal{B}}) = 0, \quad \forall \mathcal{B} \subseteq \{1, \ldots, n\} \text{ s.t. } |\mathcal{B}| \geq k,$$
$$H(A|S_{\mathcal{Z}}) = H(A), \quad \forall \mathcal{Z} \subseteq \{1, \ldots, n\} \text{ s.t. } |\mathcal{Z}| \leq z.$$

At each iteration, the Master sends $\mathbf{x}$ to all the workers. Then, each worker computes $S_i\mathbf{x}$ and sends it back to the Master. In the case where the share S_i consists of sub-shares, each worker multiplies the sub-shares by $\mathbf{x}$ and sends the result back to the master independently. Since the scheme and the computations are linear, the Master can decode $A\mathbf{x}$ after receiving enough responses. After receiving enough responses the master sends a stop message to the workers instructing them to stop computing on the remaining sub-shares. We refer to such scheme as an (n, k, z) system. We note that our scheme can be generalized to the cases where the attribute vectors $\mathbf{x}$ contain information about A, and therefore need to be hidden from the workers. We describe the generalization of our scheme to such case in [2, Appendix].

d) Encoding: We consider classical secret sharing codes [11], [12] and universal Staircase codes [13], [14]. We describe their properties that are necessary for the delay analysis. Secret sharing codes require the division of A into $k - z$ row blocks each of dimension $\frac{m}{k-z} \times \ell$ and encodes them into n shares of identical dimension. Any k shares can decode A. Similarly, Staircase codes encode A into n shares of $\frac{m}{(k-z)} \times \ell$ each with the additional requirement that each share is divided into $b = \text{LCM}\{k - z + 1, \ldots, n - z\}$ sub-shares, where $\text{LCM}\{a, b, c\}$ denotes the least common multiple of a, b and c. Decoding A requires a fraction $\alpha_d b$ sub-shares, $\alpha_d \triangleq \frac{(k-z)}{(d-z)}$, from any of the d shares, $d \in \{k, \ldots, n\}$. We provide a detailed explanation on the construction of Staircase codes in Section IV. We show that Staircase codes outperform classical codes in terms of incurred delays.

e) Delay model: Let T_A be the random variable representing the time spent to compute $A\mathbf{x}$ at one worker. We assume a mother runtime distribution $F_{T_A}(t)$ that is shifted exponential with rate λ and a constant shift c. This is a popular model for modeling service time in a compute cluster [6], [15], and primarily motivated by its analytical tractability. Furthermore, the shifted exponential distribution captures the two parts of the service completion time: the constant part of the task-dependent service time at each server, and the stochasticity in service due to uncorrelated background processes at each server. For each $i \in \{1, \ldots, n\}$, we let T_i denote the time spent by worker $\mathtt{W}_i$ to execute its task. Due to the encoding, each task given to a worker is $k - z$ times smaller than A, or $T_i = \frac{T_A}{(k-z)}$. It follows that F_{T_i} is a scaled distribution of F_{T_A}. That is, for $t \geq c/(k - z)$,

$$F_{T_i}(t) \triangleq F_{T_A}((k - z)t) = 1 - e^{-\lambda(k-z)(t-\frac{c}{k-z})}. \quad (1)$$

We assume that the T_i's, $i = 1, \ldots, n$, are independent and identically distributed (*iid*). For an (n, k, z) system using Staircase codes, we assume that T_i is evenly distributed among the sub-tasks.[1] That is, the time spent by a worker $\mathtt{W}_i$ on one sub-task is equal to T_i/b, and the time spent on $b\alpha_d = b\frac{k-z}{d-z}$ sub-tasks is $\alpha_d T_i$.

Let $T_{(i)}$ be the i^{th} order statistic of the T_i's and $T_{\text{SC}}(n, k, z)$ be the time the Master waits until it can decode $A\mathbf{x}$. If the aggregate wait is due to d workers each finishing α_d fraction of its b sub-tasks, then the Master's waiting time is $\alpha_d T_{(d)}$. We can write

$$T_{\text{SC}}(n, k, z) = \min_{d \in \{k, \ldots, n\}} \left\{\alpha_d T_{(d)}\right\}. \quad (2)$$

It is useful for our analysis to look at T_i as the sum of an exponential random variable T_i' and a constant offset, i.e. $T_i = T_i' + c/(k - z)$, where $T_i' \sim \exp(\lambda(k - z))$.

From this interpretation, it is easy to verify that the d^{th} order statistic $T_{(d)}$ of $(T_1, T_2, \ldots, T_n)$ can be expressed as

$$T_{(d)} = T'_{(d)} + c/(k - z),$$

where $T'_{(d)}$ is the d^{th} order statistic of n *iid* exponential random variables with rate $\lambda(k - z)$. Therefore, we can write the Master's waiting time for Staircase codes as

$$T_{\text{SC}}(n, k, z) = \min_{d \in \{k, \ldots, n\}} \left\{\alpha_d \Big(T'_{(d)} + \frac{c}{k - z}\Big)\right\}. \quad (3)$$

For an (n, k, z) system using classical secret sharing codes, the Master's waiting time $T_{\text{SS}}(n, k, z)$ is equal to the time spent by the fastest k workers to finish their individual tasks. Hence, we can write

$$T_{\text{SS}}(n, k, z) = T_{(k)}. \quad (4)$$

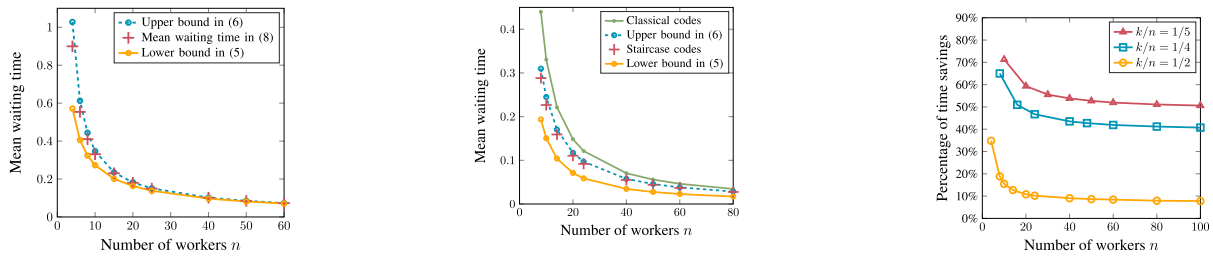
We drop the (n, k, z) notation from $T_{\text{SC}}(n, k, z)$ and $T_{\text{SS}}(n, k, z)$ when the system parameters are clear from the context.

III. Our Results

Our results characterize the delay performance of secure coded computing when using Staircase codes and compare it to classical secret sharing codes. The performance of Staircase codes is reflected in the Master's waiting time T_{SC}. Towards our goal, we establish in Theorem 1 general bounds on the Master's mean waiting time $\mathbb{E}[T_{\text{SC}}(n, k, z)]$ when using Staircase codes for all (n, k, z) systems, under the shifted exponential delay model.

