Separate Enciphering of Correlated Messages for Confidentiality in Distributed Networks

William Luh
Department of Electrical and Computer Engineering
Texas A&M University, College Station, Texas 77843
Email: luh@ece.tamu.edu

Deepa Kundur
Department of Electrical and Computer Engineering
Texas A&M University, College Station, Texas 77843
Email: deepa@ece.tamu.edu

***Abstract*—This paper studies distributed joint secrecy and compression suitable for sensor networks. A capacity region that characterizes the tradeoff between compression and secrecy is derived. We demonstrate for the two-node case that under the restriction of separate enciphering (i.e., no inter-node collaboration) unconditional secrecy by both parties cannot be achieved simultaneously. A fundamental design rule for lightweight encoder implementation critical for secrecy based on distributed source coding using syndromes and Reed-Solomon codes is presented highlighting practical feasibility.**

I. Introduction and Background

Sensor networks are characterized, in part, by their densely distributed and resource-constrained nodes. Accordingly, much recent theoretical and algorithmic research has been dedicated to the problem of nodes compressing correlated data separately without collaboration, known as distributed source coding [1]. Often, the sensor data collected may be of a sensitive nature requiring confidentiality; a common threat model for sensor networks involves a small subset of adversaries who each intercepts one transmit stream, and correspondingly collude to deduce some aspects of the sensed information. In this paper, we therefore consider the additional secrecy requirement, in which nodes must separately encipher *and* compress their data such that adversaries collectively learn as little as possible. This problem is made difficult by the assumption that sensors are not to use any cryptographic keys for reasons of cost, and all channels are noiseless in the sense that wiretap codes [2] cannot be used.

The problem addressed in this work was first addressed by Deslauriers in [3] and termed *distributed encryption*. Deslauriers tacitly assumes the use of deterministic encoders, and deduced that perfect (or unconditional) secrecy cannot be achieved. In the proposed work we derive the complete capacity region (absent in [3]), which characterizes the optimal secrecy and compression tradeoffs, while also considering stochastic encoders in the necessity argument (converse theorem). Our conclusions are the same as Deslauriers: perfect secrecy is not achievable even with stochastic encoding (not considered in [3]). Since perfect secrecy is not achievable, we define a practical measure of secrecy based on weakly secure network codes [4]. Although it turns out that any good Slepian-Wolf (SW) code satisfies the optimal secrecy-compression tradeoff that we derive in this paper, we will show that *only* non-systematic SW codes may satisfy our practical measure of secrecy. We also provide the first (deterministic) lightweight encoder implementation to this problem based on cyclic code circuit implementations.

Naturally, our work uses results from the distributed source coding community [1], [5], [6]. The distributed encryption problem is also analogous to traditional secret sharing [7]–[9] or the wiretap channel II [10], [11] in that the adversaries are restricted to have access to only a small subset of the encoded data thus hindering their goal. The distributed encryption problem is distinct because there is more than one "secret" to protect, these secrets are correlated, and the problem is distributed in nature. As a result the well-known security strategies of [7]–[11] such as inter-node collaboration, shared cryptographic keys, and common randomness that are used to achieve perfect secrecy do not apply to the addressed formulation. For example, in [10] a random codeword is chosen from a coset to confuse the adversary; in our problem randomness is not centralized, nor coordinated across the different sensors, and so wiretap channel II codes cannot be employed.

The advantage of the proposed joint enciphering and compression, is that it introduces no additional encoding complexity beyond what is incurred by distributed source coding. Hence if resources are available, traditional key-based cryptography can be applied after the proposed encoding.

In Sect. II, we summarize notation and formulate the distributed encryption problem. Our first contribution, in which we derive optimal secrecy-compression tradeoffs (or capacity region) is presented in Sect. III. Our second contribution is detailed in Sect. IV where we define what it means for a practical scheme to be secure, and outline a design rule for lightweight encoder implementation.

II. Preliminaries

A. Notation

Unless otherwise stated, let upper-case letters denote random variables, e.g. X, caligraphic upper-case letters denote finite sets, e.g. $\mathcal{X}$, lower-case letters denote realizations, e.g. x, and superscripted letters denote vectors, e.g. x^n. The probability mass function (pmf) is denoted using P_X. A Markov chain X, Y, Z in that order is denoted $X \leftrightarrow Y \leftrightarrow Z$ if and only if the joint pmf can be factored as $P_{X,Y,Z} = P_{X|Y} P_{Z|Y}$.

