Low-Complexity Secret Sharing Schemes Using Correlated Random Variables and Rate-Limited Public Communication

Rumia Sultana and Rémi A. Chou
Department of Electrical Engineering and Computer Science
Wichita State University, Wichita, KS 67260
Emails: {rxsultana@shockers.wichita, remi.chou@wichita}.edu

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517753

***Abstract*—We consider secret sharing where a dealer wants to share a secret with several participants such that predefined subsets of participants can reconstruct the secret and all other subsets of participants cannot learn any information about the secret. To this end, the dealer and the participants have access to samples of correlated random variables and a one-way (from the dealer to the participants), authenticated, public, and rate-limited communication channel. For this problem, we propose the first constructive and low-complexity coding scheme able to handle arbitrary access structures. Our construction relies on a vector quantization coupled with distribution approximations with polar codes to handle the reliability constraints, followed by universal hashing to handle the security constraints. We stress that our coding scheme does not require symmetry or degradation assumptions on the correlated random variables, and does not need a pre-shared secret among the participants and dealer. Our result is also optimal in the special case of rate-unlimited public communication when all the participants are needed to reconstruct the secret.**

I. Introduction

Secret sharing has been introduced in [1] and [2]. Basic secret sharing models consist of a dealer that distributes a secret among a set of participants with the constraint that only pre-defined sets of participants can recover the secret, while any other sets of colluding participants cannot learn any information about the secret. In this paper, unlike in [1], [2], we consider a secret sharing problem where noisy resources are available to the dealer and the participants. Specifically, the participants and dealers have access to samples of correlated random variables. Additionally, we do not assume that secure channels are available, but only assume that a one-way (from the dealer to the participants), authenticated, public, and rate-limited communication channel is available. Such secret sharing models that rely on noisy resources have been introduced in [3] for wireless channels and for source models in [4]–[6]. Channel models, e.g., [3], are similar to compound wiretap channel models [7], while source models, e.g., [4], [5] are related to compound secret-key generation, e.g., [8], [9], and biometric systems with a multiuser access structure [10].

In this paper, we propose the first constructive and low-complexity secret sharing scheme for source models with arbitrary access structures and rate-limited public communication in the asymptotic regime, i.e., when the number of source samples observed by the users approaches infinity. Our construction relies on vector quantizations coupled with distribution approximations (implemented with polar codes) to handle the reliability constraints, followed by universal hashing to handle the security constraints. We do *not* make any assumptions on the correlation of the random variables, e.g., symmetry or degradation assumptions on the source, and do *not* need a pre-shared secret among the participants and dealer.

While no constructive coding scheme has been proposed in the literature for secret sharing with arbitrary access structures, several works have focused on the simpler problem of secret-key generation between two parties from correlated random variables and public communication [11], [12]. Specifically, constructive coding schemes that achieve optimal secret-key rates for this problem have been developed in the case of rate-unlimited public communication by successively handling the reliability requirement and the secrecy requirement by means of source coding with side information and universal hashing, respectively [13]–[15]. While such methods lead to low-complexity coding schemes for unlimited public communication, their application to rate-limited public communication [16], [17] requires vector quantization for which, the construction of low-complexity schemes is challenging. For this reason and the fact that the secret sharing problem in this paper involves more than two parties, these works do not provide a constructive solution to the secret sharing problem with rate-limited public communication considered in this paper. Going back to the problem of secret-key generation between two parties, another approach that jointly handles the reliability and secrecy requirements via polar codes also yields optimal secret-key rates for rate-unlimited communication [18], [19] and rate-limited communication [18]. For arbitrary source correlations, [18], [19] only consider two parties and, unfortunately, these works do not seem to easily extend to the secret sharing problem in this paper. Additionally, a pre-shared seed is also required for known polar coding schemes to ensure strong secrecy. While this pre-shared secret has a negligible rate, such a resource is forbidden in this paper. Finally, note that constructive coding schemes for secret-key generation involving more than two parties have also been proposed in [18], [20]–[22] but only when the correlations of the random variables observed by the participants have specific

This work was supported in part by NSF grant CCF-1850227.

structures. Hence, these works cannot be applied to our setting as we consider a source with arbitrary correlations. We can, however, highlight that a recent construction in [23] provides a low-complexity and constructive coding scheme for a secret sharing channel model.

The remainder of the paper is organized as follows. We state the problem in Section III. We present our main results in Section IV. In Sections V and VI, we introduce an auxiliary result and its proof, respectively. We sketch the proofs of the main results in Sections VII and VIII. Finally, we provide concluding remarks in Section IX.

II. Notation

For $a, b \in \mathbb{R}$, define $[a] \triangleq [1, \lceil a \rceil] \cap \mathbb{N}$, $[\![a, b]\!] \triangleq [\lfloor a \rfloor, \lceil b \rceil] \cap \mathbb{N}$, and $[a]^+ \triangleq \max(0, a)$. The components of a vector $X^{1:N}$ of length $N \in \mathbb{N}$ are denoted with superscripts, i.e., $X^{1:N} \triangleq (X^1, X^2, \ldots, X^N)$. For any set $\mathcal{A} \subset [N]$, let $X^{1:N}[\mathcal{A}]$ be the components of $X^{1:N}$ whose indices are in $\mathcal{A}$. For a probability distribution p_X defined over the alphabet $\mathcal{X}$, define $\mu_{p_X} \triangleq \min_{x \in \mathcal{X}} p_X(x)$. For a set $\mathcal{S}$, let $2^{\mathcal{S}}$ denote the power set of $\mathcal{S}$. In this paper, all the logarithms are taken base two. Finally, let $\bigtimes$ denote the Cartesian product.

