Weakly Secure Data Exchange with Generalized Reed Solomon Codes

Muxi Yan*, Alex Sprintson*, and Igor Zelenko†
*Department of Electrical and Computer Engineering, Texas A&M University
†Department of Mathematics, Texas A&M University

***Abstract*—We focus on secure data exchange among a group of wireless clients. The clients exchange data by broadcasting linear combinations of packets over a lossless channel. The data exchange is performed in the presence of an eavesdropper who has access to the channel and can obtain all transmitted data. Our goal is to develop a *weakly secure* coding scheme that prevents the eavesdropper from being able to decode any of the original packets held by the clients. We present a randomized algorithm based on Generalized Reed-Solomon (GRS) codes. The algorithm has two key advantages over the previous solutions: it operates over a small (polynomial-size) finite field and provides a way to verify that constructed code is feasible. In contrast, the previous approaches require exponential field size and do not provide an efficient (polynomial-time) algorithm to verify the secrecy properties of the constructed code. We formulate an algebraic-geometric conjecture that implies the correctness of our algorithm and prove its validity for special cases. Our simulation results indicate that the algorithm is efficient in practical settings.**

I. Introduction

In the Weakly Secure Data Exchange (WSDE) problem [1] a group of wireless clients C need to exchange a set of packets X using a lossless broadcast channel accessible to all clients. Initially, each client holds a subset of the packets and needs to obtain the rest of the packets in X. The data exchange is performed over multiple rounds; in each round a client broadcasts a linear combination of packets to other clients. All transmissions are observed by an eavesdropper whose goal is to decode one or more packets from X using the transmitted data as well as its side information. Our goal is to design a coding scheme that allows each client to obtain all packets in X while protecting these packets from the eavesdropper.

Fig. 1 presents an example of an instance of the WSDE problem. In this example the clients are exchanging packets in the set $X = \{x_1, \dots, x_5\}$, such that clients 1, 2, and 3 have subsets of packets $\{x_1, x_2, x_3\}$, $\{x_2, x_3, x_4\}$, and $\{x_3, x_4, x_5\}$ available to them at the beginning of data exchange, respectively. A possible solution to this problem is for clients 1,2, and 3 to broadcast linear combinations $x_1+2x_2+x_3$, $x_2+2x_3+x_4$, and $x_3+2x_4+x_5$, respectively (all operations are over field $\mathbb{F}_5$ of size 5). Note that by observing the transmitted messages, the eavesdropper will not be able to decode any of the original packets.

We also consider a more general setting in which the eavesdropper might already have a set of packets $Z \subset X$ as a side information, in this case its goal is to obtain packets in $X \setminus Z$. In this setting, set Z is arbitrary, but its size is limited by a parameter $g < |X|$. For example, in Fig. 1 the eavesdropper will not be able to obtain any linear combination that contains two packets, so even if it has any single packet as a prior side information, it will not be able to obtain any additional packet from X.

In our previous work [1] we proposed two algorithms that provide weakly secure solutions for the WSDE problem. The main disadvantage of the proposed algorithms is that they require an exponential field size, which limits their applicability in practical settings. In addition, the algorithms presented in [1] do not provide an efficient means to verify whether the proposed solution is secure; they only guarantee that the solution is secure with high probability. In this work, we propose a new randomized algorithm which requires a small (polynomial size) field. The proposed algorithm uses a construction based on Generalized Reed-Solomon (GRS) codes. This construction achieves a major reduction in the search space. We present a combinatorial conjecture that implies that our construction succeeds with high probability.