1930-529X/07/$25.00 © 2007 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2007 proceedings.

$H(X)$ is the entropy of X, $H(X|Y)$ is the conditional entropy of X given Y, and $I(X;Y)$ is the mutual information between X and Y [12]. Matrices are given by upper-case bold letters, e.g. $\mathbf{A}$.

B. Problem Formulation

Let $S_A^k \in \mathcal{S}_A^k$ and $S_B^k \in \mathcal{S}_B^k$ denote Alice and Bob's messages resp., which are generated by the joint discrete memoryless source (DMS) given by (1).

$$P_{S_A,S_B}^k(s_A^k, s_B^k) = \prod_{i=1}^{k} P_{S_A,S_B}(s_{A,i}, s_{B,i}) \tag{1}$$

Our problem is summarized in Fig. 1. Alice and Bob are to encipher their S_A^k, S_B^k *separately without cooperation* creating $X_A^n \in \mathcal{X}_A^n$ and $X_B^N \in \mathcal{X}_B^N$ resp. (note, n and N may be different, and N is not a RV). The base station receives both X_A^n and X_B^N, and its goal is to reconstruct S_A^k and S_B^k with negligible probability of error. Let the triple (f_A, f_B, φ) denote Alice's (possibly stochastic) encoder, Bob's (possibly stochastic) encoder, and the base station's decoder resp. Here $f_A : \mathcal{S}_A^k \to \mathcal{X}_A^n$, $f_B : \mathcal{S}_B^k \to \mathcal{X}_B^N$, and $\varphi : \mathcal{X}_A^n \times \mathcal{X}_B^N \to \mathcal{S}_A^k \times \mathcal{S}_B^k$. Also $P_{f_A(S_A^k),f_B(S_B^k)|S_A^k,S_B^k} = P_{f_A(S_A^k)|S^{k}{}_A} P_{f_B(S_B^k)|S_B^k}$ from the separate encoding requirement, and hence $X_B^N \leftrightarrow S_B^k \leftrightarrow S_A^k \leftrightarrow X_A^n$ forms a Markov chain.[1] The rate of Alice and Bob's enciphered messages are defined as

$$R_A \triangleq \frac{\log_2 \|f_A\|}{k} \tag{2}$$

$$R_B \triangleq \frac{\log_2 \|f_B\|}{k}. \tag{3}$$

Here $\|f_A\|$ denotes the number of possible outputs from Alice's encoder, and similarly $\|f_B\|$ for Bob's encoder.

In Fig. 1, the eavesdropper Eve is allowed to select either X_A^n or X_B^N, but not both. Depending on which enciphered message Eve selects, the equivocation rate of Eve with respect to (w.r.t.) Alice and Bob are defined as

$$\Delta_A \triangleq \frac{H(S_A^k|X_A^n)}{k} \tag{4}$$

$$\Delta_B \triangleq \frac{H(S_B^k|X_B^N)}{k}. \tag{5}$$

The idea is that if Eve intercepts Alice's X_A^n, then she is *only* interested solely in Alice's S_A^k and not Bob's S_B^k, and vice versa.[2] Equivocation rates of $\Delta_A = H(S_A)$ for Alice and $\Delta_B = H(S_B)$ for Bob are desired as this implies $H(S_j^k|X_j^n) = kH(S_j) = H(S_j^k)$ for $j = A$ or $j = B$, which means Eve is no better off with X_j^n than she was without it.