Theorem 1 (Bounds on the Master's Mean Waiting Time $\mathbb{E}[T_{SC}]$): Let H_n be the n^{th} harmonic sum defined as $H_n \triangleq \sum_{i=1}^{n} \frac{1}{i}$, with the notation $H_0 \triangleq 0$. The mean waiting time of

[1]Therefore, we make two assumptions on the service time of the workers: (1) the distribution of service times at each worker is *iid*, (2) at each worker, the service time of a sub-task is proportional to the fraction of the total task at the worker. Accordingly, the parameters of the sub-task distribution (shift c and mean $1/\lambda$) vary linearly with the sub-task size. We observe that the service time of sub-tasks of the task computed at a given worker are proportional to their fractional size, and therefore are not independent. These assumptions make the problem more amenable to theoretical analysis. In Section IX, we compare our model to traces obtained from Amazon cloud and show that our model provides insightful engineering guidelines.



(a) Systems with fixed $n - k = 2$. (b) Systems with fixed $k/n = 1/2$. (c) Savings: systems with fixed k/n.

Fig. 2. Theoretical upper and lower bounds for systems with rate of the exponential random variable $\lambda = 1$, shift $c = 1$ and no colluding workers, i.e., $z = 1$. Figure 2a compares the bounds derived in Theorem 1 to the theoretical mean waiting time for $(k + 2, k, 1)$ derived in Corollary 2. Observe that the upper bound in (5) is a good approximation of the mean waiting time in (8). Figure 2b compares the bounds in (5) and (6) to the simulated mean waiting time for (n, k, z) systems with fixed rate $k/n = 1/2$. We obtain the mean waiting time by averaging over 10000 iterations for each value of n. Figure 2c compares the upper bound in (5) to the mean waiting time of classical secret sharing in (9). The savings are computed as the normalized difference between the waiting time of Staircase codes and classical secret sharing codes, i.e., $(\mathbb{E}[T_{\text{SS}}] - \mathbb{E}[T_{\text{SC}}])/\mathbb{E}[T_{\text{SS}}]$.

the Master $\mathbb{E}[T_{SC}]$ *for an* (n, k, z) *Staircase coded system is upper bounded by*

$$\mathbb{E}[T_{SC}] \leq \min_{d\in\{k,\ldots,n\}} \left(\frac{H_n - H_{n-d}}{\lambda(d-z)} + \frac{c}{d-z} \right). \tag{5}$$

The lower bounded is given in (6)*, as shown at the bottom of this page.*

We derive in Section VI a general integral expression (19) leading to the CDF $F_{T_{\text{SC}}}(t)$ of T_{SC}, the waiting time of the Master for all (n, k, z) systems. Using the general integral expression, we derive the exact expression of the CDF $F_{T_{\text{SC}}}(t)$ for systems with $n = k + 1$ and $n = k + 2$ as stated in the next theorem.

Theorem 2 (Exact Expression of $\mathbb{E}[T_{SC}]$ *for Systems With Up to* 2 *Stragglers): The mean waiting time of the Master for* $(k + 1, k, z)$ *and* $(k + 2, k, z)$ *systems is given in* (7) *and* (8)*, as shown at the bottom of this page, respectively.*

To give insights into the theoretical bounds above, we compare in Figure 2a bounds (5) and (6) for the case of $n = k+2$ to the exact expression in (8). We see that the upper bound in (5) is closer to the actual value and the gap between the two bounds closes as n increases. We also establish the comparison for fixed rate regimes, in particular rate $k/n = 1/2$. Since here $n \geq k + 2$, we compare in Figure 2b the bounds to numerical results obtained by simulation and observe the same behavior as before. We also plot in the same figure the mean waiting time for classical secret sharing codes obtained from (4) and given by

$$\mathbb{E}[T_{\text{SS}}] = \frac{H_n - H_{n-k}}{\lambda(k-z)} + \frac{c}{k-z}. \tag{9}$$

This allows to verify that Staircase codes always outperform classical secret sharing codes. In Figure 2c, we plot the lower bound on the relative savings brought by Staircase codes for systems with rate $k/n = 1/2, 1/4, 1/5$. For instance, for rate $1/4$, the savings are lower bounded by 40% for large n. We supplement our theoretical results in Section VIII with an extensive array of simulations in addition to measurement results obtained by implementation on Amazon EC2 clusters. The savings obtained in the implementation on Amazon cloud are within the savings predicted by the theoretical model.

IV. Staircase Codes

We briefly explain the encoding and decoding of Staircase codes. Let A be an $m \times \ell$ matrix with elements drawn uniformly at random from a finite alphabet, e.g., a finite field[2] $GF(q)$. An (n, k, z) Universal[3] Staircase code [13], [14] allows the Master to encode A into n shares and distribute them to n workers. In addition to privacy against any z colluding workers, Staircase codes enjoy the *secret reconstruction with minimum communication cost* property. The Master can reconstruct the secret by contacting any set

[2]The computation can be carried over the reals by using unbiased quantization as in [38] and references within.

[3]We only describe Universal Staircase codes [14] and shall refer to them as Staircase codes.

$$\mathbb{E}[T_{\text{SC}}] \geq \frac{c}{n-z} + \max_{d\in\{k,\ldots,n\}} \sum_{i=0}^{k-1} \binom{n}{i} \sum_{j=0}^{i} \binom{i}{j} \frac{2(-1)^j}{\lambda\left(2(n-i+j)(d-z) + (n-d)(n-d+1)\right)}. \tag{6}$$

$$\mathbb{E}\left[T_{\text{SC}}(k+1,k,z)\right] = \frac{c}{k-z+1} + \frac{1}{\lambda} \sum_{i=1}^{k+1} (-1)^i \binom{k+1}{i} \left[\frac{i \exp\left(\frac{-\lambda c}{k-z}\right)}{(k-z)i+1} - \frac{1}{(k-z+1)i} \right]. \tag{7}$$

$$\mathbb{E}[T_{\text{SC}}(k+2,k,z)] = \mathbb{E}[T_{\text{SC}}(k+2,k+1,z)] + \sum_{i=2}^{k+2} \frac{(-1)^i}{\lambda} \binom{k+2}{i} \binom{i}{2} \left[\frac{\exp\left(-\frac{4\lambda c}{k-z}\right)}{(k-z)i+4} - \frac{2\exp\left(-\frac{3\lambda c}{k-z}\right)}{(k-z)i+3} \right]. \tag{8}$$

TABLE II

THE STRUCTURE OF THE MATRIX M_{SC} THAT CONTAINS THE SECRET AND KEYS IN THE UNIVERSAL STAIRCASE CODE CONSTRUCTION [14]

$$M_{\text{SC}} = \begin{bmatrix} \mathcal{S} & \mathcal{D}_1 & \mathcal{D}_2 & \cdots & \mathcal{D}_{h-1} \\ \mathcal{R}_1 & \mathcal{R}_2 & \mathcal{R}_3 & \cdots & \mathcal{R}_h \\ & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

staircase structure

$M_1 \quad M_2 \quad M_3 \quad \cdots \quad M_h$

of d, $k \leq d \leq n$, workers and downloading a part of their shares. The information theoretic lower bound on the amount of information downloaded from each worker is referred to as communication cost $CC(d)$ and is given [16] by

$$CC(d) = \frac{k-z}{d-z}. \tag{10}$$

Encoding: Let V be an $n \times n$ Vandermonde matrix defined over $GF(q)$, $q \geq n$. Let M_{SC} be the matrix defined in Table II and detailed next. The encoding of Staircase codes consists of multiplying V by M_{SC} to obtain the matrix $C = VM_{\text{SC}}$. The n rows of C form the n shares.