Related work. The original Direct Data Exchange (DDE) problem was proposed by El Rouayheb et al. [2]. This paper established upper and lower bounds on the number of transmissions required for data exchange using linear network coding. Several solutions for the DDE problem have been proposed in [3], [4], and [5]. Courtade and Wesel [6] generalized the problem for settings with partial connectivity, i.e., settings in which some of the clients might not be able to receive messages due to the losses in the broadcast channel. References [7] and [8] considered settings with different

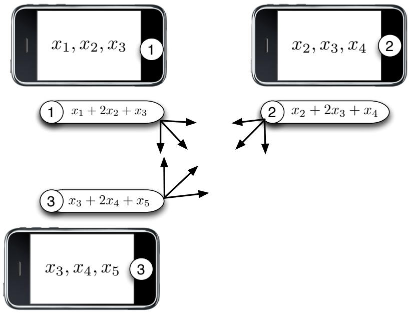

Fig. 1: Example of weakly secure solution against eavesdropper whose side information includes a single packet. All transmitted linear combinations are over field $\mathbb{F}_5$.

The work of Alex Sprintson was partially supported the NSF under grant CNS-0954153 and by the AFOSR under contract No. FA9550-13-1-0008.

978-1-4799-5186-4/14/$31.00 ©2014 IEEE

transmission costs.

Weakly secure network coding was introduced by Bhattad and Narayanan [9]. Their work focuses on weakly secure network coding for traditional (wired) networks and proposes an algorithm for multicast settings. In [10], Dau et al. considered the problem of finding weakly secure solutions for the *Index Coding* problem.

Halbawi et al. [11] used a similar approach for correcting adversarial errors in *simple multiple access networks*, i.e., networks in which a destination node receives information from multiple sources via a set of possibly adversarial relay nodes. A very recent independent work by Dau et al. [12] proposes a similar conjecture and shows its correctness for small instances.

II. Model and Problem Definition

An instance of the WSDE problem includes a set of k wireless clients, $C = \{c_i, i \in [k]\}$ and a set of n packets $X = \{x_i, i \in [n]\}$, where $[i]$ denotes the set $\{1, \cdots, i\}$. We assume that the packets in X are randomly and uniformly distributed over the underlying finite field $\mathbb{F}_q$ of size q. Initially, each client in C has access to a subset of packets $S_i \subseteq X$. We refer to S_i as the *side information* set of client c_i. For clarity of presentation, we assume that each packet belongs to the side information set of at least one client. The goal of the clients is to exchange data such that each one of them will be able to obtain all packets in X. To this end, the clients use a lossless channel which allows a client to broadcast data to all other clients in C. We use the term *message* to refer to the symbols transmitted over the broadcast channel, in contrast to *packet* that refers to the original uncoded packets $x_1, \ldots, x_n$. The communication is performed in rounds, such that at round i one of the clients, referred to as c_{t_i}, broadcasts message p_i which is a linear combination of packets that belong to its side information set S_{t_i}.[1] More specifically,

$$p_i = \sum_{j: x_j \in S_{t_i}} \gamma_{ij} x_j, \qquad (1)$$

where t_i is denoted as the index of the client that transmits in round i, and γ_{ij} is the coefficient of packet x_j in message p_i. For convenience we set $\gamma_{ij} = 0$ for all $j \in \{j, x_j \notin S_{t_i}\}$. Note that the transmitted message p_i in round i can be specified by a vector $\gamma_i = \begin{bmatrix} \gamma_{i1} & \gamma_{i2} & \ldots & \gamma_{in} \end{bmatrix}$. We refer to vector γ_i as the *encoding vector* of message p_i.

We denote by P the set of all messages transmitted, i.e.

$$P = \{p_1, p_2, \ldots, p_\mu\},$$

where μ is the total number of transmission rounds.

We also construct the *encoding matrix* Γ that includes vectors $\gamma_1, \gamma_2, \ldots, \gamma_\mu$ as rows, i.e. $\Gamma = \begin{bmatrix} \gamma_1^T & \gamma_2^T & \ldots & \gamma_\mu^T \end{bmatrix}^T$.

In addition, we denote by $u_i, i \in [n]$ the unit encoding vector that corresponds to a single packet x_i, i.e.,

$$u_i = [\underbrace{0 \quad \ldots \quad 0}_{i-1 \text{ zeros}} \quad 1 \quad \underbrace{0 \quad \ldots \quad 0}_{n-i \text{ zeros}}].$$

We denote by $U_i = \left[u_{i_1}^T, u_{i_2}^T, \ldots, u_{i_{|S_i|}}^T \right]^T$ the matrix whose rows are unit encoding vectors that correspond to the packets $x_{i_1}, x_{i_2}, \ldots, x_{i_{|S_i|}}$ in the side information set S_i of c_i.

