ISIT'2013 1569720123

Codes for Limited View Adversarial Channels

Reihaneh Safavi-Naini, Pengwei Wang
Department of Computer Science
University of Calgary, Canada
Email: [rei, pengwei]@ucalgary.ca

***Abstract*—Channels with adversarial errors have been widely considered in recent years. In this paper we propose a new type of adversarial channel that is defined by two parameters ρ_r and ρ_w, specifying the read and write power of the adversary: for a codeword of length n, adversary can read $\rho_r n$ components and add an error vector of weight up to $\rho_w n$ to the codeword. We give our motivations, define performance criteria for codes that provide reliable communication over these channels, and describe two constructions, one deterministic and one probabilistic, for these codes. We disscus our results and outline our direction for future research.**

I. Introduction

In adversarial channels, the adversary uses a family of possible strategies, and the channel law is not known to the encoder and decoder. Using the terminology of [5] we refer to these channels as *unknown channels*. For example in Hamming model of adversarial channel, the adversary can corrupt a fraction ρ of a codeword symbols in an arbitrary way. Much less is known about adversarial channels. For example the highest rate of information transmission is Hamming model is unknown although it is known that it is much less than Shannon's capacity of binary symmetric channels [7]. In adversarial channels such as [9], [3], [4], the channel has full access to the sent codeword and can use knowledge to choose an error vector that maximizes the probability of decoding failure, where decoder error probability can be taken in average or worst case forms.

In this paper we introduce a new type of adversarial channels that we call *Limited View Adversarial Channel (LVAC),* and refer to codes that provide reliable communication over these channels as *LV-codes.* An LVAC adversary sees (reads) part of a sent codeword and uses this limited view of the codeword to choose an error vector that will be *added* to the sent codeword. In the following we describe our motivations and outline our results. The model raises many interesting new questions for future research.

Motivations. Our main motivation for studying LV-codes is a cryptographic primitive known as Reliable Message Transmission (RMT) [2]. In an RMT system a sender is connected to a receiver through a set of n node disjoint paths in a network, t of which are controlled by an adversary with unlimited computational power. The goal of RMT protocols is to provide reliability for transmission. Secure Message Transmission (SMT) protocols use the same adversary model and require reliability and security both. RMT and SMT provide reliability and security in information theoretic setting and have been used as a building block of other primitives such as multiparty computation. Transcript of a 1-round RMT protocol is an n-vector of values that are sent over paths, and can be seen as a codeword, and so it is natural to study 1-round RMT protocols as codes over adversarial channels, and use the wealth of coding theory research to find bounds and constructions for these protocols. This intuition is also supported by a general construction of a 1-round RMT protocol [8] using a message authentication code and Folded Reed Solomon code.

Our results.
We initiate the study of limited view adversarial channels and codes. A (ρ_r, ρ_w) limited view adversary can see a fraction ρ_r of a sent codeword of length n, and add an error vector of weigh at most $\rho_w n$, where addition is component wise and over a an algebraic group underlying the code. LV codes must provide reliable communication against this adversary, and good code must have high rate of information transmission. We model and define the average and maximum probability of the adversary's success, and give relationships among parameters ρ_r and ρ_w, n and minimum distance of the code. We propose two constructions. The first construction is a deterministic non-linear code that has the highest possible ρ_r and ρ_w that is allowed by the derived bounds. The second construction is a randomized construction that is based on a 1-round RMT construction in [8]. In this latter construction ρ_r and ρ_w exceed the derived bounds, (bounds are for deterministic constructions). An interesting result of this work is *the first, and the only,* construction of a deterministic 1-round RMT.

Related works.
A survey of unknown channels is given in [6]. The closest adversarial model to LVDC is γ-oblivious ρ-channels in [5]. γ-oblivious ρ-channels are mainly studied for binary case and are defined as follows. Let $\rho \in (0, 1/2)$ be a constant. A ρ-channel W is an adversarial channel that for a binary input x of length n, outputs a binary vector y of length n such that $d_H(x, y) \leq \rho n$. The channel corruption can be specified by an error vector e which in general depends on the input x and for each input is specified by a probability distribution on the set of all binary n-tuples. In γ-oblivious channel a subset $2^{(1-\gamma)n}$ distinct distributions are used for all 2^{Rn} inputs and so the same error distribution may be used for more than one input. In an LVAC, the adversary sees a fraction ρ_r of coordinates and so the adversary cannot distinguish between codewords, and hence distribution of errors, that coincide in those positions. On the other hand each codeword c belongs

to $\binom{n}{\rho_r n}$ coordinate subsets, and there are more than one error distribution for one codeword. Guruswami and Smith [3] considered additive adversarial channels where an adversarial error of weight up to ρn is added to the sent codeword. In their model the adversary can choose the distribution on the messages and has full knowledge of the codeword.