To construct the matrix M defined in Table II, an (n, k, z) Staircase code requires dividing the data matrix A into $b(k-z)$ matrices $A_1, \ldots, A_{(k-z)b}$ each of dimension[4] $m/b(k-z) \times \ell$, where $b = \text{LCM}\{k+1, \ldots, n-z\}$. Let $d_1 = n, d_2 = n-1, \ldots, d_h = k$ denote the number of workers contacted by the Master, with $h = n-k+1$. Let $b_i \triangleq d_i - z$ for $i = 1, \ldots, h$. The data matrices are arranged in a $b_1 m/(k-z)b \times \ell(k-z)b/b_1$ matrix $\mathcal{S}$. To ensure secrecy, the construction uses zb matrices $R_1, \ldots, R_{zb}$ of dimension $m/b(k-z) \times \ell$ each and with elements drawn independently and uniformly at random from $GF(q)$. The random matrices $R_1, \ldots, R_{zb}$ are partitioned into h matrices $\mathcal{R}_i$, $i = 1, \ldots, h$, each of dimension $zm/(k-z)b \times \ell(k-z)b/b_i b_{i-1}$ with $b_0 = 1$. The matrix M_{SC} is the concatenation of h matrices M_i, $i = 1, \ldots, h$. Each matrix M_i consists of the b_i sub-tasks downloaded by the Master when decoding from d_i workers, i.e., when there are $n - d_i$ stragglers.

The elements appearing in each matrix $\mathcal{D}_j$ are the elements of the $(n-j+1)^{th}$ row of $\begin{bmatrix} M_1 & M_2 \ldots M_j \end{bmatrix}$ rearranged to obtain the dimension of $\mathcal{D}_j$ as $mb_{j+1}/(k-z)b \times \ell(k-z)b/b_j b_{j+1}$ for $j = 1, \ldots, h-1$. The $\mathbf{0}$'s are the all zero matrices used to complete the M_i's to $nm(k-z)b$ rows.

The structure of the matrix M_{SC}, called Staircase structure, allows the Master to decode the secret and achieve optimal communication and read overheads CO and RO for all d, $k \leq d \leq n$.

Decoding: The Master contacts any d_i workers, $i = 1, \ldots, h$ and downloads $\begin{bmatrix} M_1 \ \ldots \ M_i \end{bmatrix}$ from each contacted worker. The Master is guaranteed [14, Theorem 2] to decode the secret.

[4]If the number of rows in A is not divisible by b, one can use zero padding or the representation of A in a smaller field $GF(q_1)$ such that $q = q_1^b$.

In the setting of secure distributed computing, the Master encodes A and sends the n shares to the workers. To compute $A\mathbf{x}$, M sends $\mathbf{x}$ to the workers and waits for the first d, $k \leq d \leq n$, workers to send part of their shares multiplied by $\mathbf{x}$. Since the multiplication is linear, the Master can decode $A\mathbf{x}$ and part of the random matrices $R_i\mathbf{x}$, $i = 1, \ldots, zb$.

V. BOUNDS ON THE MASTER'S MEAN WAITING TIME FOR ALL (n, k, z) SYSTEMS

We derive an upper and a lower bound on the Master's mean waiting time $\mathbb{E}[T_{\text{SC}}(n, k, z)]$ for all (n, k, z) systems, i.e., we prove Theorem 1.

A. Proof of the Upper Bound on the Mean Waiting Time

Proof: We use Jensen's inequality to upper bound the mean waiting time $\mathbb{E}[T_{\text{SC}}]$. Since $\min$ is a convex function, we can use Jensen's inequality to upper bound the mean waiting time,

$$\begin{aligned} \mathbb{E}[T_{\text{SC}}] &= \mathbb{E}\left[\min_{d \in \{k, \ldots, n\}} \left\{ \alpha_d T'_{(d)} + \frac{c}{d-z} \right\}\right] \\ &\leq \min_{d \in \{k, \ldots, n\}} \left\{ \alpha_d \mathbb{E}\left[T'_{(d)}\right] + \frac{c}{d-z} \right\}. \end{aligned} \tag{11}$$

We need the following theorem in order to derive an exact expression of the mean of the d^{th} order statistic of n *iid* exponential random variables.

Theorem (Renyi [44]): The d^{th} order statistic $T'_{(d)}$ of n *iid* exponential random variables T'_i is equal to the following random variable in the distribution

$$T'_{(d)} \triangleq \sum_{j=1}^{d} \frac{T'_j}{n-j+1}.$$

Using Renyi's theorem, the mean of the d^{th} order statistic $\mathbb{E}\left[T'_{(d)}\right]$ can be written as

$$\mathbb{E}[T'_{(d)}] = \mathbb{E}[T'_j] \sum_{j=0}^{d-1} \frac{1}{n-j} = \frac{H_n - H_{n-d}}{\lambda(k-z)}. \tag{12}$$

From equations (11) and (12), the mean waiting time is upper bounded by

$$\mathbb{E}[T_{\text{SC}}] \leq \min_{d \in \{k, \ldots, n\}} \left\{ \frac{H_n - H_{n-d}}{\lambda(d-z)} + \frac{c}{d-z} \right\}.$$

We give an intuitive behavior of the upper bound. The harmonic number can be approximated by $H_n \approx \log(n) + \gamma$, where $\gamma \approx 0.577218$ is called the Euler-Mascheroni constant. Alternatively, we can use the upper and lower bounds $\log(n) < H_n < \log(n+1)$ on the Harmonic number H_n, to upper bound the mean waiting time

$$\begin{aligned} &\mathbb{E}[T_{\text{SC}}] < \min\{\Psi_1, \Psi_2\}, \\ &\Psi_1 \triangleq \min_{d \in \{k, \ldots, n-1\}} \left\{ \frac{1}{\lambda(d-z)} \log\left(\frac{n+1}{n-d}\right) + \frac{c}{d-z} \right\}, \\ &\Psi_2 \triangleq \frac{1}{\lambda(n-z)} \log(n+1) + \frac{c}{n-z}. \end{aligned}$$