An encoding scheme is feasible if all clients in C are able to decode all packets in X after the last transmission is completed. In other words, a feasible encoding scheme satisfies the following condition:

$$u_j \in \left\langle \begin{bmatrix} \Gamma \\ U_i \end{bmatrix} \right\rangle, \forall i \in [k] \text{ and } j \in [n],$$

where $\langle \cdot \rangle$ denotes the row space of a matrix. It also implies that for all $i \in [k]$, it holds that

$$\text{rank}\left(\begin{bmatrix} \Gamma \\ U_i \end{bmatrix} \right) = n.$$

Our goal is to design a *weakly secure* encoding scheme that protects the packets in X from the eavesdropper. In contrast to strongly (or information-theoretically) secure schemes whose goal is to prevent the eavesdropper from receiving any information about packets in X, the goal of weakly secure schemes is to protect individual packets. In this context, two settings can be considered. In the first settings, the eavesdropper has no prior side information about packets in X. In this setting, for each $u_i, i \in [n]$ it must hold that $u_i \notin \langle \Gamma \rangle$. The goal is to guarantee that the eavesdropper will not be able to obtain any of the original packets by observing the messages transmitted over the channel.

In the second setting, the eavesdropper might have some prior side information about packets in X. We denote by $Z \subseteq X$ the set of packets known to the eavesdropper and by $U_Z = \begin{bmatrix} u_{z_1}^T & \ldots & u_{z_{|Z|}} \end{bmatrix}^T$ the matrix that consists of unit encoding coefficients of these packets. The set Z can be arbitrary, but its size is limited by parameter g. The scheduler knows the value of g but does not know Z. The goal of the scheduler is to prevent the eavesdropper from being able to obtain any packet in $X \setminus U_Z$. Thus, a weakly secure scheme must satisfy the following condition: for each $u_i, i \in [n]$ it holds that

$$u_i \notin \left\langle \begin{bmatrix} \Gamma \\ U_Z \end{bmatrix} \right\rangle.$$

Let $w(\Gamma)$ be the minimum Hamming weight of a vector in the row space of encoding matrix Γ. It is easy to verify (see [1]) that a scheme that uses encoding matrix Γ is weakly secure against an adversary with side information set of size g if and only if $w(\Gamma) \geq g + 2$. Thus, our goal is to find an encoding matrix Γ that maximizes $w(\Gamma)$. Indeed, a scheme with larger value of $w(\Gamma)$ can protect the transmissions against adversaries that have larger side information sets.

The Direct Data Exchange (DDE) problem is formally defined as follows:

Problem DDE. *Find an encoding matrix $\Gamma \in \mathbb{F}_q^{\mu \times n}$ that satisfies the following conditions:*

1) *For each row γ_i in Γ, there exists $j \in [k]$ such that γ_i is a linear combination of vectors in U_j.*
2) *For all $i \in [k]$ it holds that*

$$\text{rank}\left(\begin{bmatrix} \Gamma \\ U_i \end{bmatrix} \right) = n. \qquad (2)$$

[1] In general, a message could be a linear function of side information of c_{t_i} and as well of the message previously transmitted over the channel. However, it is easy to verify that this does not provide any advantage in our setting.

3) *The number of rows μ of Γ is minimized.*

The first condition guarantees that at each round i, there exists a client c_{t_i} that can transmit message $p_i = \sum_{j:x_j \in S_{t_i}} \gamma_{ij} x_j$. The second condition guarantees that each client can decode all packets in X. Finally, the third condition ensures that the total number of transmission is minimum.

The DDE problem has been the subject of many previous works [3]–[6]. In particular, references [4], [5] present efficient polynomial-time solutions for this problem.