II. Background and Model

A codebook $C[n, N, d]$ is a set of vectors $C = \{c_1, \cdots, c_N\}$ in $\mathbb{F}_q^n$ with minimum distance d. The rate of the code is $R = \log_q N/n$. The codebook is used to encode a set $\mathcal{M}$ of N messages having a probability distribution $\Pr(m)$. We consider codes with *deterministic encoding and decoding functions:*[1]

$$Enc : \mathcal{M} \to C, \quad Dec : \mathbb{F}_q^n \to \{\mathcal{M}, \perp\}$$

Enc takes a message $m \in \mathcal{M}$ according to the distribution $\Pr(m)$, and and outputs a codeword c. *Dec* takes a corrupted word in $\mathbb{F}_q^n$ and outputs $\hat{m} \in \mathcal{M}$, or fails and outputs $\perp$.

We use Reed Solomon (RS) codes defined over $\mathbb{F}_q$. The encoding function $RS(f, n) = (f(\alpha_1), \cdots f(\alpha_n))$ encodes a message of length $k < n$ to a codeword of length n. The message is used as coefficients of a polynomial $f(x)$. The codevector is evaluation of $f(x)$ on n known values $\alpha_1, \cdots \alpha_n$.

The first construction uses an *Algebraic Manipulation Detection Code*[1].

Definition 1: An (N, G, δ)-Algebraic Manipulation Detection code (AMD code) maps a message set $\mathcal{M}$ of size N into an additive group $\mathcal{G}$ of order G using an encoding function, $E : \mathcal{M} \to \mathcal{G}$. The code has a decoding function $D : \mathcal{G} \to \mathcal{M} \cup \{\perp\}$ such that $D(E(m)) = m$ for any $m \in \mathcal{M}$. The security of an AMD code requires that for any $m \in \mathcal{M}$, $\Delta \in \mathcal{G}$,

$$\Pr[D(E(m) + \Delta) \in \{m, \perp\}] \leq \delta \tag{1}$$

We will use the AMD code $E : \mathbb{F}_q^\ell \to \mathbb{F}_q^\ell \times \mathbb{F}_q \times \mathbb{F}_q$,

$$E(m) = (m, r, f(m, r)); \quad f(m, r) = r^{\ell+2} + \sum_{i=1}^{\ell} m_i r^i \tag{2}$$

In the above equation, r is randomly chosen from $\mathbb{F}_q$ and serves as the randomness of encoding.

A. Limited view adversary

We assume the distribution on $\mathcal{M}$ is uniform. For a set of positions $S_r = \{i_1, \cdots, i_{\rho_r n}\} \subset [n]$, let $\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]$ denote the set of codewords that have $(c_{i_1}, \cdots, c_{i_{\rho_r n}})$ in positions $\{i_1, \cdots, i_{\rho_r n}\}$, respectively. Let $\mathcal{E}$ denote the set of vectors e with $wt(e) = \rho_w n$.

Definition 2: A (ρ_r, ρ_w) limited view adversary ((ρ_r, ρ_w) LV adversary) has two capabilities: reading and writing.

- Reading: Adversary can read a ρ_r fraction of a codeword.
- Writing: Adversary can add to the codeword, an error with $wt(e) = \rho_w n$.

[1]Probabilistic (Stochastic) codes will have probabilistic encoding and deterministic decoding.

The adversary's error is defined by two position sets, $S_r = \{i_1, \cdots, i_{\rho_r n}\} \subset [n]$, and $S_w = \{j_1, \cdots, j_{\rho_w n}\} \subset [n]$ and error values $c_{j_1}, \cdots, c_{j_{\rho_w n}}$. The adversary knows the codeword is an element of a known codebook C with known encoding and decoding functions. His goal is to to make the decoder fail. For a chosen subset of position $\{i_1, \cdots, i_{\rho_r n}\}$ the adversary observes $\{c_{i_1}, \cdots, c_{i_{\rho_r n}}\}$, chooses $e \in \mathbb{F}_q^n, wt(e) \leq \rho_w n$ and adds e to the channel. For worst case adversarial error, we consider the error that results in the highest failure probability (there can be more than one error-we consider one such errors).