III. Problem Statement

Consider a dealer and J participants. Let $\mathcal{Y}_{[J]} \triangleq \bigtimes_{j \in [J]} \mathcal{Y}_j$ be the Cartesian product of J finite alphabets $(\mathcal{Y}_j)_{j \in [J]}$. Consider a discrete memoryless source $((\mathcal{X} \times \mathcal{Y}_j)_{j \in [J]}, (p_{XY_j})_{j \in [J]})$ with $\mathcal{X} \triangleq \{0, 1\}$. $(p_{XY_j})_{j \in [J]}$ is known to all parties. Let $\mathbb{A}$ be a set of subsets of $[J]$ such that for any $\mathcal{T} \subseteq [J]$, if $\mathcal{T}$ contains a set that belongs to $\mathbb{A}$, then $\mathcal{T}$ also belongs to $\mathbb{A}$, i.e., $\mathbb{A}$ is a monotone access structure [24]. We also define $\mathbb{E} \triangleq 2^{[J]} \backslash \mathbb{A}$ as the set of all colluding subsets of users who must not learn any information about the secret. In the following, for any $\mathcal{E} \in \mathbb{E}$ and $\mathcal{A} \in \mathbb{A}$, we use the notation $Y_{\mathcal{E}}^{1:N} \triangleq (Y_j^{1:N})_{j \in \mathcal{E}}$ and $Y_{\mathcal{A}}^{1:N} \triangleq (Y_j^{1:N})_{j \in \mathcal{A}}$. Moreover, the dealer can communicate with the participants over an authenticated, one-way, rate-limited, noiseless, and public communication channel.

Definition 1. *A $(2^{NR_s}, R_p, \mathbb{A}, N)$ secret sharing strategy is defined as follows:*

1) *The dealer observes $X^{1:N}$ and Participant $j \in J$ observes $Y_j^{1:N}$;*
2) *The dealer transmits $M \triangleq f(X^{1:N})$ subject to the communication constraint $H(M) \leq NR_p$;*
3) *The dealer computes a secret $S \in [2^{NR_s}]$ from $X^{1:N}$;*
4) *Any subset of participants $\mathcal{A} \in \mathbb{A}$ can compute an estimate $\hat{S}(\mathcal{A})$ of S from their observations $(Y_j^{1:N})_{j \in \mathcal{A}}$ and M.*

Definition 2. *(R_p, R_s) is said achievable if there exists a sequence of $(2^{NR_s}, R_p, \mathbb{A}, N)$ secret sharing strategies such that*

$$\lim_{N \to +\infty} \max_{\mathcal{A} \in \mathbb{A}} \mathbb{P}[\hat{S}(\mathcal{A}) \neq S] = 0 \quad (\textit{Reliability}), \tag{1}$$

$$\lim_{N \to +\infty} \max_{\mathcal{E} \in \mathbb{E}} I(S; MY_{\mathcal{E}}^{1:N}) = 0 \quad (\textit{Strong Security}), \tag{2}$$

$$\lim_{N \to +\infty} \log|\mathcal{S}| - H(S) = 0 \quad (\textit{Secret Uniformity}). \tag{3}$$

(1) means that any subset of participants in $\mathbb{A}$ is able to recover the secret, (2) means that any subset of participants in $\mathbb{E}$ cannot obtain information about the secret, while (3) means that the secret is nearly uniform, which is meant to maximize the uncertainty of guessing S by the subsets of participants in $\mathbb{E}$. The secret capacity is defined as $C_s(R_p) \triangleq \sup\{R_s : (R_p, R_s)$ is achievable$\}$.

IV. Main Results

Theorem 1. *The secret rate*

$$R_s = \max_{\substack{U \\ U - X - Y_{[J]}}} \left[\min_{\mathcal{A} \in \mathbb{A}} I(U; Y_{\mathcal{A}}) - \max_{\mathcal{E} \in \mathbb{E}} I(U; Y_{\mathcal{E}}) \right]^+$$
$$\text{subject to} \quad R_p = \max_{\mathcal{A} \in \mathbb{A}} I(U; X|Y_{\mathcal{A}})$$

is achievable with an encoder and decoders that operate over t blocks of source observation sequences of length N, and have complexity $T(tN)$, where $T(tN)$ is the complexity of field multiplication in GF(2^{tN}). Note that $T(tN) = O(tN \log(tN))$ for a wide range of values of tN [25, App. D].

Theorem 2. *The secret rate*

$$R_s = \max_{\substack{U,V \\ U - V - X - Y_{[J]}}} \left[\min_{\mathcal{A} \in \mathbb{A}} I(V; Y_{\mathcal{A}}|U) - \max_{\mathcal{E} \in \mathbb{E}} I(V; Y_{\mathcal{E}}|U) \right]^+$$
$$\text{subject to } R_p = \max_{\mathcal{A} \in \mathbb{A}} I(U; X|Y_{\mathcal{A}}) + \max_{\mathcal{A} \in \mathbb{A}} I(V; X|UY_{\mathcal{A}})$$

is achievable with an encoder and decoders that operate over t blocks of source observation sequences of length N, and have complexity $T(tN)$.

Note that the achievable rates in Theorems 1 and 2 could be obtained from [8]. However, [8] only provides an existence result and not a constructive coding scheme.