We say a quadruple (d_A, d_B, r_A, r_B) (corresponding to $(\Delta_A, \Delta_B, R_A, R_B)$) is *achievable* if there exists a (f_A, f_B, φ) such that for all $\epsilon > 0$ (arbitrarily small) and k sufficiently large the following are satisfied:

$$\begin{aligned} Pr\{(S_A^k, S_B^k) \neq (\hat{S}_A^k, \hat{S}_B^k)\} &\leq \epsilon && (6)\\ R_A &\leq r_A + \epsilon && (7)\\ R_B &\leq r_B + \epsilon && (8)\\ d_A - \epsilon \leq \Delta_A &\leq d_A && (9)\\ d_B - \epsilon \leq \Delta_B &\leq d_B && (10) \end{aligned}$$

where

$$\begin{aligned} X_A^n &= f_A(S_A^k) && (11)\\ X_B^N &= f_B(S_B^k) && (12)\\ (\hat{S}_A^k, \hat{S}_B^k) &= \varphi(X_A^n, X_B^N) && (13) \end{aligned}$$

and the separate enciphering constraint is enforced through the Markov chain (easily proved)

$$X_B^N \leftrightarrow S_B^k \leftrightarrow S_A^k \leftrightarrow X_A^n. \tag{14}$$

In addition, all parties, Alice, Bob, Eve, and base station have complete knowledge of $(f_A,\ f_B, \varphi)$.

III. The Capacity Region

The capacity region $\mathcal{R}$, defined to be the closure of the set of rate quadruples (d_A, d_B, r_A, r_B) that are achievable (see Section II-B), is described in Theorem 1 for the general distributed encryption problem.

Theorem 1: The capacity region is the closure of the union of all (d_A, d_B, r_A, r_B) satisfying

$$\begin{aligned} d_A + d_B &\leq I(S_A; S_B) && (15)\\ r_A &\geq H(S_A|S_B) && (16)\\ r_B &\geq H(S_B|S_A) && (17)\\ r_A + r_B &\geq H(S_A, S_B) && (18)\\ r_A + d_A &\geq H(S_A) && (19)\\ r_B + d_B &\geq H(S_B). && (20) \end{aligned}$$

[1]Instead of having f_A and f_B output non-negative integers as is usually the case in a source coding problem formulation, f_A and f_B output blocks of symbols to match the practical implementation in Sect. IV.

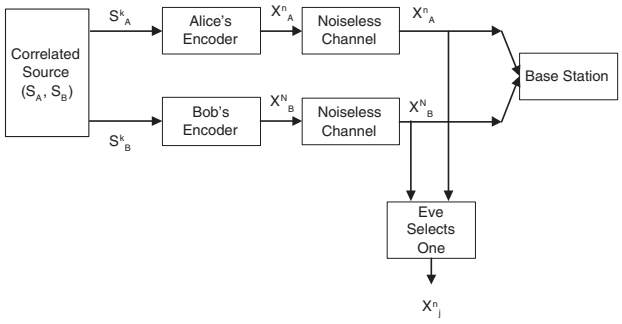

Fig. 1. Separate Enciphering by Alice and Bob with Eavesdropping by Eve

[2]This may seem a strange assumption, however it is similar to [13] in which a legitimate subscriber has complete knowledge of one random vector, but is to be kept ignorant about another correlated random vector. To abandon this assumption, it is necessary that $H(S_A) = H(S_B)$, which is usually the case in a sensor network environment.

1930-529X/07/$25.00 © 2007 IEEE

Theorem 1 also gives us the impossibility result that unconditional secrecy cannot be achieved by both Alice and Bob simultaneously. For Alice and Bob to both achieve unconditional secrecy, it would be necessary to have $d_A + d_B = H(S_A) + H(S_B)$, but (15) denies this since $I(S_A; S_B) \leq H(S_A, S_B) \leq H(S_A) + H(S_B)$.

A. Proof of Converse of Theorem 1

Assume that some (d_A, d_B, r_A, r_B) is achievable such that (6) to (13) are satisfied with the Markov constraint in (14). Then we shall show that the following bounds of (15) to (20) for *all* $\epsilon > 0$ are necessarily true.

First we prove (15) is necessarily true. To make use of (6), we call upon Fano's inequality

$$H(S_A^k, S_B^k | \hat{S}_A^k, \hat{S}_B^k) \leq h(\epsilon) + \epsilon k \log_2 |\mathcal{S}_A||\mathcal{S}_B| \triangleq k\epsilon_k \quad (21)$$

where $Pr\{(S_A^k, S_B^k) \neq (\hat{S}_A^k, \hat{S}_B^k)\} < \epsilon$, $\epsilon_k \to 0$ as $k \to \infty$ and $h(p)$ is the binary entropy function defined as

$$h(p) = -p \log_2 p - (1-p) \log_2 (1-p). \quad (22)$$

We state the following well-known lemma for Markov chains whose simple proof can be derived by the reader or be found in [2].