Lemma 1: For a code of length n and minimum distance d, we have $\rho_r n \leq n - d$ and $\rho_w n \leq d - 1$. Moreover,

A1 $\rho_w = \rho_r$, then $\rho_r n = \rho_w n \leq \frac{n-1}{2}$.
A2 $\rho_w > \rho_r$, then $\rho_r n \leq \frac{n-1}{2}$ and $\rho_w n \leq d - 1$.
A3 $\rho_w < \rho_r$, then $\rho_w n \leq \frac{n-1}{2}$ and $\rho_r n \leq n - d$.

Proof: The adversary can read at most $\rho_r n \leq n - d$ components. This is because any $n - d + 1$ components can uniquely determine the codeword. The adversary can add errors to at most $\rho_r n \leq d - 1$ components because the minimum distance of codeword is d.

A1 is true because if the adversary can add at most $\rho_r n = \rho_w n = d - 1$ (or $\rho_w n = \rho_r n = n - d$) errors, the distance should meet $d \leq \frac{1}{2}(n + 1)$ $(d \geq \frac{1}{2}(n + 1))$. So we have, $\rho_r n = \rho_w n = d - 1 \leq \frac{n-1}{2}$ $(\rho_r n = \rho_w n = n - d \leq \frac{n-1}{2})$.

A2 is true because if $n - d \leq d - 1$, then $d \geq \frac{n+1}{2}$. So we have $\rho_r n \leq \frac{n-1}{2}$. If $n - d \geq d - 1$, then $d \leq \frac{n+1}{2}$ and we have $\rho_r n < \rho_w n \leq \frac{n-1}{2}$.

A3 is proved in a similar way. ■

B. Decoding error

Definition 3: For a codeword $c \in \mathbb{F}_q^n$ and an integer r, let $B(r, c)$ be the Hamming ball of radius r centred at c.

For $e \in \mathbb{F}_q^n$, let $wt(e)$ be the number of non-zero components of e.

Definition 4: A *Bounded Distance Decoding (BDD)* algorithm Dec takes a received word $y = (y_1, \cdots, y_n)$ and outputs $c \in \mathcal{C}$, if c is the unique codeword at distance at most $wt(e)$ from y, otherwise output $\perp$.

$$Dec(y) = \begin{cases} m & if\ |\mathbf{B}(wt(e), c + e)| = 1 \\ \perp & if\ |\mathbf{B}(wt(e), c + e)| > 1 \end{cases}$$

An adversary who sees the sent codeword c, that is $\rho_r = 1$, will always succeed in making the decoder to fail: the adversary constructs $e = c - c'$ that has at most $\rho_w n$ non-zero components, and adds that to c.

An LV adversary however does not see the whole codeword. By observing $\{c_{i_1}, \cdots, c_{i_{\rho r n}}\}$, the adversary can only determine a subset of possible sent codewords. So the choice of the error vector will depend on the partial information about the sent codeword.

Definition 5: Consider a subset $\{i_1, \cdots, i_{\rho_r n}\} \subset [n]$ and assume the values seen by the adversary are $\{c_{i_1}, \cdots, c_{i_{\rho_r n}}\}$.

The adversary knows one of the codewords in $\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]$ has been sent but is uncertain which one. *Assuming uniform message distribution*, all codewords

have the same probability and because of deterministic encoding, codewords in $\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]$ all have the same probability of having been sent. For a codeword c in $\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]$, an error vector e, will cause decoder failure, if $\mathbf{B}(wt(e), c+e)$ contains a codevector other than c. Since the adversary is not certain about the sent codeword, he has to choose e such that it causes decoder failure for the most number of codewords in $\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]$ which e causes decoder failure.

Definition 6: Let,

$$A_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]) = \{c \in \mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}] \mid c' \in C, \\ c' \neq c, c' \in \mathbf{B}(wt(e), c+e)\}$$

Here $A_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}])$ is the subset of codewords in $\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]$ that will cause decoder to fail if adversary uses e.

The decode error probability for e and (possible sent codewords) $\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]$ is,

$$\delta_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]) = \frac{|A_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}])|}{|\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]|}$$

For a given $(c_{i_1}, \cdots, c_{i_{\rho_r n}})$, to choose an e that results in the highest decoder failure probability, the adversary finds,

$$\delta(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]) = \max_e \frac{|A_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}])|}{|\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]|} \\ = \frac{1}{|\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]|} \max_e |A_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}])|$$

Here $\delta(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}])$ is the highest probability of decoder failure when the set of possible sent codewords is $\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]$, and is obtained for worst case error vectors $e = \mathcal{A}(c_{i_1}, \cdots, c_{i_{\rho_r n}})$.