Next, we have the following corollary for rate-unlimited public communication and when all the participants are needed to reconstruct the secret, i.e., any strict subsets of participants in $[J]$ must not learn any information about the secret.

Corollary 1. *When $R_p = +\infty$ and $\mathbb{A} = \{[J]\}$, the secret capacity*

$$\lim_{R_p \to +\infty} C_s(R_p) = \min_{\mathcal{E} \subsetneq [J]} I(X; Y_{[J]}|Y_{\mathcal{E}})$$

is achievable with an encoder and decoders that operate over t blocks of source observation sequences of length N, and have complexity $T(tN)$.

V. Auxiliary result

In this section, we use the same notation as in Section III and provide an auxiliary result to construct a coding scheme for the secret sharing problem described in Section III. Specifically, we consider a setting similar to the one in Section III with the following modifications. Instead of considering the constraints (1), (2), and (3), the dealer creates a quantized version $\widetilde{U}^{1:N}$ of the source observations $X^{1:N}$, with the requirements that (i) $\widetilde{U}^{1:N}$ follows a pre-determined product

distribution, and (ii) any subsets of participants in the access structure can reconstruct $\widetilde{U}^{1:N}$. This result is formalized in Theorem 3 and will be used to prove Theorem 1.

Consider a discrete memoryless source $((\mathcal{X} \times \mathcal{Y}_{\mathcal{A}})_{\mathcal{A}\in\mathbb{A}}, (p_{XY_{\mathcal{A}}})_{\mathcal{A}\in\mathbb{A}})$ with $\mathcal{X} \triangleq \{0,1\}$. Define the joint probability distribution $p_{XUY_{\mathcal{A}}} \triangleq p_{XY_{\mathcal{A}}}p_{U|X}$, $\mathcal{A} \in \mathbb{A}$.

Theorem 3. *Let $N \triangleq 2^n$, $n \in \mathbb{N}$. For any $\epsilon, \delta > 0$, there exist $k, n_0 \in \mathbb{N}$ such that for any $n \geq n_0$, there exist an encoder*

$$f : \mathcal{X}^{kN} \to \mathcal{U}^{kN} \times \mathcal{M},$$

and $|\mathbb{A}|$ decoders

$$g_{\mathcal{A}} : \mathcal{M} \times \mathcal{Y}_{\mathcal{A}}^{kN} \to \mathcal{U}^{kN}, \forall \mathcal{A} \in \mathbb{A},$$

such that the public communication rate satisfies

$$R_M \leq \max_{\mathcal{A}\in\mathbb{A}} I(U; X|Y_{\mathcal{A}}) + \delta, \tag{4}$$

the probability of error at Decoder $\mathcal{A} \in \mathbb{A}$ satisfies

$$\mathbb{P}[\widetilde{U}_{1:k}^{1:N} \neq g_{\mathcal{A}}(M, Y_{\mathcal{A},1:k}^{1:N})] < \epsilon, \forall \mathcal{A} \in \mathbb{A}, \tag{5}$$

where $X_{1:k}^{1:N} \triangleq (X_i^{1:N})_{i\in[k]}$ is the source observation at the encoder with $X_i^{1:N}$, $i \in [k]$, a sequence of length N, $Y_{\mathcal{A},1:k}^{1:N} \triangleq (Y_{\mathcal{A},i}^{1:N})_{i\in[k]}$ is the source observation at Decoder $\mathcal{A} \in \mathbb{A}$, and $(\widetilde{U}_{1:k}^{1:N}, M) \triangleq f(X_{1:k}^{1:N})$. $X_{1:k}^{1:N}$ and $Y_{\mathcal{A},1:k}^{1:N}$ are distributed according to $p_{X_{1:k}^{1:N}U_{1:k}^{1:N}Y_{\mathcal{A},1:k}^{1:N}} \triangleq \prod_{i=1}^{Nk} p_{XUY_{\mathcal{A}}}$, $\mathcal{A} \in \mathbb{A}$. The joint probability distribution induced by the encoding scheme $\widetilde{p}_{U_{1:k}^{1:N}X_{1:k}^{1:N}}$ satisfies

$$\mathbb{D}\left(\widetilde{p}_{U_{1:k}^{1:N}X_{1:k}^{1:N}} \middle\| p_{U_{1:k}^{1:N}X_{1:k}^{1:N}}\right) \leq \delta_N^{(1)}, \tag{6}$$

where

$$\delta_N^{(1)} \triangleq k^{\frac{3}{2}} N \log\left(\mu_{p_{UX}}^{-1}\right) \sqrt{2\ln 2}\left[\sqrt{N\delta_N} + \sqrt{N\delta_N + \delta_N^{(2)}}\right],$$

$$\delta_N^{(2)} \triangleq 2\sqrt{2\ln 2}\sqrt{N\delta_N} \log\left(|\mathcal{U}|^{2N}|\mathcal{X}|^N / \left[\sqrt{2\ln 2}\sqrt{N\delta_N}\right]\right),$$

$$\delta_N \triangleq 2^{-N^\beta}, \beta \in]0, \frac{1}{2}[.$$

Moreover, the complexity of the encoder and decoders is $O(kN \log N)$.

Note that Theorem 3 recovers [26, Th. II.1] in the absence of vector quantization and the constraint (6).