Lemma 1: If $X \leftrightarrow Y \leftrightarrow Z$ forms a Markov chain, then

$$I(X; Y|Z) = I(X; Y) - I(X; Z). \quad (23)$$

Now we proceed with the proof and write $H(S_A^k, S_B^k)$ in two ways. First we have

$$\begin{aligned} H(S_A^k, S_B^k) &= H(S_A^k) + H(S_B^k) - I(S_A^k; S_B^k) \\ &= H(S_A^k) + H(S_B^k) - kI(S_A; S_B). \end{aligned} \quad (24)$$

We can also write

$$\begin{aligned} H(S_A^k, S_B^k) &= H(S_A^k, S_B^k | \hat{S}_A^k, \hat{S}_B^k) + I(S_A^k, S_B^k; \hat{S}_A^k, \hat{S}_B^k) \\ &\leq I(S_A^k, S_B^k; X_A^n, X_B^N) + k\epsilon_k \end{aligned} \quad (25)$$

due to Fano's inequality. Applying the chain rule gives

$$\begin{aligned} & I(S_A^k, S_B^k; X_A^n, X_B^N) \\ = \; & I(S_A^k; X_A^n, X_B^N) + I(S_B^k; X_A^n, X_B^N | S_A^k) \\ = \; & I(S_A^k; X_A^n) + I(S_A^k; X_B^N | X_A^n) + I(S_B^k; X_A^n, X_B^N | S_A^k) \\ = \; & I(S_A^k; X_A^n) + I(S_A^k; X_B^N | X_A^n) \\ & + I(S_B^k; X_A^n | S_A^k) + I(S_B^k; X_B^N | S_A^k, X_A^n) \\ = \; & I(S_A^k; X_A^n) + I(S_A^k; X_B^N | X_A^n) \\ & + I(S_B^k; X_B^N | S_A^k, X_A^n) \end{aligned} \quad (26)$$

since $I(S_B^k; X_A^n | S_A^k) = 0$. The original Markov chain of (14) induces $X_B^N \leftrightarrow S_B^k \leftrightarrow (S_A^k, X_A^n)$, and using Lemma 1 we obtain

$$\begin{aligned} I(S_B^k; X_B^N | S_A^k, X_A^n) &= I(S_B^k; X_B^N) - I(X_B^N; S_A^k, X_A^n) \\ &= I(S_B^k; X_B^N) - I(X_A^n; X_B^N) \\ &\quad - I(S_A^k; X_B^N | X_A^n) \end{aligned} \quad (27)$$

where the final equality made use of the chain rule again. Therefore applying (27) to (26) provides

$$\begin{aligned} & I(S_A^k, S_B^k; X_A^n, X_B^N) \\ = \; & I(S_A^k; X_A^n) + I(S_B^k; X_B^N) - I(X_A^n; X_B^N) \\ = \; & H(S_A^k) - H(S_A^k | X_A^n) + H(S_B^k) \\ & - H(S_B^k | X_B^N) - I(X_A^n; X_B^N) \\ \leq \; & H(S_A^k) - H(S_A^k | X_A^n) + H(S_B^k) - H(S_B^k | X_B^N) \end{aligned} \quad (28)$$

Combining (24), (25), (28) and using the definition of equivocation rate (see (4) and (5)) gives the desired upper bound of (15). Equations (16) to (18) are a result of the SW theorem [5], and (19) and (20) follow simply from the chain rule:

$$\begin{aligned} H(S_A^k) &= kH(S_A) \\ &\leq H(S_A^k, X_A^n) \\ &= H(X_A^n) + H(S_A^k | X_A^n) \\ &\leq \log_2 \|f_A\| + H(S_A^k | X_A^n). \end{aligned} \quad (29)$$

Now dividing by k and using the definitions for rate and equivocation rate (see (2) and (4)), and then using the definition of achievability of these rates (see (7) and (9)) results in

$$\begin{aligned} H(S_A) &\leq R_A + \Delta_A \\ &\leq (r_A + \epsilon) + d_A \end{aligned} \quad (30)$$

which is the desired (19) by letting $\epsilon \to 0$. Bob's rate-equivocation sum, (20), follows in the same way. ■

Remark: The same result can be obtained for deterministic encoders by using $H(X_A^n | S_A^k) = H(X_B^N | S_B^k) = 0$. However, the proof of (15) makes no such assumption so as to include the possibility of stochastic encoding. Since both deterministic and stochastic encoding regions turn out to be equal, the rest of the paper deals explicitly with deterministic encoding.