For a set of position $\{i_1, \cdots, i_{\rho_r n}\}$, we can define two types of failure probability: average and maximum.

Definition 7: Average failure probability for the subset $\{i_1, \cdots, i_{\rho_r n}\}$ is,

$$\hat{\delta}(i_1, \cdots, i_{\rho_r n}) = \sum_{c_{i_1}, \cdots, c_{i_{\rho_r n}}} \Pr(c_{i_1}, \cdots, c_{i_{\rho_r n}}) \cdot \\ \delta(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}])$$

Maximum error probability for the subset $\{i_1, \cdots, i_{\rho_r n}\}$ is,

$$\delta_m(i_1, \cdots, i_{\rho_r n}) = \max_{c_{i_1}, \cdots, c_{i_{\rho_r n}}} \delta(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}])$$

Finally the adversary will find the best choice of $\{i_1, \cdots, i_{\rho_r n}\}$ to maximize the decoding error probability $\hat{\delta}(i_1, \cdots, i_{\rho_r n})$, (or $\delta(i_1, \cdots, i_{\rho_r n})$).

Definition 8: The average and maximum error probability for the decoder of an LV code $\mathcal{C}$ is given by,

$$\hat{\delta} = \max_{i_1, \cdots, i_{\rho_r n} \subset [n]} \hat{\delta}(i_1, \cdots, i_{\rho_r n})$$

and,

$$\delta_m = \max_{i_1, \cdots, i_{\rho_r n} \subset [n]} \delta_m(i_1, \cdots, i_{\rho_r n})$$

In an $[n, N, \delta]$ limited view adversary code C, δ is the probability that a message sent by the sender is correctly decoded by the receiver, when the adversary uses their best strategy to defeat the decoder, and δ is defined as above.

Definition 9: For a (ρ_r, ρ_w) *limited view* adversarial channel, a rate R is *achievable*, if there is a family of codes $[n, N, \delta]$ indexed by $n \in \mathbb{N}$, such that $N > 2^{Rn}$ and for sufficiently large n, we have $\hat{\delta}(\delta_m) \leq \delta$. Capacity of a (ρ_r, ρ_w) *limited view* adversarial channel, is the highest achievable rate.

III. Constructions

We give two constructions of LV- codes.

The first construction is *deterministic* and the second one is randomized. The second construction can have higher ρ_r and ρ_w. *In the first construction, we assume $\rho_r = \rho_w$. In the second construction, we assume that the adversary read and write sets are the same.* We denote the set $[n] = \{1, \cdots, n\}$.

A. Construction I

The first construction is a non-linear code that uses Reed-Solomon codes and AMD codes. The LV code is defined over $\mathbb{F}_q^2$ and has $(q^2)^{k'}$ codewords. Assume the adversary reads and writes to, at most $t = min(k' - 1/2, n - k' - 1/2)$ components. That is, $\rho_r = \rho_w = min(R - \frac{1}{2n}, 1 - R - \frac{1}{2n})$.

Sender:

1) A message is a vector $(m_1, \cdots, m_{2k-2}, m_{2k-1})$, $m_i \in \mathbb{F}_q$. We assume the message is uniformly distributed over $\mathbb{F}_q^{2k-1}$. Sender chooses $m_{2k-1} \neq \{\alpha_{i,j}\}_{i=1\cdots n, j=1,2}$[2]
   $\mathcal{S}$ uses the AMD code in Section II to encode the message, as AMD code $(m_1, \cdots, m_{2k-2}, m_{2k-1}, \gamma)$ where, $\gamma = m_{2k-1}^{2k} - \sum_{i=1}^{2k-2} m_i m_{2k-1}^i$.
   In other words, m_{2k-1} serves as randomness in encoding.
2) The AMD codes is encoded using an RS code of length $2n$. The AMD code $(m_1, \cdots, m_{2k-2}, m_{2k-1}, \gamma)$ defines a polynomial, $f(X) = \gamma + \sum_{i=1}^{2k-1} m_i X^i$. We use $\mathsf{RS}(f, 2n)$ to generate a codevector of length $2n$. The codeword of the LV code is by interpolating pairs of consecutive elements of this code as elements of $\mathbb{F}_q^2$. That is, considering a pair $c_j = (c_{j,1}, c_{j_2}) = (f(\alpha_{2j-1}), f(\alpha_{2j})), j = 1, \cdots n$ of evaluations, as an element of $\mathbb{F}_q^2$. The LV code has q^{2k-1} codewords (γ is AMD tag) and is defined over $\mathbb{F}_q^2$ and so the equivalent dimension of the code is $(q^2)^{k'} = q^{2k-1}$, that is $k' = k - 1/2$.