VI. Proof of Theorem 3

A. Notation

Let $N \triangleq 2^n$, $n \in \mathbb{N}$. Let $G_n \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n}$ be the source polarization matrix defined in [27]. Fix a sequence of joint probability distribution $(p_{UXY_{\mathcal{A}}})_{\mathcal{A}\in\mathbb{A}}$. Define $U^{1:N} \triangleq V^{1:N}G_n$. For $\delta_N \triangleq 2^{-N^\beta}$, $\beta \in]0, \frac{1}{2}[$ define the following sets

$$\mathcal{H}_U \triangleq \left\{i \in [N] : H(V^i|V^{1:i-1}) \geq \delta_N\right\},$$
$$\mathcal{V}_{U|X} \triangleq \left\{i \in [N] : H(V^i|V^{1:i-1}X^{1:N}) \geq 1 - \delta_N\right\},$$
$$\mathcal{H}_{U|Y_{\mathcal{A}}} \triangleq \left\{i \in [N] : H(V^i|V^{1:i-1}Y_{\mathcal{A}}^{1:N}) \geq \delta_N\right\}, \mathcal{A} \in \mathbb{A}.$$

We refer to [28], [29] for an interpretation of these sets in terms of randomness extraction, source coding, and source resolvability. We write the access structure as $\mathbb{A} = \{\mathcal{A}_j \subseteq [J] : j \in [|\mathbb{A}|]\}$, and adopt, for convenience, in our coding scheme and its analysis a slightly different notation than in the statement of Theorem 3. Specifically, we define $B_j \triangleq Y_{\mathcal{A}_j}$ and $B_{j,1:k}^{1:N} \triangleq Y_{\mathcal{A}_j,1:k}^{1:N}$. This notation is convenient as the analysis of our coding scheme is done by induction over the number of decoders, i.e., the cardinality of $\mathbb{A}$. Then, for a finite set of integers $\mathcal{S} \subseteq [|\mathbb{A}|]$, consider the following notation for a set of $|\mathcal{S}|$ decoders $(g_{\mathcal{S},j}^{1:kN})_{j\in\mathcal{S}}$. The superscript indicates the length of the output, the first subscript $\mathcal{S}$ indicates that all the decoders are indexed by $\mathcal{S}$, the second subscript $j \in \mathcal{S}$ indicates that $g_{\mathcal{S},j}^{1:N}$ is the decoding function for the decoder indexed by $j \in \mathcal{S}$, i.e., the decoder that has access to the source observations $B_{j,1:k}^{1:N} = Y_{\mathcal{A}_j,1:k}^{1:N}$. Similarly, consider the following notation for an encoder $f_{\mathcal{S}}^{1:kN}$. The superscript indicates the length of the input, the subscript $\mathcal{S}$ indicates that all the decoders are indexed by $\mathcal{S}$.

B. Preliminary Results

Algorithm 1 describes the construction of $\widetilde{U}_{1:k}^{1:N}$ from the random variables $X_{1:k}^{1:N}$ and a vector R_1 of $|\mathcal{V}_{U|X}|$ uniformly distributed bits. For N large enough, one can prove that

$$\mathbb{D}\left(\widetilde{p}_{U_{1:k}^{1:N}X_{1:k}^{1:N}} \| p_{U_{1:k}^{1:N}X_{1:k}^{1:N}}\right) \leq \delta_N^{(1)}, \tag{7}$$

where $\delta_N^{(1)}$ is defined in Theorem 3. The proof is omitted due to space constraints. Note that in Line 4 of Algorithm 1, some of the random drawings can be simplified by deterministic decisions similar to [30].

Algorithm 1 Construction of $\widetilde{U}_{1:k}^{1:N}$

Require: Source observations $(X_i^{1:N})_{i\in[k]}$, where $X_i^{1:N}$ is of length N and corresponds to Block $i \in [k]$; a vector R_1 of $|\mathcal{V}_{U|X}|$ uniformly distributed bits.

1: **for** Block $i = 1$ to k **do**
2: $R_i \leftarrow R_1$
3: $\widetilde{V}_i^{1:N}[\mathcal{V}_{U|X}] \leftarrow R_i$
4: Given $X_i^{1:N}$, successively draw the remaining bits of $\widetilde{V}_i^{1:N}$ according to $\widetilde{p}_{V_i^{1:N}X_i^{1:N}} \triangleq \prod_{j=1}^{N} \widetilde{p}_{V_i^j|V_i^{1:j-1}X_i^{1:N}} p_{X^{1:N}}$ with

$$\widetilde{p}_{V_i^j|V_i^{1:j-1}X_i^{1:N}}(v_i^j|\widetilde{V}_i^{1:j-1}X_i^{1:N}) \triangleq \begin{cases} p_{V^j|V^{1:j-1}X^{1:N}}(v_i^j|\widetilde{V}_i^{1:j-1}X_i^{1:N}) & \text{if } j \in \mathcal{H}_U \backslash \mathcal{V}_{U|X}, \\ p_{V^j|V^{1:j-1}}(v_i^j|\widetilde{V}_i^{1:j-1}) & \text{if } j \in \mathcal{H}_U^c. \end{cases}$$

5: Construct $\widetilde{U}_i^{1:N} \triangleq \widetilde{V}_i^{1:N}G_n$
6: **end for**
7: **return** $\widetilde{U}_{1:k}^{1:N} \triangleq (\widetilde{U}_i^{1:N})_{i\in[k]}$