■

B. Proof of the Lower Bound on the Mean Waiting Time

Proof: Recall that $T_{SC} = \min\{\alpha_d T_{(d)} : d \in \{k,\ldots,n\}\} = \min\{\alpha_d T'_{(d)} + \frac{c}{d-z} : d \in \{k,\ldots,n\}\}$. Since the minimum of the sum is greater than the sum of the minimums, we can lower bound the waiting time T_{SC} in terms of residual waiting time $T'_{\text{SC}} \triangleq \min\{\alpha_d T'_{(d)} : d \in \{k,\ldots,n\}\}$, as

$$T_{\text{SC}} = \min_{d\in\{k,\ldots,n\}} \{\alpha_d T'_{(d)} + \frac{c}{d-z}\} \geq T'_{\text{SC}} + \frac{c}{(n-z)}.$$

Since the mean of a continuous random variable can be computed by integrating the tail probability, we lower bound $\mathbb{E}[T'_{\text{SC}}]$ by lower bounding the tail probability of T'_{SC} exceeding any threshold value t. We observe that T'_{SC} is greater than t, if and only if the d^{th} order statistic $T'_{(d)}$ is greater than $\frac{t}{\alpha_d}$ for each $d \in \{k,\ldots,n\}$. That is,

$$\{T'_{\text{SC}} > t\} = \bigcap_{d=k}^{n} \left\{T'_{(d)} > \frac{t}{\alpha_d}\right\}.$$

Recall that $t\alpha_d^{-1}(k-z) = t(d-z)$ is increasing in d, and so is $T'_{(d)}$. For the residual service times $T'_1,\ldots,T'_n$, we consider $\mathcal{C}_d(t)$ defined as the following set

$$\left\{T'_{(k)} > \frac{t}{\alpha_d}\right\} \bigcap_{i=d+1}^{n} \left\{T'_{(i)} - T'_{(i-1)} > \frac{t}{\alpha_i} - \frac{t}{\alpha_{i-1}}\right\}.$$

For each $d \in \{k,\ldots,n\}$, we observe that $\mathcal{C}_d(t) \subseteq \{T'_{\text{SC}} > t\}$ since $\{T'_{(k)} > t\alpha_d^{-1}\} \subseteq \cap_{j=k}^{d}\{T'_{(j)} > t\alpha_j^{-1}\}$. It follows that, $\Pr\{T'_{\text{SC}} > t\} \geq \max_{d\in\{k,\ldots,n\}} \Pr(\mathcal{C}_d(t))$. Next, we evaluate $\Pr(\mathcal{C}_d(t))$ explicitly. To this end, we first observe that ${\alpha_j}^{-1} - {\alpha_{j-1}}^{-1} = (k-z)^{-1}$ identically for each $j \in \{1,\ldots,n\}$. Further, we apply Renyi's theorem and independence of residual times T'_is to write

$$\Pr(\mathcal{C}_d(t)) = \Pr\left\{T'_{(k)} > \frac{t}{\alpha_d}\right\} \prod_{j=d+1}^{n} \Pr\left\{\frac{T'_j}{n-j+1} > \frac{t}{(k-z)}\right\}. \quad (13)$$

In the following, we would use $F(t) = 1 - e^{-\lambda t}$ for $t \geq 0$ to represent the cumulative distribution function (CDF) and $\bar{F}(t) = 1 - F(t)$ to represent the complementary cumulative distribution function (CCDF), of an exponential random variable with rate λ. It follows that the CCDF for the residual service time T'_j is $\Pr\{T'_j > t\} = \bar{F}((k-z)t)$. Utilizing the exponential form, we can write (14), as shown at the bottom of this page.

From definition, it follows that $\alpha_k = 1$. Further, the k^{th} order statistic of n residual service times exceeds a threshold if and only if at most $k-1$ different residual service times are less than the threshold, c.f., Lemma 3. That is,

$$\Pr\left\{T'_{(k)} > t\right\} = \sum_{i=0}^{k-1} \binom{n}{i} F\left((k-z)t\right)^i \bar{F}\left((k-z)t\right)^{n-i}. \quad (15)$$

Since $F(t) = 1 - \bar{F}(t)$, using the binomial expansion, we have

$$F\left((k-z)t\right)^i = \sum_{j=0}^{i} \binom{i}{j} (-1)^j \bar{F}\left((k-z)t\right)^j. \quad (16)$$

Exploiting the exponential form of $\bar{F}(t)$, aggregating results from (14), (15) and (16), we can re-write (13) as (17), shown at the bottom of this page. The proof follows from the integral $\int_0^\infty e^{-xt}dt = \frac{1}{x}$, the linearity of integrals, and the lower bound (18), as shown at the bottom of this page. ■

Lemma 3: Marginal complementary distribution of d^{th} order statistic $T'_{(d)}$ of n iid *random variables $(T'_1,\ldots,T'_n)$ with common distribution $f_{T'}(t)$ is given by*

$$\Pr\{T'_{(d)} > t\} = \sum_{i=0}^{d-1} \binom{n}{i} F_{T'}(t)\bar{F}_{T'}(t)^{n-i}.$$

We note the cumulative distribution function (CDF) of f by $F_{T'}(t) \triangleq f_{T'}(T' < t)$ and the complementary cumulative distribution function (CCDF) of f by $\bar{F} \triangleq f_{T'}(T' > t) = 1 - F_{T'}(t)$.

Proof: The d^{th} order statistic is greater than t, if and only if at most $d-1$ out of n *iid* random variables $(T_1,\ldots,T_n)$ can be less than t, and the rest are greater than t. ■

VI. Distribution of the Master's Waiting Time for all (n,k,z) Systems

Now we are ready to derive an integral expression for the probability distribution of T_{SC}, the Master's waiting time when using Staircase codes.

Theorem 4 (Integral Expression Leading to $F_{T_{SC}}(t)$): The distribution of the Master's waiting time T_{SC} of an (n,k,z) system using Staircase codes is given in (19)*, as shown at the bottom of the next page.*

We denote the residual service time at each worker $\mathbb{W}_i$, $i = 1,\ldots,n$, by the random variable $T'_i = T_i - \frac{c}{k-z}$, and the associated distribution by $F(y_i) \triangleq F_{T'}(y_i) = 1 - \exp(-\lambda y_i)$ for $y_i > 0$. For $i = k,\ldots,n$, we define t_i as

$$\prod_{j=d+1}^{n} \Pr\left\{\frac{T'_j}{n-j+1} > \frac{t}{(k-z)}\right\} = \bar{F}\left(\sum_{j=d+1}^{n} (n-j+1)t\right) = \bar{F}\left(\frac{(n-d)(n-d+1)t}{2}\right). \quad (14)$$

$$\Pr(\mathcal{C}_d(t)) = \sum_{i=0}^{k-1} \binom{n}{i} \sum_{j=0}^{i} \binom{i}{j} (-1)^j \bar{F}\big(t(n-i+j)(d-z) + t(n-d)(n-d+1)/2\big). \quad (17)$$

$$\mathbb{E}[T'_{\text{SC}}] = \int_0^\infty \Pr\{T'_{\text{SC}} > t\}\, dt \geq \int_0^\infty \max_{d\in\{k,\ldots,n\}} \Pr(\mathcal{C}_d(t))dt \geq \max_{d\in\{k,\ldots,n\}} \int_0^\infty \Pr(\mathcal{C}_d(t))dt. \quad (18)$$

$t_i \triangleq \max\left\{\left(\frac{i-z}{k-z}\right)\left(t-\frac{c}{i-z}\right),0\right\}$. *We denote by* $A(t)$ *the set of ordered variables* $(y_k,\ldots,y_n)$ *defined as*

$\{0 \le y_k \le y_{k+1} \le \cdots \le y_n : t_k < y_k, \ldots, t_n < y_n\}$.