Receiver:

1) For every k positions $\{l_1, \cdots, l_k\} \in [n]$, $\mathcal{R}$ finds $\{y_{l_1} \cdots, y_{l_k}\}$ with $y_{l_j} = (y_{l_j,1}, y_{l_j,2})$ (in total $2k$ evaluations of $f(x)$). By solving a system of linear equation $\{f'(\alpha_{2l_j-1}) = y_{l_j,1}, f'(\alpha_{2l_j}) = y_{l_j,2}\}_{j=1\cdots k}$, $\mathcal{R}$ obtains a candidate AMD code $(m_1', \cdots, m_{2(k-1)}', m_{2k-1}', \gamma')$.

[2]This effectively results in a slightly lower rate. We make this assumption to simplify the analysis.

2) Verify, $\gamma' = (m'_{2k-1})^{2k} - \sum_{i=1}^{2k-2} m'_i (m'_{2k-1})^i$. If there is a unique message vector that passes the verification, output the message; otherwise output $\perp$.

Theorem 1: For $\rho_r = \rho_w = min(R - \frac{1}{2n}, 1 - R - \frac{1}{2n})$ the code above has $q^{2k'}$ codewords, and $\delta \leq \mathcal{O}(\binom{n}{k}n)/q$.

Proof: Assume the adversary reads the set of positions $\mathsf{S}_r = \{i_1 \cdots i_t\}$, and adds error e with non-zero elements on $\mathsf{S}_w = \{j_1, \cdots, j_t\}$, and $y_j = c_j + e_j$ for $j = \{j_1 \cdots j_t\}$. First we calculate the size of set $|\mathcal{C}[c_{i_1}, \cdots, c_{i_t}]|$. The set of codewords with the same components $\{c_{i_1} \cdots c_{i_t}\}$ on S_r consists of $c = \mathsf{RS}(f, 2n)$ with coefficients of $f(X)$ satisfying:

$$\begin{bmatrix} c_{i_1,1} - m_{2k-1}\alpha_{i_1,1}^{2k-1} \\ c_{i_1,2} - m_{2k-1}\alpha_{i_1,2}^{2k-1} \\ \vdots \\ c_{i_t,1} - m_{2k-1}\alpha_{i_t,1}^{2k-1} \\ c_{i_t,2} - m_{2k-1}\alpha_{i_t,2}^{2k-1} \\ m_{2k-1}^{2k} \end{bmatrix} = \begin{bmatrix} 1 & \alpha_{i_1,1} & \cdots & \alpha_{i_1,1}^{2k-2} \\ 1 & \alpha_{i_1,2} & \cdots & \alpha_{i_1,2}^{2k-2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha_{i_t,1} & \cdots & \alpha_{i_t,1}^{2k-2} \\ 1 & \alpha_{i_t,2} & \cdots & \alpha_{i_t,2}^{2k-2} \\ 1 & m_{2k-1} & \cdots & m_{2k-1}^{2k-2} \end{bmatrix} \times [\gamma, m_1, \cdots, m_{2k-2}]^T \quad (3)$$

where $[.]^T$ is the transpose of the vector $[.]$. For each m_{2k-1}, there are $q^{2k-2t-2}$ solutions for $(\gamma, m_1, \cdots, m_{2k-2})$. So there are at least $(q - 2n)q^{2k-2t-2}$ solutions for $(m_1, \cdots, m_{2k-1}, \gamma)$. Each message corresponds to a codeword with components $\{c_{i_1}, \cdots, c_{i_t}\}$ on S_r and so $|\mathcal{C}[c_{i_1}, \cdots, c_{i_t}]| = (q - 2n)q^{2k-2t-2}$.