B. Proof Sketch of Direct Part of Theorem 1

The bounds of (16) to (18) are again from the SW theorem. For simplicity we approximate the size and probabilities of strongly typical sets and their members resp. for a block length $k \to \infty$. Using the SW theorem proof in [12], Alice has approx. 2^{kR_A} bins containing typical strings s_A^k. Since the assignment of s_A^ks to the bins are uniform, asymptotically we expect approx. $2^{kH(S_A)}/2^{kR_A}$ typical strings per bin. This means if Alice sends a bin index using kR_A bits or a rate of R_A, Eve is confused as to which of the $2^{kH(S_A)}/2^{kR_A}$ strings is the correct one. As the typical strings are almost uniformly distributed in the asymptotic scenario, Eve's equivocation rate is $\Delta_A = H(S_A) - R_A$. Bob performs the same encoding using a different set of bins and so $\Delta_B = H(S_B) - R_B$. This immediately shows $\Delta_A + R_A = H(S_A)$ and $\Delta_B + R_B = H(S_B)$ ((19) and (20)) are achievable. The SW theorem allows us to achieve approximately $R_A + R_B = H(S_A, S_B)$, and so we also have $\Delta_A + \Delta_B = H(S_A) + H(S_B) - H(S_A, S_B) = I(S_A; S_B)$, which gives (15). ■

The direct part of the proof cannot be implemented in practice, as the bin-based encoders require large look-up table, countering our efforts for efficient implementation.

1930-529X/07/$25.00 © 2007 IEEE

IV. Practical Distributed Encryption Codes

In this section we derive low-cost implementations based on linear distributed source codes.

A. Operational Overview

Assume that all strings of a fixed-length over $GF(q)$ are equiprobable, and that the Hamming distance between Alice and Bob's strings (not exceeding a threshold t) is the correlation model. Recall that in a distributed source code (DSC), linear codes may be used to partition the space of fixed-length strings into bins analogous to the SW proof in [12].

To encode, let $\mathbf{G} = \left(\frac{\mathbf{G}_1}{\mathbf{G}_2}\right)$ be a $h \times k$ generator matrix of a t-error-correcting linear code such that $\mathbf{G}_1$ is $h_1 \times k$ and $\mathbf{G}_2$ is $h_2 \times k$ and $h_1 + h_2 = h$. The corresponding parity check matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ may be used to simultaneously compress and secure (to be defined in the next subsection) Alice and Bob's strings, e.g. the output of the encoders are $\mathbf{H}_1 s_A^k$ and $\mathbf{H}_2 s_B^k$, where $d_H(s_A^k, s_B^k) \leq t$, and $d_H(\cdot,\cdot)$ is the Hamming distance.

Since any t symbols may differ between S_A^k and S_B^k, $H(S_A|S_B) = \frac{t}{k}\log_2 q$ when $\mathcal{S}_A = \mathcal{S}_B = GF(q)$. It can be shown that the rates are $R_A = \frac{k-h_1}{k}\log_2 q$, and $R_B = \frac{k-h_2}{k}\log_2 q$ (these rates should satisfy the SW theorem before proceeding) while the equivocation rates are $\Delta_A = \frac{h_1}{k}\log_2 q$ and $\Delta_B = \frac{h_2}{k}\log_2 q$ [10]. Therefore the equivocation sum is $\Delta_A + \Delta_B = \frac{h}{k}\log_2 q$ since $h_1 + h_2 = h$, however the upper bound in (15) is $I(S_A; S_B) = H(S_A) - H(S_A|S_B) = (1 - \frac{t}{k})\log_2 q$. It can be shown that this scheme does not achieve the optimal rate-equivocation tradeoff when a bounded distance decoder is used. For example, $t \leq \frac{k-h}{2}$ (the Singleton bound), and since $k \geq h$, $I(S_A; S_B) \geq \frac{(\frac{k+h}{2})}{k}\log_2 q \geq \frac{h}{k}\log_2 q = \Delta_A + \Delta_B$ follows.