The Weakly Secure Data Exchange (WSDE) problem is formulated in a similar way, but instead of the condition (3) (minimizing μ), the goal is to maximize $w(\Gamma)$. In fact, as discussed in [1], these two problems are related in the way that a scheme that maximizes $w(\Gamma)$ also minimizes μ. Note that the Singleton bound implies that the maximum achievable value of $w(\Gamma)$ is bounded by $n - \mu + 1$.

In the following sections, we present an algorithm that obtains a solution for WSDE problem that requires a small (polynomial size) field.

III. Reduction

Our approach is to find find an optimal solution Γ to Problem DDE and then modify it to obtain a solution to Problem WSDE. We begin by creating a modified instance of the problem through several reduction steps. These reduction steps will provide us a network instance with more structure which simplifies our presentation. Our reduction satisfies the following conditions:

(C1) The minimum number of transmissions needed to satisfy the requests of all clients is the same for both instances (i.e., the optimal solution to Problem DDE has the same size for both original and modified instances).

(C2) Any solution Γ for the modified instance which is secure against an adversary with side information set of size g is also a secure solution for the original instance.

Our goal is to construct a modified instance that has a solution to Problem DDE with the following properties:

(P1) Each client either broadcasts a single message or never broadcasts a message;

(P2) Each client that never transmits has exactly $n - \mu$ packets in its side information set;

(P3) Each client that transmits has exactly $n - \mu + 1$ packets in its side information set.

Our reduction first finds a solution to Problem DDE for the given network instance. Then we apply several steps to modify the problem instance in order to satisfy properties (P1)-(P3).

The goal of the first step is to create a modified problem instance that satisfies property (P1). For each client c_i that makes $\mu_i > 1$ transmissions in the DDE solution, c_i is substituted by μ_i clients with the same side information set S_i such that each one of the clients transmits exactly one time. Note that this reduction step satisfies conditions (C1) and (C2).

The goal of the next step is to ensure that any client that does not transmit satisfies property (P2). For each client c_i with $\mu_i = 0$ and $|S_i| > n - \mu$, we find a subset of packets $S_i^* \subseteq S_i$ of size $|S_i^*| = n - \mu$ whose corresponding unit encoding vectors $U_i^* = \{u_j : x_j \in S_i^*\}$ satisfy $\dim(U_i^* \cup \{\gamma_1, \ldots, \gamma_\mu\}) = n$. Such subset can be found efficiently in polynomial time. We set side information of c_i as $S_i = S_i^*$. After replacement of side information of all clients that do not transmit, property (P2) is satisfied. Note that this procedure maintains conditions (C1) and (C2).

In the third step, we modify the side information and transmissions of clients that transmit once, which ensures that property (P3) holds. If for some client c_i it holds that $\mu_i = 1$ and $|S_i| > n - \mu + 1$, we find a subset of packets $S_i^* \subseteq S_i$ of size $|S_i^*| = n - \mu + 1$ whose corresponding unit encoding vectors $U_i^* = \{u_j : x_j \in S_i^*\}$ satisfy $\dim(U_i^* \cup \{\gamma_1, \ldots, \gamma_\mu\} \backslash \{\gamma_{j^*}\}) = n$, where γ_{j^*} is the message transmitted by client i, i.e., $t_{j^*} = i$. We set side information of c_i as $S_i = S_i^*$. Also, note that γ_{j^*} is in the linear span of $U_i^* \cup \{\gamma_1, \ldots, \gamma_\mu\} \setminus \{\gamma_{j^*}\}$. In other words, γ_{j^*} can be written as

$$\gamma_{j^*} = \sum_{j:u_j \in U_i^*} \beta_j u_j + \sum_{j \neq j^*} \beta_j' \gamma_j, \tag{3}$$

where β_j and β_j' are coefficients that belong to $\mathbb{F}_q$. We then modify γ_{j^*} to $\sum_{j:u_j \in U_i^*} \beta_j u_j$. Note that this replacement does not change row space of Γ as well as the total number of transmissions, so conditions (C1) and (C2) are satisfied. We satisfy property (P3) by applying this transformation to all clients in the network.