We apply Theorem 4 to get the mean waiting time of the Master and the exact distribution of the waiting time for systems with $n = k+1$ and $n = k+2$ in Theorem 2 and Corollary 5, respectively.

Corollary 5 (Exact Expression of $F_{T_{SC}}(t)$ for Systems With Up to 2 Stragglers): The distribution of the Master's waiting time for $(k+1,k,z)$ and $(k+2,k,z)$ systems is given in (20) *and* (21)*, as shown at the bottom of this page, respectively. Both distributions are defined for $t>0$, and $F_{T'}(t) \triangleq 1-\exp(-\lambda(k-z)t)$.*

We omit the proof of Corollary 5 since it follows from simply integrating (19) and defer the proof of Theorem 2 to the technical report [2].

Proof of Theorem 4: Let T'_i denote the residual service time of worker i with the offset $\frac{c}{k-z}$. The sequence $(T'_1,\ldots,T'_n)$ of residual service times of n workers is assumed to be *iid* and distributed exponentially with rate $\lambda(k-z)$ with the tail-distribution function $\bar{F}_{T'}(t) \triangleq e^{-\lambda(k-z)t}$ for $t>0$.

Since the common distribution of residual service times is absolutely continuous with respect to the Lebesgue measure, the corresponding probability density exists and is denoted by $f_{T'}(t) = dF_{T'}(t)/dt = \lambda(k-z)e^{-\lambda(k-z)t}$ for $t \geqslant 0$. Further, we know that the order statistics $(T'_{(1)},\ldots,T'_{(n)})$ of residual times $(T'_1,\ldots,T'_n)$ is identical for all their $n!$ permutations. Hence, for any $0 \le y_1 \le \ldots \le y_n$, we can write $f_{T'_{(1)},\ldots,T'_{(n)}}(y_1,\ldots,y_n) = n! f_{T'_1,\ldots,T'_n}(y_1,\ldots,y_n) = n!\prod_{i=1}^n f_{T'}(y_i)$. The product form of joint density follows from the independence of the residual service times.

In terms of $\alpha_j = \frac{k-z}{j-z}$, the order statistics of residual times $T'_{(j)}$, and the offset $\frac{c}{k-z}$, we can write

$$\{T_{\text{SC}} > t\} = \bigcap_{j=k}^{n}\left\{T'_{(j)} > \frac{t}{\alpha_j} - \frac{c}{j-z}\right\}.$$

For each $k \le j \le n$, we define $t_j \triangleq \max\left\{\frac{t}{\alpha_j}-\frac{c}{j-z},0\right\}$, $y_{n+1} \triangleq \infty$, and $\hat{A}(t) \triangleq \cap_{j=k}^{n+1}\{t_j < y_j \le y_{j+1}\} \cap_{j=1}^{k-1}\{0 \le y_j \le y_{j+1}\}$. In terms of t_j, y_{n+1} and $\hat{A}(t)$, we can write the tail distribution as in (22), as shown at the bottom of this page.

First, we compute the integral with respect to ordered non-negative real variables $(y_1,\ldots,y_{k-1})$ over the region $B_{k-1} \triangleq \cap_{j=1}^{k-1}\{0 \le y_j \le y_{j+1}\}$, a projection of $\hat{A}(t)$ on $(k-1)$ dimensional space spanned by $(y_1,\ldots,y_{k-1})$.

Claim 6: For each $k>1$, we have

$$I_k \triangleq \int_{B_{k-1}} dF_{T'}(y_{k-1})\ldots dF_{T'}(y_1) = \int_0^{y_k}\cdots\int_0^{y_2}\prod_{i=1}^{k-1} dF_{T'}(y_i) = \frac{F_{T'}(y_k)^{k-1}}{(k-1)!}.$$

Since the projection of $\hat{A}(t)$ on $(n-k+1)$ dimensional space spanned by $(y_k,\ldots,y_n)$ is equal to $A(t)$, it follows that the integration of the first part is equal to $n!\int_{(y_k,\ldots,y_n)\in A(t)} dF_{T'}(y_n)\ldots dF_{T'}(y_k)$, giving us the result. ∎

VII. Interplay Between Code Design and Latency

Universal Staircase codes allows the master to decode $A\mathbf{x}$ from any random number d of workers, $k \le d \le n$. The downside is that the universal construction requires a large number of sub-tasks $b = \text{LCM}\{k-z+1,\ldots,n-z\}$. In many applications, there may be an overhead associated with excessive divisions into sub-tasks. We show that we can reduce the number of sub-tasks at the expense of a small increase of the Master's waiting time. Using the so-called Δ-Universal Staircase codes [14] reduces the number of sub-tasks at the expense of limiting the Master to a set $\Delta \subseteq \{k,\ldots,n\}$ of number of workers allowing the Master to decode $A\mathbf{x}$. In other words, the Master can decode $A\mathbf{x}$ by downloading enough information from any d workers, $d \in \Delta$. The number of sub-tasks assigned to each worker is reduced from $b = \text{LCM}\{k-z+1,\ldots,n-z\}$ to the least common multiple of all $d_i \in \Delta$. It remains to prove that d is concentrated around its mean. Hence, restricting d to an interval Δ centered around its mean, leads to a reduction in the Master's waiting time.

Next, we prove that the number of workers d that minimize the waiting time is concentrated around its average.

Lemma 7: For an (n,k,z) system, the probability distribution of the distance between d and its average is

$$\Pr\{|d-\mathbb{E}[d]| > t\} \le 2e^{-2t^2/n(n-k)^2}.$$

We prove Lemma 7 by showing that the number of workers d that first finish the aggregate computation is concentrated around its mean, using McDiarmid's inequality. Recall that $d : \mathbb{R}_+^n \to \{k,\ldots,n\}$ is a function of the compute times $T_1,\ldots,T_n$.