Next we upper bound the size of set $|A_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}])|$. For a vector $x = (x_1 \cdots x_n)$, let $\mathsf{Supp}(x)$ denote the subset of positions where $x_i \neq 0$. Let,

$$A_e^{(l_1,\cdots,l_k)}(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}]) = \{c \in \mathcal{C}[c_{i_1}, \cdots, c_{i_t}] \mid c' \in C, c' \neq c, \mathsf{Supp}(c + e - c') \subseteq [n] \backslash \{l_1, \cdots, l_k\}\}$$

Because $t = \min(k - 1, n - k)$, we have $t \leq n - k$. So there is

$$A_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}]) \subseteq \{c \in \mathcal{C}[c_{i_1}, \cdots, c_{i_t}] \mid c' \in C, c' \neq c, c' \in \mathbf{B}(n - k, c + e)\} = \bigcup_{l_1,\cdots,l_k \in [n]} A_e^{(l_1,\cdots,l_k)}(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}])$$

For each sets of k positions $\{l_1, \cdots, l_k\} \in [n]$ we calculate the number of codewords in $A_e^{(l_1,\cdots,l_k)}\mathcal{C}[c_{i_1}, \cdots, c_{i_t}]$.

The received word y on a set of k positions $\{l_1, \cdots, l_k\} \in [n]$ is

$$\{y_{l_1,1}, y_{l_1,2}, \cdots, y_{l_k,1}, y_{l_k,2}\} = \{c_{l_1,1}, c_{l_1,2}, \cdots, c_{l_k,1}, c_{l_k,2}\} + \{e_{l_1,1}, e_{l_1,2}, \cdots, e_{l_k,1}, e_{l_k,2}\}$$

The set of $2k$ linear equations,

$$\begin{bmatrix} y_{l_1,1} \\ y_{l_1,2} \\ \vdots \\ y_{l_k,1} \\ y_{l_k,2} \end{bmatrix} = \begin{bmatrix} 1 & \alpha_{l_1,1} & \cdots & \alpha_{l_1,1}^{2k-1} \\ 1 & \alpha_{l_1,2} & \cdots & \alpha_{l_1,2}^{2k-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha_{l_k,1} & \cdots & \alpha_{l_k,1}^{2k-1} \\ 1 & \alpha_{l_k,2} & \cdots & \alpha_{l_k,2}^{2k-1} \end{bmatrix} \times \begin{bmatrix} \gamma' \\ m'_1 \\ \vdots \\ m'_{2k-1} \end{bmatrix} \quad (4)$$

has full rank and gives a unique solution,

$$\{m'_1, \cdots, m'_{2k-1}, \gamma'\} = \{m_1, \cdots, m_{2k-1}, \gamma\} + \{\Delta m_1, \cdots, \Delta m_{2k-1}, \Delta\gamma\}$$

If the vector $\{m'_1, \cdots, m'_{2k-1}, \gamma'\}$ passes AMD verification,

$$\gamma + \Delta\gamma = (m_{2k-1} + \Delta m_{2k-1})^{2k} - \sum_{i=1}^{2k-2} (m_i + \Delta m_i)(m_{2k-1} + \Delta m_{2k-1})^i \quad (5)$$

then there exists a codewode c which is encoded from $\{m_1, \cdots, m_{2k-1}, \gamma\}$ by $\mathsf{RS}(\cdot)$ such that the codeword is in $\mathcal{C}[c_{i_1}, \cdots, c_{i_t}]$ and there exists $c' = \mathsf{RS}(f', 2n)$ with $f'(X) = \gamma' + \sum_{i=1}^{2k-1} m'_i X^i$ and $c' \in \mathbf{B}(n - k, c + e)$. So if the vector $\{m_1, \cdots, m_{2k-1}, \gamma\}$ satisfies Eq. 3 and Eq. 5, the encoded codeword c is in $A_e^{(l_1,\cdots,l_k)}(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}])$.

Next we calculate the size of solution $\{m_1, \cdots, m_{2k-1}, \gamma\}$ in Eq. 3 and Eq. 5. The Eq. 5 is equivalent to,

$$\gamma + \sum_{i=1}^{2k-2} m_i(m_{2k-1} + \Delta m_{2k-1})^i = (m_{2k-1} + \Delta m_{2k-1})^{2k} - \sum_{i=1}^{2k-2} \Delta m_i (m_{2k-1} + \Delta m_{2k-1})^i - \Delta\gamma \quad (6)$$

We add Eq. 6 to Eq. 3 and form a new equations system with $2t + 2$ equations. We bound the size $|A_e^{(l_1,\cdots,l_k)}(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}])|$ for two cases, (i) $\Delta m_{2k-1} = 0$, and (ii) $\Delta m_{2k-1} \neq 0$.