C. Proof of Equations (4), (5), and (6)

This proof is by induction over the cardinality of $\mathbb{A}$. Assume that $|\mathbb{A}| = 1$. The encoder and decoder for the case $|\mathbb{A}| = 1$ are defined in Algorithms 2 and 3, respectively. One can prove

Algorithm 2 Encoding when $|\mathbb{A}|= 1$

Require: Source observations $X_{1:k}^{1:N} \triangleq (X_i^{1:N})_{i\in[k]}$, where $X_i^{1:N}$ is of length N and corresponds to Block $i \in [k]$; a vector R_1 of $|\mathcal{V}_{U|X}|$ uniformly distributed bits

1: Transmit R_1 to Bob over the public channel
2: Construct $\widetilde{U}_{1:k}^{1:N}$ using Algorithm 1 with input the random variables $X_{1:k}^{1:N}$ and R_1
3: Define

$$M \triangleq (\widetilde{V}_i^{1:N}[\mathcal{H}_{U|B_1}\backslash\mathcal{V}_{U|X}])_{i\in[k]}.$$

4: Transmit M to Bob over the public channel
5: Define $f_{\{1\}}^{1:kN}(X_{1:k}^{1:N}) \triangleq (\widetilde{U}_{1:k}^{1:N}, M)$

Algorithm 3 Decoding when $|\mathbb{A}|= 1$

Require: Source observations $B_{1,1:k}^{1:N} \triangleq (B_{1,i}^{1:N})_{i\in[k]}$, where $B_{1,i}^{1:N}$ is of length N and corresponds to Block $i \in [k]$; the message M and R_1.

1: **for** Block $i = 1$ to k **do**
2: Form $\widetilde{V}_i^{1:N}[\mathcal{H}_{U|B_1}]$ from M and R_1
3: Create $\hat{U}_{1,i}^{1:N}$, an estimate of $\widetilde{U}_i^{1:N}$, from $(\widetilde{V}_i^{1:N}[\mathcal{H}_{U|B_1}], B_{1,i}^{1:N})$ using the successive cancellation decoder of [27].
4: **end for**
5: Define $\hat{U}_{1,1:k}^{1:N} \triangleq (\hat{U}_{1,i}^{1:N})_{i\in[k]}$
6: Define $g_{\{1\},1}^{1:kN}(M, B_{1,1:k}^{1:N}) \triangleq \hat{U}_{1,1:k}^{1:N}$

that the communication rate satisfies

$$\lim_{N\to+\infty} R_{\{1\}} = I(U, X|B_1) + \frac{H(U|X)}{k}.$$

The proof is omitted due to space constraints.

Next, by considering a coupling [31] between $\widetilde{p}_{U_i^{1:N}}$ and $p_{U^{1:N}}$, $i \in [k]$, such that $\mathbb{P}[\mathcal{E}_i] = \mathbb{V}(\widetilde{p}_{U_i^{1:N}}, p_{U^{1:N}})$, where $\mathcal{E}_i \triangleq \{\widetilde{U}_i^{1:N} \neq U^{1:N}\}$, one can show that

$$\mathbb{P}[\hat{U}_{1,1:k}^{1:N} \neq \widetilde{U}_{1:k}^{1:N}] \leq k(\sqrt{2\log 2}\sqrt{N\delta_N} + N\delta_N).$$

The proof is omitted due to space constraints.

Finally, since $\widetilde{U}_{1:k}^{1:N}$ is constructed with Algorithm 1 as indicated in Line 2 of Algorithm 2, Equation (6) holds by (7). This proves the theorem for $|\mathbb{A}|= 1$.

Now suppose that the theorem holds for $|\mathbb{A}|= L$. Fix $\epsilon > 0$ and $\delta > 0$. By the induction hypothesis, there exist k_1, n_1 such that for any $n \geq n_1$, there is an encoder

$$f_{[L]}^{1:k_1N} : \mathcal{X}^{k_1N} \to \mathcal{U}^{k_1N} \times \mathcal{M}_{[L]},$$

and L decoders

$$g_{[L],l}^{1:k_1N} : \mathcal{M}_{[L]} \times \mathcal{Y}_l^{k_1N} \to \mathcal{U}^{k_1N}, \forall l \in [L],$$

such that, the communication rate satisfies

$$R_{[L]} \leq \max_{l\in[L]} I(U; X|B_l) + \frac{\delta}{2},$$

and for k_2 large enough such that

$$\left(1 + k_2^{-1}\right)\left(\max_{l\in[L+1]} I(U; X|B_l) + \frac{\delta}{2}\right) + k_2^{-1}H(U|X) \leq \max_{l\in[L+1]} I(U; X|B_l) + \delta,$$

the probability of error satisfies

$$\mathbb{P}[\widetilde{U}_{1:k_1}^{1:N} \neq g_{[L],l}^{1:k_1N}(M_{[L]}, B_{l,1:k_1}^{1:N})] \leq \frac{\epsilon}{k_2}, \forall l \in [L],$$

where $(\widetilde{U}_{1:k_1}^{1:N}, M_{[L]}) \triangleq f_{[L]}^{1:k_1N}(X_{1:k_1}^{1:N})$. Next, define, as in the case $|\mathbb{A}|= 1$, the encoder

$$f_{\{L+1\}}^{1:k_1N} : \mathcal{X}^{k_1N} \to \mathcal{U}^{k_1N} \times \mathcal{M}_{\{L+1\}},$$

and the decoder

$$g_{\{L+1\},L+1}^{1:k_1N} : \mathcal{M}_{\{L+1\}} \times \mathcal{Y}_{L+1}^{k_1N} \to \mathcal{U}^{k_1N}.$$