IV. Randomized Algorithm

In this section we present a randomized algorithm, referred to as Algorithm 1, for Problem WSDE. The first step of the algorithm is to modify the problem instance by applying the reduction presented in Section III. The resulting instance satisfies conditions (P1), (P2), and (P3). We denote by Γ the optimal solution to Problem DDE for this instance and by μ the corresponding number of transmissions. For $i \in [\mu]$, t_i denotes the client that transmits at round i according to this solution.

The next step is to randomly select values of $\alpha_1, \ldots, \alpha_n$ from the underlying field $\mathbb{F}_q$. The values are selected in such a way that $\alpha_i \neq \alpha_j$ for $i \neq j$. This is easy to implement, i.e., by first randomly selecting α_1 from $\mathbb{F}_q$, then selecting α_2 from $\mathbb{F}_q \setminus \{\alpha_1\}$ and so on.

We construct a Vandermonde matrix G with the parameters $\alpha_1, \ldots, \alpha_n$:

$$G(\alpha_1, \ldots, \alpha_n) = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1^{\mu-1} & \alpha_2^{\mu-1} & \ldots & \alpha_n^{\mu-1} \end{bmatrix}.$$

Note G is the generator matrix of a Generalized Reed-Solomon (GRS) code, which is a Maximum Distance Separable (MDS) code.

Next, we identify a $\mu \times \mu$ transform matrix T such that product $\Gamma' = [\gamma_{ij}'] = T \cdot G$ satisfies the following condition: for each $i \in [\mu]$ and $j \in [n]$: if $x_j \notin S_{t_i}$ then $\gamma_{ij}' = 0$. Note that the transform matrix T can be identified through elementary row operations and that T is unique up to a scalar multiplicative factor.

Finally, if T is a full rank matrix, then the algorithm returns $T \cdot G$ as a solution. Otherwise, the algorithm terminates with a failure. Alternatively, the algorithm can return to Step 3 and proceed with a different choice of $\alpha_1, \ldots, \alpha_n$, resulting in a Las Vegas type algorithm.

It is easy to verify that if T is a full rank matrix, then $\Gamma' = T \cdot G$ is an optimum WSDE solution. Indeed since G is a generator matrix of an MDS code, so is Γ'. Since every client in the network has at least $n - \mu$ side information packets, it can decode all the packets in X. Also, since Γ' satisfies the Singleton bound, the minimum Hamming weight of a vector in its row space is at least $n - \mu + 1$.

We note that Algorithm 1 utilizes a matrix completion approach. Indeed, we first obtain a solution with an encoding matrix Γ and the set of transmitting clients $\{t_i : i = 1, \ldots, \mu\}$ (client t_i is transmitting at round i). Our goal is then to complete the matrix Γ' subject to the following constraint:

$$\text{for all } i \in [\mu] \text{ and } j \in [n] \text{ if } x_j \notin S_{t_i} \text{ then } \gamma'_{ij} = 0. \quad (4)$$

Any other elements of Γ' can be assigned to any element of the finite field $\mathbb{F}_q$. Here, Γ' is an incomplete matrix that has zeros in places specified by (4). We complete Γ' by applying linear transformation T to an MDS code such that the resulting matrix $\Gamma' = [\gamma'_{ij}]$ satisfies (4).

Note that matrix T is uniquely determined by the incomplete matrix Γ; we denote it by T_Γ in the sequel. In particular, matrix T can be explicitly written in the following way. In matrix Γ, we replace all zero entries in the j'th column with α_j for all $j = 1, \ldots, n$. Let N_i be a subset of $\{\alpha_1, \ldots, \alpha_n\}$ consisting of those α_j's that appear in the i'th row. Then, we can write the (i, j)'th entry r_{ij} of T_Γ as follows:

$$r_{ij} = (-1)^{j-1} \sum_{A \subseteq N_i : |A| = j-1} \prod_{\ell : \alpha_\ell \in A} \alpha_\ell .$$

From now on incomplete encoding matrices Γ for given μ will be called *configurations of size μ*. Note that $\det(T_\Gamma)$ can be considered as a polynomial of $\alpha_1, \ldots \alpha_n$. Our goal is to characterize all configurations Γ for which $\det(T_\Gamma)$ is not (identically) equal to zero.