Let $\Delta m_{2k-1} = 0$. We consider two cases: if m_{2k-1} satisfies $\sum_{i=1}^{2k-2} \Delta m_i m_{2k-1}^i = \Delta\gamma$, the equation system for $\{\gamma, m_1, \cdots, m_{2k-2}\}$ has $q^{2k-2t-2}$ solutions. On other hand, if $\sum_{i=1}^{2k-2} \Delta m_i m_{2k-1}^i \neq \Delta\gamma$, the equation system does not have any solution for $\{\gamma, m_1, \cdots, m_{2k-2}\}$. So there is

$$\begin{aligned} &|A_e^{(l_1,\cdots,l_k)}(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}])| \\ \leq &|\{m_{2k-1} \mid \sum_{i=1}^{2k-2} \Delta m_i m_{2k-1}^i = \Delta\gamma]\}| \times q^{2k-2t-2} + \\ &|\{m_{2k-1} \mid \sum_{i=1}^{2k-2} \Delta m_i m_{2k-1}^i \neq \Delta\gamma]\}| \times 0 \\ = &(2k-2)q^{2k-2t-2} \end{aligned}$$

Let $\Delta m_{2k-1} \neq 0$. If m_{2k-1} satisfies $m_{2k-1} + \Delta m_{2k-1} = \{\alpha_{i,j}\}_{i=1\cdots n, j=1,2}$, the equation system for $(\gamma, m_1, \cdots, m_{2k-2})$ has $q^{2k-2t-2}$ solutions, and if $m_{2k-1} + \Delta m_{2k-1} \neq \{\alpha_{i,j}\}_{i=1\cdots n, j=1,2}$, the equation system has $q^{2k-2t-3}$ solutions. So there is

$$\begin{aligned} &|A_e^{(l_1,\cdots,l_k)}(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}])| \\ \leq &|\{m_{2k-1} \mid m_{2k-1} + \Delta m_{2k-1} = \{\alpha_{i,j}\}_{\substack{i=1\cdots n \\ j=1,2}}\}| \times q^{2k-2t-2} \\ &+ |\{m_{2k-1} \mid m_{2k-1} + \Delta m_{2k-1} \neq \{\alpha_{i,j}\}_{\substack{i=1\cdots n \\ j=1,2}}\}| \\ &\times q^{2k-2t-3} = 4n \times q^{2k-2t-2} + (q - 4n)q^{2k-2t-3} \\ \leq &(4n+1)q^{2k-2t-2} \end{aligned}$$

This means that for any set of k positions $\{l_1, \cdots, l_k\}$, we have $|A_e^{(l_1,\cdots,l_k)}(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}])| \leq (4n+1)q^{2k-2t-2}$.

There are total $\binom{n}{k}$ of the set of k positions $\{l_1, \cdots, l_k\} \subset [n]$ and so $|A_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_t}])| \leq \binom{n}{k}(4n+1)q^{2k-2t-2}$.

We have $q > 4n+1$ if we choose q to be large. So the probability of decoding error is

$$\delta = \frac{|A_e(\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}])|}{|\mathcal{C}[c_{i_1}, \cdots, c_{i_{\rho_r n}}]|} = \frac{\binom{n}{k}(4n+1)}{q-2n} \leq \frac{8\binom{n}{k}n}{q}$$

■

The error probability for this code is a function of n and so to have low δ the field size q must be chosen appropriately.

B. Construction II

We give a randomized construction that is based on the RMT scheme[8]. For this construction we assume the set of read and write components are the same. For randomized codes, ρ_r, ρ_w may be higher than the bounds in Lemma 1.

Assume the adversary controls at most $t < \frac{n}{2}$components. The code uses a Message Authentication Code (MAC) and a Folded Reed Solomon codes (FRS codes). More details on the construction can be found in [8].

Sender:

1) The sender generates n keys r_j, $j \in [n]$, for a MAC and constructs a tag for the message. To construct the keys, $\mathcal{S}$ generates $t+1$ random polynomials $r = \{P_1(X), \cdots, P_{t+1}(X)\}$ over $\mathbb{F}_q^s$, each of degree t. $\mathcal{S}$ also chooses $\{\alpha_1, \alpha_2, \cdots, \alpha_n\} \in \mathbb{F}_q^s$ and makes them public. The keys are,

   $$r_j = \{P_1(\alpha_j), P_2(\alpha_j), \cdots, P_{t+1}(\alpha_j)\}, j = 1, \cdots, n$$