To decode, the base station receives both syndrome vectors, and looks in both cosets to find the two words that satisfy the Hamming distance t bound. Since the super-code is t-error-correcting, [6] showed there is a unique pair.[3]

B. Secrecy Model for Practical Implementation

In the previous section we reviewed linear distributed source coding without considering a measure of secrecy. In this section we propose the practical measure of secrecy used in the rest of this paper.

Suppose Eve intercepts Alice's $\mathbf{H}_1 s_A^k$. Eve should not be able to solve for any symbol in s_A^k. Furthermore, if Eve guesses g symbols of s_A^k correctly, it is desirable that Eve not learn any additional symbols from s_A^k. Surprisingly, a condition to prevent against this form of cryptanalysis is related to the weakly secure network coding problem [4]. We give a simplified condition pertaining to our distributed encryption problem.

Lemma 2: Let I_i be a $1 \times k$ vector of 0s except the i^{th} position is a 1. If Eve's number of guesses g is restricted to $g < k - h_1$ and $g < k - h_2$ for Alice and Bob's messages resp., and

$$\left(\frac{\mathbf{H}_1}{\frac{I_{j_1}}{\vdots}\atop I_{j_{h_1}}}\right) \tag{31}$$

is full rank for any distinct h_1 I_is, then Eve's guesses do not provide her with anymore information; similarly for $\mathbf{H}_2$ using any distinct h_2 I_is. Furthermore, Eve cannot solve for any symbols when she does not guess. (Theorems 2, Lemmas 2 and 3 from [4]).

We easily extended Lemma 2 to Proposition 1.

Proposition 1: The full rank condition of (31) is satisfied iff any h_1 columns of $\mathbf{H}_1$ has a non-zero determinant, and any h_2 columns of $\mathbf{H}_2$ has a non-zero determinant

Proof: We will show this is true for $\mathbf{H}_1$ and the same line of reasoning follows for $\mathbf{H}_2$. Assume that $\mathbf{H}_1$ fails Lemma 2. Then there exists h_1 distinct I_is, $\{I_{j_1}, \ldots, I_{j_{h_1}}\}$, such that for some $(\beta_1, \ldots, \beta_{k-h_1}) \in GF(q)^{k-h_1} - \{(0, \ldots, 0)\}$ and $(\gamma_1, \ldots, \gamma_{k-h_1}) \in GF(q)^{k-h_1} - \{(0, \ldots, 0)\}$ we have

$$(\beta_1, \ldots, \beta_{k-h_1})\mathbf{H}_1 = \gamma_1 I_{j_i} + \cdots + \gamma_{h_1} I_{j_{h_1}} \tag{32}$$

i.e. $\{I_{j_1}, \ldots, I_{j_{h_1}}\}$ is in the row space of $\mathbf{H}_1$. This is then equivalent to

$$(\beta_1, \ldots, \beta_{k-h_1})\mathbf{H}_1^{[\{1,\ldots,k\}-\{j_1,\ldots,j_{h_1}\}]} = (0, \ldots, 0) \tag{33}$$

where $\mathbf{H}_1^{[\{1,\ldots,k\}-\{j_1,\ldots,j_{h_1}\}]}$ represents the $k - h_1$ columns with indices *not* from $\{j_1, \ldots, j_{h_1}\}$. There is a non-trivial solution for $(\beta_1, \ldots, \beta_{k-h_1})$ only if $\det \mathbf{H}_1^{[\{1,\ldots,k\}-\{j_1,\ldots,j_{h_1}\}]} = 0$.

Conversely, if there exists $k - h_1$ columns of $\mathbf{H}_1$ that has a zero determinant, then there exists $(\beta_1, \ldots, \beta_{k-h_1}) \in GF(q)^{k-h_1} - \{(0, \ldots, 0)\}$, such that a linear combination of the rows of $\mathbf{H}_1$ using these β_is as the scalars will result in a vector with 0s in precisely the positions of these $k - h_1$ columns (since a square matrix has a trivial or zero nullspace iff it is full rank [14]). This means the other h_1 positions can actually be written as a linear combination of h_1 I_is of the same h_1 positions, and hence $\mathbf{H}_1$ fails Lemma 2. ■

C. Partitioning a Channel Code for Secrecy

First we show the importance of properly partitioning a generator matrix for secrecy by showing a "bad" partition.