Definition 1. *We say that a configuration Γ satisfies* no rectangle condition *if it does not contain an $a \times b$ all zeros submatrix with $a + b = \mu + 1$.*

Note that there is another description of matrix T_Γ: if $P_i(\lambda)$ is a polynomial of degree $\mu - 1$ such that the set of roots equal to N_i, then the ith row of T_Γ is nothing but the row of the coefficients of this polynomial (from the lowest degree to the highest one). This description easily implies the following

Proposition 1. *If $\det(T_\Gamma) \neq 0$, then Γ satisfies no rectangle condition.*

Proof. Assume that Γ contains a $a \times b$ all zeros submatrix with $a + b = \mu + 1$ and let $i_1, \ldots i_a$ be the indices of rows involved in this submatrix. Then the polynomials $\{P_{i_s}\}_{s=1}^{a}$ have b common zeros and therefore a common factor of degree b (in λ). Dividing all these polynomials by this common factor we get a polynomials of degree $\mu - 1 - b = a - 2$. Since the vector space of polynomial of degree $a - 2$ is $(a-1)$-dimensional, the resulting a polynomials are linearly dependent. Therefore the original a polynomials are linearly dependent and therefore $\det(T_\Gamma) = 0$, which is a contradiction. □

Algorithm 1 Randomized Algorithm

```
1: Obtain reduced network instance and corresponding DDE solution $\Gamma \in \mathbb{F}_q^{\mu \times n}$ using the procedure described in Section III
2: Randomly choose $\alpha_1, \ldots, \alpha_n$ from underlying field $\mathbb{F}_q$ such that $\alpha_i \neq \alpha_j$ for $i \neq j$ and $\alpha_i \neq 0$ for all $i \in [n]$
3: Construct the Vandermonde matrix $G = [g_{ij}] \in \mathbb{F}_q^{\mu \times n}$ with parameters $\alpha_1, \ldots, \alpha_n$
4: Find a $\mu \times \mu$ matrix $T$ such that $\Gamma' = [\gamma'_{ij}] = TG$ satisfies the following condition: for each $i \in [\mu]$ and $j \in [n]$ if $x_j \notin S_{t_i}$ then $\gamma'_{ij} = 0$
5: If $T$ is a full-rank matrix then return $\Gamma'$
6: Otherwise terminate with a failure
```

Conjecture 1. *The converse of Proposition 1 holds, i.e. if Γ satisfies no rectangle condition then $\det(T_\Gamma) \neq 0$.*

This conjecture was also stated recently in [12]. In our previous work [1] we proved that if the matrix Γ satisfies no rectangle condition, then a random assignment of $\alpha_1, \ldots, \alpha_n$ yields, with high probability, a full rank matrix T_Γ. Thus, if the conjecture is true, our algorithm will succeed with high probability.

We proceed by presenting a partial inductive procedure. We use this procedure for proving the conjecture in the case $\mu = 3$ over finite field $\mathbb{F}_q$ with large enough field size q.

First, there is a simple description of $\det(T_\Gamma)$. Let $(i_0, i_1, \ldots i_{\mu-1})$ be a permutation of $(1, \ldots, \mu)$. Now we mark some zero entries in the configuration Γ according to the following rule: we do not mark any zero entry in i_0'th row, we mark one zero entry row i_1, two zero entries in row i_2 and so on. Assume that zero entries in Γ are replaced by α_j's as before. Then $(-1)^{\mu(\mu-1)/2} \det(T_\Gamma)$ is equal to the sum of all monomials obtained from the product of all α_j corresponding to the marked zeros with the coefficient equals to the sign of the permutation $(i_0, i_1, \ldots i_{\mu-1})$ (note that the sum is taken over all possible markings as above). Note that the same monomial may correspond to different marking and therefore may be canceled.