From the case $|\mathbb{A}|= 1$, there exists n_2 such that for $n \geq n_2$, the communication rate satisfies

$$R_{\{L+1\}} \leq I(U; X|B_{L+1}) + \frac{\delta}{2},$$

and the probability of error satisfies

$$\mathbb{P}[\widetilde{U}_{1:k_1}^{1:N} \neq g_{\{L+1\},L+1}^{1:k_1N}(M_{\{L+1\}}, B_{L+1,1:k_1}^{1:N})] \leq \frac{\epsilon}{k_2},$$

where $(\widetilde{U}_{1:k_1}^{1:N}, M_{\{L+1\}}) \triangleq f_{\{L+1\}}^{1:k_1N}(X_{1:k_1}^{1:N})$. We now describe the encoder and decoders for the case $|\mathbb{A}|= L + 1$. Let $k \triangleq k_1k_2$, $n_0 \triangleq \max(n_1, n_2)$. The encoder is defined in Algorithm 4, the first L decoders and Decoder $L+1$ are defined in Algorithms 5 and 6, respectively. With this construction, one can show that (4), (5), and (6) hold. The proof is omitted due to space constraints.

Algorithm 4 Encoding for the case $|\mathbb{A}|= L + 1$

Require: Source observations $X_{1:k_2}^{1:k_1N} \triangleq (X_i^{1:k_1N})_{i\in[k_2]}$, where $X_i^{1:k_1N}$ is of length k_1N and corresponds to Block $i \in [k_2]$; a vector R_1 of $|\mathcal{V}_{U|X}|$ uniformly distributed bits

1: Transmit R_1 to Bob over the public channel
2: Construct $\widetilde{U}_{1:k_2}^{1:k_1N}$ using Algorithm 1 from the random variables $X_{1:k_2}^{1:k_1N}$ and R_1
3: Define $(\cdot, M_{[L],i}) \triangleq f_{[L]}^{1:k_1N}(X_i^{1:k_1N})$ for $i \in [k_2]$
4: Define $(\cdot, M_{\{L+1\},i}) \triangleq f_{\{L+1\}}^{1:k_1N}(X_i^{1:k_1N})$ for $i \in [k_2]$
5: Define

$$M_{[L+1]} \triangleq \left[M_{[L],1}, (M_{[L],i+1} \oplus M_{\{L+1\},i})_{i=1}^{k_2-1}, M_{\{L+1\},k_2}\right],$$

where $\oplus$ denotes bitwise modulo two addition. If the two sequences in the sum have different length, then the shorter sequence is padded with zeros.

6: Transmit M to Bob over the public channel
7: Define $f_{[L+1]}^{1:kN}(X_{1:k_2}^{1:k_1N}) \triangleq (\widetilde{U}_{1:k_2}^{1:k_1N}, M_{[L+1]})$

Algorithm 5 Decoding at Decoder $l \in [L]$

Require: Source observations $B_{l,1:k_2}^{1:k_1N} \triangleq (B_{l,i}^{1:k_1N})_{i\in[k_2]}$, where $B_{l,i}^{1:k_1N}$ is of length k_1N and corresponds to Block $i \in [k_2]$; the message $M_{[L+1]}$ and R_1

1: Create an estimate of $\widetilde{U}_1^{1:k_1N}$

$$\hat{U}_{l,1}^{1:k_1N} \triangleq g_{[L],l}^{1:k_1N}(M_{[L],1}, B_{l,1}^{1:k_1N}),$$

where $M_{[L],1}$ is contained in $M_{[L+1]}$

2: **for** Block $i = 1$ to $k_2 - 1$ **do**

3: Define $(\cdot, \hat{M}_{\{L+1\},i}) \triangleq f_{\{L+1\}}^{1:k_1N}(\hat{U}_{l,i}^{1:k_1N})$

4: Since $M_{[L+1]}$ contains $M_{[L],i+1} \oplus M_{\{L+1\},i}$, the decoder can form

$$\hat{U}_{l,i+1}^{1:k_1N} \triangleq g_{[L],l}^{1:k_1N}\Big(M_{[L],i+1} \oplus M_{\{L+1\},i} \oplus \hat{M}_{\{L+1\},i}, B_{l,i+1}^{1:k_1N}\Big)$$

5: **end for**

6: Define $\hat{U}_{l,1:k_2}^{1:k_1N} \triangleq (\hat{U}_{l,i}^{1:k_1N})_{i\in[k_2]}$

7: Define $g_{[L+1],l}^{1:kN}(M_{[L+1]}, B_{l,1:k_2}^{1:k_1N}) \triangleq \hat{U}_{l,1:k_2}^{1:k_1N}$

Algorithm 6 Decoding at Decoder $L + 1$

Require: Source observations $B_{L+1,1:k_2}^{1:k_1N} \triangleq (B_{L+1,i}^{1:k_1N})_{i\in[k_2]}$, where $B_{L+1,i}^{1:k_1N}$ is of length k_1N and corresponds to Block $i \in [k_2]$; the message $M_{[L+1]}$ and R_1

1: Create an estimate

$$\hat{U}_{L+1,k_2}^{1:k_1N} \triangleq g_{\{L+1\},L+1}^{1:kN}(M_{\{L+1\},k_2}, B_{L+1,k_2}^{1:k_1N}),$$

where $M_{\{L+1\},k_2}$ is contained in $M_{[L+1]}$

2: **for** Block $i = k_2 - 1$ to 1 **do**

3: Define $(\cdot, \hat{M}_{[L],i+1}) \triangleq f_{[L]}^{1:k_1N}(\hat{U}_{L+1,i+1}^{1:k_1N})$