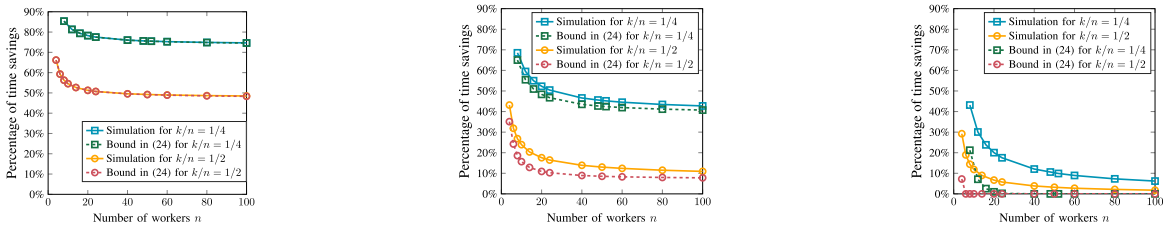
$$d(T_1,\ldots,T_n) \triangleq \arg\min\left\{\frac{k-z}{i-z}T_{(i)} : i \in \{k,\ldots,n\}\right\}.$$

$$F_{T_{\text{SC}}}(t) = 1 - n!\int_{(y_k,\ldots,y_n)\in A(t)} \frac{F_{T'}(y_k)^{k-1}}{(k-1)!} dF_{T'}(y_k)\ldots dF_{T'}(y_n) \quad \text{for } t>0. \tag{19}$$

$$F_{T_{\text{SC}}(k+1,k,z)}(t) = F_{T'}(t_{k+1})^{k+1} + F_{T'}(t_k)^k \bar{F}_{T'}(t_{k+1})(k+1). \tag{20}$$

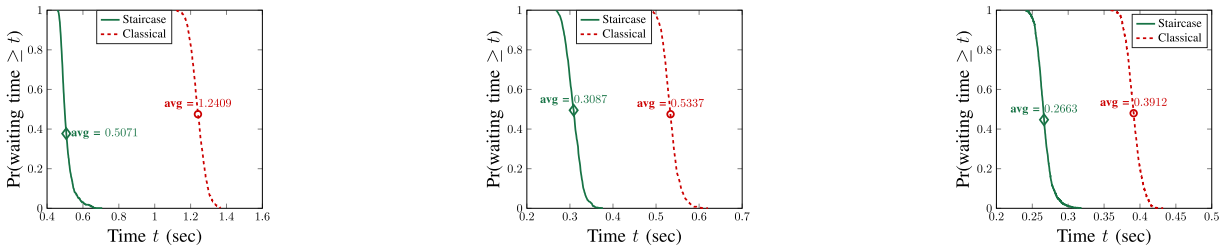
$$F_{T_{\text{SC}}(k+2,k,z)}(t) = F_{T'}(t_{k+2})^{k+2} + (k+2)\bar{F}_{T'}(t_{k+2})\Big[F_{T'}(t_{k+1})^{k+1} + (k+1)F_{T'}(t_k)^k(\bar{F}_{T'}(t_{k+1}) - \frac{1}{2}\bar{F}_{T'}(t_{k+2}))\Big]. \tag{21}$$

$$\Pr\{T_{\text{SC}} > t\} = \int_{y\in\hat{A}(t)} dF_{T'_{(1)},\ldots,T'_{(n)}}(y) = n!\int_{t_n}^{\infty}\cdots\int_{t_k}^{y_{k+1}}\prod_{i=k}^{n} dF_{T'}(y_i)\left(\int_0^{y_k}\cdots\int_0^{y_2}\prod_{i=1}^{k-1} dF_{T'}(y_i)\right). \tag{22}$$



(a) Savings for $\lambda c = 100$. (b) Savings for $\lambda c = 1$. (c) Savings for $\lambda c = 0.001$.

Fig. 3. Savings for the fixed rate regime, $k/n = 1/2$ and $1/4$. The lower bound on the savings of Staircase codes obtained from (24) is compared to the numerical values obtained by simulations. We consider systems with no colluding workers, i.e., $z = 1$, we fix $\lambda = 1$ and vary c. For instance, for systems with rate $k/n = 1/2$ and $\lambda c = 100$ Staircase codes can provide up to 66% reduction in the mean waiting time.



(a) $(n, k, z) = (4, 2, 1)$. (b) $(n, k, z) = (10, 5, 1)$. (c) $(n, k, z) = (20, 10, 1)$.

Fig. 4. Empirical complementary CDF of the Master's waiting time (and its average) observed on Amazon EC2 clusters for systems with rate $k/n = 1/2$. The data matrix A is a 378000×250 matrix with entries generated uniformly at random from $\{1, \ldots, 255\}$. Staircase codes bring 59% reduction in the mean waiting time for $n = 4$. Those numbers were obtained by repeating the multiplication process 1000 times.

Claim 8: The number of workers d that minimize the waiting time is a bounded difference function of compute times with constants $(n-k, \ldots, n-k)$. That is, for each $i \in [n]$ taking $t, t^i \in \mathbb{R}_+^n$ such that $t_j = t_j^i$ for each $j \in [n] \setminus \{i\}$ and $t_i \neq t_i^i$,

$$\sup\{|g(t) - g(t^i)| : t, t^i \in \mathbb{R}_+^n\} \leq n - k. \tag{23}$$

The claim follows from the fact that $d \in \{k, \ldots, n\}$. Therefore, we can apply the McDiarmid's inequality to obtain the concentration bound on d.

VIII. Simulations

We use the normalized difference between the mean waiting time of Staircase codes and classical secret sharing codes as a performance metric for Staircase codes. We refer to this metric as the savings. Using the result of Theorem 1, we can get a lower and an upper bound on the savings brought by Staircase codes. The lower bound on the savings is given in (24).

$$1 - \frac{\mathbb{E}[T_{\text{SC}}]}{\mathbb{E}[T_{\text{SS}}]} \geq 1 - \min_{d \in \{k, \ldots, n\}} \left\{ \frac{(k-z)(\lambda c + H_n - H_{n-d})}{(d-z)(\lambda c + H_n - H_{n-k})} \right\}. \tag{24}$$

To get an idea of the actual savings and the tightness of the bound in (24), we ran numerical simulations of the mean waiting time induced by the use of Staircase codes. By looking at (24), we notice that the bound depends on λ and c only through[5] λc (our simulations show that the actual savings also have a strong dependency on λc). Therefore, we consider three cases for λc : large values of λc ($\lambda c = 100$), medium values of λc ($\lambda c = 1$) and small values of λc ($\lambda c = 0.001$). We ran the simulations for two regimes:

- *Fixed rate k/n*: the plots can be seen in Figure 3. We deduce from the plots that the lower bound is tighter for large values of λc. Moreover, the savings increase with the decrease of the rate k/n and the increase of λc. Note that for large values of λc, the lower bound in (24) converges to $1 - k/n$.
- *Fixed number of parities $n-k$*: We deduce from the plots that similarly to the fixed rate regime the lower bound is tight for large values of λc and that the savings increase with the increase of the number of parities $n-k$ and with the increase of λc. However, we observe that the savings vanish asymptotically with n in this regime. Due to space constraints, the plots are omitted and can be found in [2].