A way to prove the conjecture is to choose a marking which corresponds to a monomial that does not cancel by other markings. First we prove the following general statement:

Lemma 2. ***(Partial induction step)*** *Assume that Conjecture 1 holds for all configurations of size μ. Consider a configuration Γ of size $\mu + 1$ satisfying no rectangle condition and there is at least one column in Γ with μ zero entries. Then, $\det(T_\Gamma) \neq 0$.*

Proof. By an appropriate permutation of rows and columns we can assume that the first column of Γ contains μ zeros in the first μ rows. Removing the first column and the last row from Γ, we obtain a configuration $\widetilde{\Gamma}$ of size μ. Since Γ satisfies no rectangle condition for $\mu + 1$, then $\widetilde{\Gamma}$ satisfies no rectangle condition for μ (otherwise, if $\widetilde{\Gamma}$ contains a $a \times b$ zero submatrix with $a + b = \mu + 1$, then we can attach the corresponding part of the first column of Γ to this submatrix to get $a \times (b + 1)$ zero submatrix of Γ, which contradict our no rectangle condition for Γ). If Conjecture 1 holds for $\widetilde{\Gamma}$, then the polynomial representing $\det(T_{\widetilde{\Gamma}})$ contains at least one nonzero monomial. Consider the marking of zero entries in $\widetilde{\Gamma}$, corresponding to this monomial. Then, mark also all μ zero entries in the first column. Then, the monomial corresponding to this new marking of zero entries in Γ in the polynomial representation of $\det(T_\Gamma)$ is not canceled. Indeed, assuming that it can be canceled, the canceling monomial must contain factor α_1^μ and must correspond to the marking with no zero entries marked in the last row of Γ, which implies that the monomial in $\widetilde{\Gamma}$ is also canceled. □

The previous lemma is far to cover all possible cases if we want to make an induction in μ, but it might be useful to at least cases of small μ. We say a configuration Γ is *totally sparse* if all of its columns have at most one zero. Obviously, monomial corresponding to any marking of totally sparse configuration cannot be canceled, so $\det(T_\Gamma) \neq 0$.

The case $\mu = 1$ is void. In the case of $\mu = 2$ the only configuration satisfying no rectangle condition (up to column/row permutations) is totally sparse. Now, consider the case $\mu = 3$. In this case either there exists a column with 2 zero entries and we can use Lemma 2 for $\mu = 2$, or the configuration is totally sparse.

Lemma 2, together with some other techniques, can be used to prove the conjecture for larger μ. We postpone the proofs to future publications.

Suppose Conjecture 1 is true. We evaluate the probability of success of this algorithm. Suppose we choose $\alpha_1, \ldots, \alpha_n$ independently with identical distribution over field $\mathbb{F}_q$. With the same field size, probability of success of this scheme is smaller than Algorithm 1. The algorithm succeeds if it satisfies

$$\det(T) \prod_{i,j \in [n]: i \neq j} (\alpha_i - \alpha_j) \neq 0. \tag{5}$$

Since the elements of ith row of T are the coefficients of polynomial $\prod_{\ell \in N_i}(x - \alpha_\ell)$, it can be verified that in each element of matrix T, degree of any one of the variables $\alpha_1, \ldots, \alpha_n$ is at most 1. Thus, the degree of any variable in $\det(T)$ is at most μ. It then follows that each variable has at most degree of $n + \mu$ in the polynomial in (5). By [13], when selecting $\alpha_1, \ldots, \alpha_n$ independently and uniformly, the probability that the polynomial in (5) does not evaluate to zero is lower bounded by $\left(1 - \frac{n+\mu}{q}\right)^n$. Note if we select $q = n(n + \mu) < 2n^2$, we get the lower bound probability as $\left(1 - \frac{1}{n}\right)^n$. The probability evaluates to constant $1/e$ as $n \to \infty$. So when n is large, the field size increases quadratically with the input size n. This result shows a big advantage comparing with the previous work in [1], which requires the field size increasing exponentially with the input size n.