4: Since $M_{[L+1]}$ contains $M_{\{L+1\},i} \oplus M_{[L],i+1}$, the decoder can form

$$\hat{U}_{L+1,i}^{1:k_1N} \triangleq g_{\{L+1\},L+1,i}^{1:k_1N}\big(M_{\{L+1\},i} \oplus M_{[L],i+1} \oplus \hat{M}_{[L],i+1}, B_{L+1,i}^{1:k_1N}\big)$$

5: **end for**

6: Define $\hat{U}_{L+1,1:k_2}^{1:k_1N} \triangleq (\hat{U}_{L+1,i}^{1:k_1N})_{i\in[k_2]}$

7: Define $g_{[L+1],L+1}^{1:kN}(M_{[L+1]}, B_{L+1,1:k_2}^{1:k_1N}) \triangleq \hat{U}_{L+1,1:k_2}^{1:k_1N}$

VII. Proof sketch of Theorem 1

We only describe the coding scheme, its analysis is omitted due to space constraints. Using Theorem 3, consider a coding scheme $\mathcal{C}$ that achieves the public communication rate $R_p = \max_{\mathcal{A}\in\mathbb{A}} I(U;X|Y_{\mathcal{A}})$ for the joint probability distribution $(p_{UXY_{\mathcal{A}}})_{\mathcal{A}\in\mathbb{A}} \triangleq (p_{U|X}p_{XY_{\mathcal{A}}})_{\mathcal{A}\in\mathbb{A}}$. For $\mathcal{A} \in \mathbb{A}$, let $\widetilde{p}_{U^{1:N}X^{1:N}Y_{\mathcal{A}}^{1:N}M}$ be the probability distribution of the random variables $(\widetilde{U}^{1:N}, X^{1:N}, Y_{\mathcal{A}}^{1:N}, M)$ induced by the coding scheme $\mathcal{C}$. Repeat t times and independently the coding scheme $\mathcal{C}$. For these t repetitions, the first outputs of the encoder are denoted by $\widetilde{U}^{1:tN}$, and the estimate of $\widetilde{U}^{1:tN}$ at Decoder $\mathcal{A} \in \mathbb{A}$ is denoted by $\hat{U}_{\mathcal{A}}^{1:tN}$.

Consider a hash function $F : \mathcal{R}\times\{0,1\}^{tN} \longrightarrow \{0,1\}^{tNR_s}$, where R_s is the secrecy rate (that can be shown to be the rate in Theorem 1). The encoder forms $S \triangleq F(R, \widetilde{U}^{1:tN})$, where R represents the uniformly random choice of the hash function in a family of two-universal hash functions and is transmitted over the public communication channel. Hence, R is made available to all parties. Note that by an hybrid argument, e.g., [32], the communication rate associated with the transmission of R can be made negligible. Finally, Decoder $\mathcal{A} \in \mathbb{A}$ forms $\hat{S}(\mathcal{A}) \triangleq F(R, \hat{U}_{\mathcal{A}}^{1:tN})$ for $\mathcal{A} \in \mathbb{A}$.

Note that F can be implemented with complexity $T(tN)$ [33], where $T(tN)$ is the complexity of field multiplication in GF(2^{tN}).

VIII. Proof Sketch of Theorem 2

The proof of Theorem 2 largely relies on the proof of Theorem 1. The proof idea is as follows. First, using Theorem 3, we construct a coding scheme $\mathcal{C}_U$ where the dealer generates $\widetilde{U}^{1:N}$ and the message M_U with rate

$$R_U = \max_{\mathcal{A}\in\mathbb{A}} I(U;X|Y_{\mathcal{A}})$$

from the source observation $X^{1:N}$ such that any subset of the participants $\mathcal{A} \in \mathbb{A}$ can form $\hat{U}_{\mathcal{A}}^{1:N}$ from $(M_U, Y_{\mathcal{A}}^{1:N})$ and $\lim_{N\to\infty} \mathbb{P}[\widetilde{U}^{1:N} \neq \hat{U}_{\mathcal{A}}^{1:N}] = 0$. Next, using a slightly modified version of Theorem 3 we construct a coding scheme $\mathcal{C}_V$, where the dealer generates $\widetilde{V}^{1:N}$ and the message M_V with rate

$$R_V = \max_{\mathcal{A}\in\mathbb{A}} I(V;X|UY_{\mathcal{A}})$$

from the source observation $X^{1:N}$ such that any subset of the participants $\mathcal{A} \in \mathbb{A}$ can form $\hat{V}_{\mathcal{A}}^{1:N}$ from $(M_V, Y_{\mathcal{A}}^{1:N}, \widetilde{U}^{1:N})$ and $\lim_{N\to\infty} \mathbb{P}[\widetilde{V}^{1:N} \neq \hat{V}_{\mathcal{A}}^{1:N}] = 0$.