[5]Note that for $c = 0$ we go to the exponential model and the savings would depend only on λ.

IX. Implementation and Validation of the Theoretical Model

We describe a representative sample of our implementation on Amazon EC2 clusters and discuss our observations. We present traces for systems with fixed rate $k/n = 1/2$ (Figure 4). We noticed that the straggler behavior, and therefore the savings, can depend on the date and time of the implementation. We refer interested readers to [2] where we highlight this dependence by presenting traces of one system implemented at different date and times.

Discussion on the theoretical model: Before giving the details, we summarize our findings. We observe that the savings of the system on EC2 can surpass the numerical values resulting from our theoretical model in Section II for

large sizes of the matrix A. However, for small sizes of A, the savings in practice can be less.

The difference between the theoretical results and the implementations can be attributed to several reasons. First, in our model we assume in (2) that the total service time of a task does not change when divided into b sub-tasks, each requiring the same service time. Whereas, our implementation on Amazon shows that the download time decreases faster than linearly with the size of the sub-task for large sub-tasks. Second, for small sub-tasks, we noticed an additional overhead of sending the results of multiple sub-tasks. This overhead becomes non-negligible when the task is small. Third, we have assumed a homogeneous setting where all workers have the same behavior which is not always the case in practice.

Despite these differences, our adopted theoretical model is more amenable to theoretical analysis and provides insightful engineering guiding principles.

We present the implementation of $(4, 2, 1)$, $(10, 5, 1)$ and $(20, 10, 1)$ systems on Amazon EC2 clusters. We use M4.large EC2 instances [45] from Amazon web services (AWS) for our implementation. We assign the Master's job to an instance located in Virginia and the workers job to instances located in Ohio. We plot in Figures 4a, 4b and 4c the empirical complementary CDF of the Master's waiting time for Staircase codes and classical secret sharing codes for $(4, 2, 1)$, $(10, 5, 1)$ and $(20, 10, 1)$ systems, respectively. The average savings brought by Staircase codes are 59%, 42% and 32% for systems with $n = 4$, $n = 10$ and $n = 20$ workers, respectively. Note that for this set of implementation, the Master's data A is a matrix of size 378000×250 with entries generated uniformly at random from $\{1, \ldots, 255\}$. We run 1000 multiplications of A by a randomly generated vector $\mathbf{x}$. In the technical report [2] we present the trace of a $(4, 2, 1)$ system implemented at different dates and times on Amazon EC2 clusters.

X. Conclusion and Open Problems

We consider the problem of secure coded computing. We propose the use of a new family of secret sharing codes called Staircase codes that reduces the delays caused by stragglers. We show that Staircase codes always lead to smaller waiting time compared to classical secret sharing codes, e.g., Shamir secret sharing codes. The reason behind reducing the delays is that Staircase codes allow flexibility in the number of stragglers up to a given maximum, and universally achieve the information theoretic limit on the download cost by the Master, leading to latency reduction. We consider the shifted exponential model for the workers's response time. In our analysis, we find upper and lower bounds on the Master's mean waiting time. We characterize the distribution of the Master's waiting time, and its mean, for systems with $n = k - 1$ and $n = k - 2$. Moreover, we derive an expression that can give the exact distribution, and the mean, of the waiting time of the Master. We supplement our theoretical study with extensive implementation on Amazon EC2 clusters.

While Staircase codes reduce the Master's waiting time by minimizing the download cost, they are not designed to minimize latency. The problem of designing codes that minimize the latency remains open in general. Another open problem, which we leave for future work, is when malicious workers corrupt the results sent to the Master.

References

[1] R. Bitar, P. Parag, and S. El Rouayheb, "Minimizing latency for secure distributed computing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2900–2904.

[2] R. Bitar, P. Parag, and S. El Rouayheb, "Minimizing latency for secure coded computing using secret sharing via staircase codes," 2018, *arXiv:1802.02640*. [Online]. Available: http://arxiv.org/abs/1802.02640

[3] *SETI at Home*. Accessed: Feb. 5, 2017. [Online]. Available: https://setiathome.berkeley.edu

[4] *Folding at Home*. Accessed: Feb. 5, 2017. [Online]. Available: https://foldingathome.stanford.edu

[5] Z. Brakerski and V. Vaikuntanathan, "Efficient fully homomorphic encryption from (standard) LWE," *SIAM J. Comput.*, vol. 43, no. 2, pp. 831–871, 2014.

[6] K. Lee, M. Lam, R. Pedarsani, D. Papailiopoulos, and K. Ramchandran, "Speeding up distributed machine learning using codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1514–1529, Mar. 2018.

[7] S. Dutta, V. Cadambe, and P. Grover, "Short-dot: Computing large linear transforms distributedly using coded short dot products," in *Proc. 29th Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 2092–2100.

[8] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[9] J. Dean and L. A. Barroso, "The tail at scale," *Commun. ACM*, vol. 56, no. 2, p. 74, Feb. 2013.

[10] M. J. Atallah and K. B. Frikken, "Securely outsourcing linear algebra computations," in *Proc. 5th ACM Symp. Inf., Comput. Commun. Secur. ASIACCS*, New York, NY, USA, 2010, pp. 48–59.

[11] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, 1979.

[12] R. J. McEliece and D. V. Sarwate, "On sharing secrets and Reed–Solomon codes," *Commun. ACM*, vol. 24, no. 9, pp. 583–584, Sep. 1981.

[13] R. Bitar and S. El Rouayheb, "Staircase codes for secret sharing with optimal communication and read overheads," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1396–1400.

[14] R. Bitar and S. El Rouayheb, "Staircase codes for secret sharing with optimal communication and read overheads," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 933–943, Feb. 2018.

[15] G. Liang and U. C. Kozat, "TOFEC: Achieving optimal throughput-delay trade-off of cloud storage using erasure codes," in *Proc. IEEE INFOCOM - IEEE Conf. Comput. Commun.*, Apr. 2014, pp. 826–834.

[16] W. Huang, M. Langberg, J. Kliewer, and J. Bruck, "Communication efficient secret sharing," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7195–7206, Dec. 2016.

[17] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "Straggler mitigation in distributed matrix multiplication: Fundamental limits and optimal coding," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1920–1933, Mar. 2020.

[18] Z. Jia and S. A. Jafar, "Cross subspace alignment codes for coded distributed batch computation," 2019, *arXiv:1909.13873*. [Online]. Available: http://arxiv.org/abs/1909.13873

[19] J. Dean *et al.*, "Large scale distributed deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1223–1231.