Another advantage of our algorithm is in the complexity of checking whether the random algorithm succeeded. In the algorithm in previous work [1], it requires checking whether all $\mu \times \mu$ submatrices of the encoding matrix are full rank, which takes time exponential to input size. However in Algorithm 1 we only need to check one matrix (T) to be full rank. This can be done in polynomial running time.

Simulation results. To evaluate the probability of success of our algorithm we performed a numerical study. Our simulation results show that for $\mu = 8$ and $n = 16$ the probability of success is 0.71 and 0.84 for $q = 32$ and $q = 64$, respectively. For $\mu = 9$ and $n = 18$, the probability of success is 0.45 and 0.83 for $q = 32$ and $q = 64$, respectively. These results support our conjecture and indicate that the probability of success of our algorithm is high and increases with field size.

V. Conclusion

In this work we propose an algorithm to problem WSDE as an improvement of previous work. The improved random algorithm requires field size that increases linearly with respect to the input size, which is a very significant improvement comparing to exponential field size requirement in the previous work. In addition, the verification of success of the algorithm we propose only has polynomial complexity, outperforming the algorithm in previous work which requires exponential time complexity. We present a conjecture which, if correct, implies correctness of our algorithm. We proved the correctness of the algorithm for the small cases and performed simulations whose results support our conjecture.

References

[1] M. Yan and A. Sprintson, “Algorithms for weakly secure data exchange,” in *Proceedings of IEEE International Symposium on Network Coding (NetCod 2013)*, Calgary, Alberta, Canada, June 2013.

[2] S. E. Rouayheb, A. Sprintson, and P. Sadeghi, “On coding for cooperative data exchange,” in *Proceedings of IEEE Information Theory Workshop (ITW)*, Cairo, Egypt, January 2009.

[3] A. Sprintson, P. Sadeghi, G. Booker, and S. E. Rouayheb, “A randomized algorithm and performance bounds for coded cooperative data exchange,” in *Proceedings of 2010 IEEE International Symposium on Information Theory (ISIT)*, Austin, Texas, U.S.A., June 2010.

[4] ——, “Deterministic algorithm for coded cooperative data exchange,” in *Proceedings of 7th International ICST Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine)*, Houston, Texas, U.S.A., November 2010.

[5] N. Milosavljevic, S. Pawar, S. E. Rouayheb, M. Gastpar, and K. Ramchandran, “An optimal divide-and-conquer solution to the linear data exchange problem,” in *Proceedings of 2011 IEEE International Symposium on Information Theory (ISIT)*, 2011.

[6] T. A. Courtade and R. D. Wesel, “Coded cooperative data exchange in multihop networks,” arXiv:1203.3445.

[7] S. E. Tajbakhsh, P. Sadeghi, and R. Shams, “A generalized model for cost and fairness analysis in coded cooperative data exchange,” in *Proceedings of IEEE International Symposium on Network Coding (NetCod 2011)*, 2011.

[8] D. Ozgul and A. Sprintson, “An algorithms for cooperative data exchange with cost criterion,” in *Proceedings of 2011 Information Theory and Application Workshop (ITA)*, 2011.

[9] K. Bhattad and K. R. Narayanan, “Weakly secure network coding,” in *Proceedings of Workshop on Network Coding, Theory and Applications (NetCod ’05)*, Riva del Garda, Italy, April 2005.

[10] D. S. Hoang, “On index coding with side information,” Ph.D. dissertation, Nanyang Technological University, 2012.

[11] W. Halbawi, T. Ho, H. Yao, and I. Duursma, “Distributed Reed-Solomon Codes for Simple Multiple Access Networks,” arXiv:1310.5187.

[12] S. H. Dau, W. Song, and C. Yuen, “On the existence of mds codes over small fields with constrained generator matrices,” arXiv:1401.3807.

[13] T. Ho, “Networking from a network coding perspective,” Ph.D. dissertation, Massachusetts Institute of Technology, 2004.