Proposition 2: If $\mathbf{G}_1$ contains any column of 0s then Alice's encoder is not secure. If $\mathbf{G}_2$ contains any column of 0s then Bob's encoder is not secure.

Proof: Suppose the i^{th} column in $\mathbf{G}_1$ is a 0-column. Then the row space of $\mathbf{H}_1$ contains I_i, since the row space of $\mathbf{H}_1$ is the dual space of $\mathbf{G}_1$, which contains those vectors orthogonal to all vectors in the row space of $\mathbf{G}_1$. Therefore $\mathbf{H}_1$ fails Lemma 2. ■

Proposition 2 immediately disqualifies the partitioning technique found in [15], since [15] partitions a systematic generator matrix $\mathbf{G}$; no matter how a systematic matrix is partitioned any partition will always have a 0-column.

[3] In this paper we do not address decoding complexity or the decoding algorithm since our goal is inexpensive implementation of the encoder.

1930-529X/07/$25.00 © 2007 IEEE

Towards this end we informally show that Reed-Solomon (RS) codes can always be used when partitioned properly. RS codes are cyclic, and so Alice and Bob can take advantage of the simple cyclic code circuits implementations.

Proposition 3: Reed-Solomon codes may be used for secure distributed encryption codes as defined by Lemma 2 when they are partitioned appropriately.

Instead of giving a formal proof, we shall give an example, which will illustrate the ideas. First, recall that the dual code of a RS code is another RS code. It will be easiest to work with the dual code. Let us take the $(15, 11)$ RS code over $GF(16)$. A parity check matrix of this code is given by $\mathbf{A}$,

$$\mathbf{A} = \begin{pmatrix} 1 & \xi & \dots & \xi^{14} \\ 1 & \xi^2 & \dots & \xi^{28} \\ 1 & \xi^3 & \dots & \xi^{42} \\ 1 & \xi^4 & \dots & \xi^{56} \end{pmatrix} \tag{34}$$

where ξ is a primitive element in $GF(16)$. If instead we use $\mathbf{A}$ as a generator matrix, i.e. $\mathbf{G} = \mathbf{A}$, then we still have a RS code that is now $(15, 4)$ with minimum distance 12, and so can correct up to $t = 5$ errors. Therefore let us assume that Alice and Bob process 15 symbols over $GF(16)$, S_A^{15} and S_B^{15} resp. such that $d_H(S_A^{15}, S_B^{15}) \leq t = 5$. Let $\mathbf{G}_1$, $\mathbf{G}_2$ be the top two rows and bottom two rows of $\mathbf{A}$ resp.; the reader can check that the resulting rates satisfy the SW theorem. Then the parity check polynomial $h_1(x)$ is the recipricol of $(x - \xi)(x - \xi^2)$, i.e. $h_1(x) = (x - \xi^{-1})(x - \xi^{-2})$, and similarly $h_2(x) = (x - \xi^{-3})(x - \xi^{-4})$ [16]. This means the generator polynomials are $g_1(x) = \frac{x^{15}-1}{(x-\xi^{-1})(x-\xi^{-2})}$ and $g_2(x) = \frac{x^{15}-1}{(x-\xi^{-3})(x-\xi^{-4})}$ since RS codes are cyclic. This also means that $g_1(x)$ and $g_2(x)$ are generator polynomials for RS codes, since both polynomials have consecutive powers of ξ as roots, which by definition is a RS code.[4] Finally since both partitions are RS codes, any parity check matrices $\mathbf{H}_1$ (resp. $\mathbf{H}_2$) will have the property that any h_1 (resp. h_2) columns results in a non-singular square matrix [16], and so satisfies Proposition 1, making $\mathbf{H}_1$ and $\mathbf{H}_2$ secure matrices. Finally we could put $\mathbf{H}_1$ and $\mathbf{H}_2$ in a nonsysematic cyclic form [16] to take advantage of a feed-forward shift-register circuit implementation. Alternatively, at a higher cost but faster, we could also put $\mathbf{H}_1$ and $\mathbf{H}_2$ in a form that uses the partial syndrome circuit implementation. The base station's decoder would have to be modified to correspond to the encoder implementation of course.