[20] J. Chen, X. Pan, R. Monga, S. Bengio, and R. Jozefowicz, "Revisiting distributed synchronous SGD," 2016, *arXiv:1604.00981*. [Online]. Available: http://arxiv.org/abs/1604.00981

[21] L. Huang, S. Pawar, H. Zhang, and K. Ramchandran, "Codes can reduce queueing delay in data centers," in *Proc. IEEE Int. Symp. Inf. Theory Proc.*, Jul. 2012, pp. 2766–2770.

[22] G. Joshi, Y. Liu, and E. Soljanin, "Coding for fast content download," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2012, pp. 326–333.

[23] S. Kadhe, E. Soljanin, and A. Sprintson, "Analyzing the download time of availability codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, 2015, pp. 1467–1471.

[24] R. Tandon, Q. Lei, A. G. Dimakis, and N. Karampatziakis, "Gradient coding," in *Proc. 29th Conf. Neural Inf. Process. Syst. (NIPS)*, 2016.

[25] S. Li, M. A. Maddah-Ali, and A. S. Avestimehr, "A unified coding framework for distributed computing with straggling servers," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6.

[26] S. Li, M. A. Maddah-Ali, Q. Yu, and A. S. Avestimehr, "A fundamental tradeoff between computation and communication in distributed computing," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 109–128, Jan. 2018.

[27] W. Halbawi, N. Azizan, F. Salehi, and B. Hassibi, "Improving distributed gradient descent using Reed–Solomon codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2027–2031.
[28] S. Dutta, V. Cadambe, and P. Grover, "Coded convolution for parallel and distributed computing within a deadline," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2403–2407.
[29] Y. Yang, P. Grover, and S. Kar, "Computing linear transformations with unreliable components," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3729–3756, Jun. 2017.
[30] R. Cramer, I. B. Damgård, and J. B. Nielsen, *Secure Multiparty Computation and Secret Sharing*, 1st ed. New York, NY, USA: Cambridge Univ. Press, 2015.
[31] R. G. D'Oliveira, S. El Rouayheb, and D. Karpuk, "Gasp codes for secure distributed matrix multiplication," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Feb. 2019, pp. 1107–1111.
[32] W.-T. Chang and R. Tandon, "On the capacity of secure distributed matrix multiplication," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.
[33] J. Kakar, S. Ebadifar, and A. Sezgin, "Rate-efficiency and straggler-robustness through partition in distributed two-sided secure matrix computation," 2018, *arXiv:1810.13006*. [Online]. Available: http://arxiv.org/abs/1810.13006
[34] H. Yang and J. Lee, "Secure distributed computing with straggling servers using polynomial codes," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 1, pp. 141–150, Jan. 2019.
[35] J. Kakar, A. Khristoforov, S. Ebadifar, and A. Sezgin, "Uplink-downlink tradeoff in secure distributed matrix multiplication," 2019, *arXiv:1910.13849*. [Online]. Available: http://arxiv.org/abs/1910.13849
[36] W.-T. Chang and R. Tandon, "On the upload versus download cost for secure and private matrix multiplication," 2019, *arXiv:1906.10684*. [Online]. Available: http://arxiv.org/abs/1906.10684
[37] Q. Yu, S. Li, N. Raviv, S. M. M. Kalan, M. Soltanolkotabi, and S. A. Avestimehr, "Lagrange coded computing: Optimal design for resiliency, security, and privacy," in *Proc. 22nd Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2019, pp. 1215–1225.
[38] J. So, B. Guler, A. Salman Avestimehr, and P. Mohassel, "CodedPrivateML: A fast and privacy-preserving framework for distributed machine learning," 2019, *arXiv:1902.00641*. [Online]. Available: http://arxiv.org/abs/1902.00641
[39] Q. Yu and A. Salman Avestimehr, "Harmonic coding: An optimal linear code for privacy-preserving gradient-type computation," 2019, *arXiv:1904.13206*. [Online]. Available: http://arxiv.org/abs/1904.13206
[40] H. Takabi, E. Hesamifard, and M. Ghasemi, "Privacy preserving multi-party machine learning with homomorphic encryption," in *Proc. 29th Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 1–5.
[41] R. Hall, S. E. Fienberg, and Y. Nardi, "Secure multiple linear regression based on homomorphic encryption," *J. Off. Statist.*, vol. 27, no. 4, p. 669, 2011.
[42] L. Kamm, D. Bogdanov, S. Laur, and J. Vilo, "A new way to protect privacy in large-scale genome-wide association studies," *Bioinformatics*, vol. 29, no. 7, pp. 886–893, Apr. 2013.
[43] S. Gade and N. H. Vaidya, "Private learning on networks: Part II," 2017, *arXiv:1703.09185*. [Online]. Available: http://arxiv.org/abs/1703.09185
[44] A. Rényi, "On the theory of order statistics," *Acta Math. Hungarica*, vol. 4, nos. 3–4, pp. 191–231, 1953.
[45] *Amazon Web Services*. Accessed: Mar. 27, 2018. [Online]. Available: https://aws.amazon.com/ec2

Rawad Bitar (Member, IEEE) received the Diploma degree in computer and communication engineering from the Faculty of Engineering, Lebanese University, Roumieh, Lebanon, in 2013, the M.S. degree from the Doctoral School, Lebanese University, Tripoli, Lebanon, in 2014, and the Ph.D. degree in electrical engineering from Rutgers University, New Jersey, NJ, USA, in 2020. He is currently a Post-Doctoral Researcher with the Technical University of Munich. His research interests are in the broad area of information theory and coding theory with a focus on coding for information theoretically secure distributed systems with application to machine learning.

Parimal Parag (Member, IEEE) received the B.Tech. and M.Tech. degrees from IIT Madras, India, in 2004, and the Ph.D. degree from Texas A&M University in 2011, all in electrical engineering. He was a Senior System Engineer (corporate R&D) with Assia Inc., Redwood City, CA, USA, from 2011 to 2014. He is currently an Assistant Professor with the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru, India. His research interests lie in the design and analysis of large scale distributed systems. He was a coauthor of the 2018 IEEE ISIT Student Best Paper. He was a recipient of the 2017 Early Career Award from the Science and Engineering Research Board, India.

Salim El Rouayheb (Member, IEEE) received the Diploma degree in electrical engineering from the Faculty of Engineering, Lebanese University, Roumieh, Lebanon, in 2002, the M.S. degree from the American University of Beirut, Lebanon, in 2004, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 2009. He was a Post-Doctoral Research Fellow with UC Berkeley from 2010 to 2011 and a Research Scholar with Princeton University from 2012 to 2013. He was an Assistant Professor with the ECE Department, Illinois Institute of Technology, from 2013 to 2017. He is currently an Assistant Professor with the ECE Department, Rutgers University, New Jersey, NJ, USA. His research interests are in the broad area of information theory and coding theory with a focus on network coding, coding for distributed storage, and information theoretic security. He was a recipient of the NSF Career Award.