Then, similar to the proof sketch of Theorem 1, consider a hash function $F : \mathcal{R} \times \{0,1\}^{tN} \longrightarrow \{0,1\}^{tNR_s}$, where R_s is the secrecy rate (that can be shown to be the rate in Theorem 2). The encoder forms $S \triangleq F(R, \widetilde{V}^{1:tN})$, where R represents the uniformly random choice of the hash function in a family of two-universal hash functions and is transmitted over the public communication channel. Finally, Decoder $\mathcal{A} \in \mathbb{A}$ forms $\hat{S}(\mathcal{A}) \triangleq F(R, \hat{V}_{\mathcal{A}}^{1:tN})$ for $\mathcal{A} \in \mathbb{A}$.

IX. Conclusion

We considered secret sharing from correlated random variables and rate-limited public communication. For this problem, we proposed the first constructive and low-complexity coding scheme able to handle arbitrary access structures. Our construction relies on a vector quantization coupled with distribution approximations with polar codes to handle the reliability constraints, followed by universal hashing to handle the security constraints. We stress that our coding scheme does not require symmetry or degradation assumptions on the correlated random variables, and does not need a pre-shared secret among the participants and dealer. Our result is also optimal in the special case of rate-unlimited public communication when all the participants are needed to reconstruct the secret.

REFERENCES

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

[2] G. R. Blakley, "Safeguarding cryptographic keys," in *Managing Requirements Knowledge, International Workshop*, 1979, pp. 313–313.

[3] S. Zou, Y. Liang, L. Lai, and S. Shamai, "An information theoretic approach to secret sharing," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3121–3136, 2015.

[4] I. Csiszár and P. Narayan, "Capacity of a shared secret key," in *Proc. of IEEE Int. Symp. Inf. Theory*, 2010, pp. 2593–2596.

[5] R. A. Chou, "Secret sharing over a public channel from correlated random variables," in *Proc. of IEEE Int. Symp. Inf. Theory*, 2018, pp. 991–995.

[6] V. Rana, R. A. Chou, and H. Kwon, "Secret sharing from correlated gaussian random variables and public communication," in *IEEE Information Theory Workshop (ITW)*, 2021.

[7] Y. Liang, G. Kramer, and H. V. Poor, "Compound wiretap channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 1–12, 2009.

[8] N. Tavangaran, H. Boche, and R. F. Schaefer, "Secret-key generation using compound sources and one-way public communication," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 1, pp. 227–241, 2016.

[9] M. Bloch, "Channel intrinsic randomness," in *Proc. of IEEE Int. Symp. Inf. Theory*, 2010, pp. 2607–2611.

[10] R. A. Chou, "Biometric systems with multiuser access structures," in *Proc. of IEEE Int. Symp. Inf. Theory*, 2019, pp. 807–811.

[11] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, 1993.

[12] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[13] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1915–1923, 1995.

[14] C. Cachin and U. M. Maurer, "Linking information reconciliation and privacy amplification," *Journal of Cryptology*, vol. 10, no. 2, pp. 97–110, 1997.

[15] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2000, pp. 351–368.

[16] R. A. Chou and M. R. Bloch, "Separation of reliability and secrecy in rate-limited secret-key generation," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4941–4957, 2014.

[17] S. Nitinawarat and P. Narayan, "Secret key generation for correlated Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3373–3391, 2012.

[18] R. A. Chou, M. R. Bloch, and E. Abbe, "Polar coding for secret-key generation," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6213–6237, 2015.

[19] J. M. Renes, R. Renner, and D. Sutter, "Efficient one-way secret-key agreement and private channel coding via polarization," in *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2013, pp. 194–213.

[20] S. Nitinawarat, C. Ye, A. Barg, P. Narayan, and A. Reznik, "Secret key generation for a pairwise independent network model," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6482–6489, 2010.

[21] S. Nitinawarat and P. Narayan, "Perfect omniscience, perfect secrecy, and steiner tree packing," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6490–6500, 2010.

[22] R. A. Chou and A. Yener, "Secret-key generation in many-to-one networks: An integrated game-theoretic and information-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5144–5159, 2019.

[23] R. A. Chou, "Unified framework for polynomial-time wiretap channel codes," *arXiv preprint arXiv:2002.01924*, 2020.

[24] J. Benaloh and J. Leichter, "Generalized secret sharing and monotone functions," in *Conference on the Theory and Application of Cryptography*. Springer, 1988, pp. 27–35.

[25] M. Hayashi and T. Tsurumaru, "More efficient privacy amplification with less random seeds via dual universal hash function," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 2213–2232, 2016.

[26] M. Ye and A. Barg, "Universal source polarization and an application to a multi-user problem," in *Proc. of the Annual Allerton Conf. on Communication, Control, and Computing*, 2014, pp. 805–812.

[27] E. Arikan, "Source polarization," in *Proc. of IEEE Int. Symp. Inf. Theory*, 2010, pp. 899–903.

[28] R. A. Chou and M. R. Bloch, "Polar coding for the broadcast channel with confidential messages: A random binning analogy," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2410–2429, 2016.

[29] R. A. Chou, M. R. Bloch, and J. Kliewer, "Empirical and strong coordination via soft covering with polar codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5087–5100, 2018.

[30] R. A. Chou and M. R. Bloch, "Using deterministic decisions for low-entropy bits in the encoding and decoding of polar codes," in *Proc. of the Annual Allerton Conf. on Communication, Control, and Computing*, 2015, pp. 1380–1385.

[31] D. Aldous, "Random walks on finite groups and rapidly mixing markov chains," in *Séminaire de Probabilités XVII 1981/82*. Springer, 1983, pp. 243–297.

[32] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Annual Cryptology Conference*. Springer, 2012, pp. 294–311.

[33] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 143–154, 1979.