Before leaving this example, we note that had we not started with the dual form of matrix $\mathbf{A}$, partitioning any RS generator matrix *may not* lead to RS code partitions. This is seen easily by noting that if we had started with a systematic generator matrix for a RS code and partitioned it as in [15], then the resulting partitions will not be secure viz. Proposition 2. This means the dual space of the partitioned codes do not have parity check matrices satisfying Proposition 1, which implies that the partitions are not RS codes.

[4]Note that in the SW code construction of [6] it does not matter what error correcting capabilities the two subcodes $\mathbf{G}_1$, $\mathbf{G}_2$ have; only the error correcting capabilities of the super-code $\mathbf{G}$ is important.

To summarize, the general idea used in the example above is to find a generator matrix for a RS code such that partitioning it by separating the top and bottom rows will still result in individual RS codes. This means the two newly created RS codes have parity check matrices that satisfy the non-zero determinant condition required in Proposition 1. We demonstrated a simple way to ensure this partitioning will always be secure by starting with a parity check matrix of a specific form, e.g. $\mathbf{A}$ (always exists for BCH and RS codes) of a RS code, and letting it be the generator matrix instead, i.e. the dual code. These ideas along with Proposition 3 will be formalized and generalized in a future paper.

V. Conclusion and Future Work

We derived the capacity region for the distributed encryption problem, as well as provided a practical design rule for economical encoder implementation based on distributed source codes using syndromes and Reed-Solomon codes.

Acknowledgment

The authors would like to thank Kapil Bhattad for engaging with us in helpful discussions, and the reviewers for their helpful comments.

References

[1] Z. Xiong, A. D. Liveris, and S. Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21:80–94, September 2004.

[2] A. D. Wyner. The wire-tap channel. *The Bell System Technical Journal*, 54(8):1355–1387, October 1975.

[3] I. Deslauriers. Distributed encryption and the Slepian-Wolf theorem. In *Canadian Conference on Electrical and Computer Engineering*, pages 93–97, Saskatoon, Sask., Canada, May 2005.

[4] K. Bhattad and K. R. Narayanan. Weakly secure network coding. In *Workshop on Network Coding, Theory, and Applications (NETCOD)*, pages 63–68, Riva Del Garda, Italy, April 2005.

[5] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inform. Theory*, 19(4):471–480, July 1973.

[6] S. Pradhan and K. Ramchandran. Distributed source coding: Symmetric rates and applications to sensor networks. In *Proc. DCC'00*, Snowbird, UT, March 2000.

[7] A. Shamir. How to share a secret. In *Communications of the ACM 22*, number 11, pages 612–613, November 1979.

[8] E. D. Karnin, J. W. Greene, and M. E. Hellman. On secret sharing systems. *IEEE Trans. on Information Theory*, 29(1):35–41, January 1983.

[9] H. Yamamoto. On secret sharing communication systems with two or three channels. *IEEE Trans. on Information Theory*, 32(3):387–393, May 1986.

[10] L. H. Ozarow and A. D. Wyner. Wire-tap channel II. *AT & T Bell Laboratories Technical Journal*, 63(10):2135–2157, 1984.

[11] Y. Luo, C. Mitrpant, and A. J. Han Vinck. Some new characters on the wire-tap channel of type II. *IEEE Trans. on Information Theory*, 51(3):1222–1229, March 2005.

[12] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 2nd edition, 1991.

[13] H. Yamamoto. A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers. *IEEE Trans. on Information Theory*, 29(6):918–923, November 1983.

[14] C. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics, 2001.

[15] V. Stanković, A. Liveris, Z. Xiong, and C. Georghiades. Design of Slepian-Wolf codes by channel code partitioning. In *Proc. DCC'04*, Snowbird, UT, March 2004.

[16] R. E. Blahut. *Algebraic Codes for Data Transmission*, chapter 5, page 112. Cambridge University Press, New York, 2nd edition, 2003.