   The tag for a message $\mathbf{m} = \{m_1 \cdots m_{\binom{t+2}{2}-1}\}$ consists of coefficients of $A(z)$ over $\mathbb{F}_q^s$,

   $$\begin{aligned} A(X) =& \mathsf{MAC}(\mathbf{m}, r) = m_1 P_1(X) + \cdots + m_t P_t(X) \\ &+ m_{t+1} P_1(X)^2 + \cdots + m_{2t} P_t(X)^2 + \cdots + \\ &m_{\binom{t+2}{2}-1} P_{t-1}(X) P_t(X) + P_{t+1}(X) \end{aligned}$$

2) The FRS code is used to encode that tagged message constructed above. The message length (code dimension), $k = \binom{t+2}{2} - 1)s + (t+1)s$. The encoded message is,

   $$\mathbf{f} = \{m_1 \cdots m_{\binom{t+2}{2}-1}, A(z)\}$$

3) Each component of the above codeword will be appended with the associated randomness to form the corresponding component of the final code,

   $$c_j = \{f(\gamma^{ju}), f(\gamma^{ju}), \cdots, f(\gamma^{ju+u-1}), r_j\},\ 1 \leq j \leq n.$$

Receiver:

1) Receives a word y with components $y_j = \{y_{j,1}, y_{j,2}, \cdots, y_{j,u}, r'_j\}$. Uses FRS list decoding on Y and obtains a list of messages that will include the correct one.
2) Constructs the authentication vector $\mathsf{MAC}(\mathbf{m}, r_1) = A(\alpha_1), \mathsf{MAC}(x, r_2) = A(\alpha_2), \cdots, \mathsf{MAC}(x, r_n) = A(\alpha_n)$. $\mathcal{R}$ outputs $(m_0, m_1, \cdots, m_{(\binom{t+2}{2}-1)s-1})$ if this is the unique message such that at least $t+1$ verification equations pass. Otherwise output $\perp$.

Theorem 2: The code above provides protection against a (ρ_r, ρ_w) limited view adversary with $\rho_r = \rho_w = t/n < \frac{1}{2}$. The code is an $[n, q^{(\binom{t+2}{2}-1)s}, \delta]$ LV code over $\mathbb{F}_q^{u+(t+1)s}$, with $\delta \leq \frac{n-t}{q}$ where, $s = 2n, u = s^2, q > nu$.

Proof: Follows from construction in [8]. ■

The adversary in construction II can add more errors because we do not need the condition that $t \leq \min(R - \frac{1}{2n}, 1 - R - \frac{1}{2n})$. Decoding algorithms for both constructions are exponential.

IV. Concluding Remarks

We proposed a new type of adversarial channels that is motivated by RMT and SMT and gave constructions for a deterministic and a randomized code. We note that in RMT and SMT, the adversarial error is replacement error: the adversary san replace a component that they write to with any arbitrary field element.

If $S_w \subset S_r$, the replacement error is equivalent to the additive error and an LV code can always give an RMT or SMT. The converse however is only true if extra limitations (write and read set are the same) will be put on LV adversary. Numerous open problems, including construction of high rate and low error probability codes with efficient decoding, as well as tight bounds on capacity of LV channels will be interesting open question for future research.

Acknowledgment

This research is in part supported by Alberta Innovates Technology Future, in the province of Alberta, Canada.

References

[1] R. Cramer, Y. Dodis, S. Fehr, C. Padró, D. Wichs, "Detection of Algebraic Manipulation with Applications to Robust Secret Sharing and Fuzzy Extractors", *EUROCRYPT*, pp. 471–488, 2008.

[2] D. Dolev, C. Dwork, O. Waarts, and M. Yung, "Perfectly Secure Message Transmission", *Journal of the ACM*, vol. 40(1), pp. 17-47, 1993.

[3] V. Guruswami, A. Smith, "Codes for Computationally Simple Channels: Explicit Constructions with Optimal Rate", *FOCS*, pp. 723–732, 2010.

[4] M. Langberg, "Private Codes or Succinct Random Codes That Are (Almost) Perfect", *FOCS*, pp. 325–334, 2004.

[5] M. Langberg, "Oblivious Communication Channels and Their Capacity", *IEEE Transaction Information Theory*, Vol. 54(1), pp. 424–429, 2008.

[6] A. Lapidoth and P. Narayan, "Reliable Communication Under Channel Uncertainty", *IEEE Transaction Information Theory*, vol. 44(6), pp. 2148–2177, 1998.

[7] C. E. Shannon, "A Mathematical Theory of Communication", *Bell System Tech Journal*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.

[8] R. Safavi-Naini, M. Tuhin, P. Wang. "A General Construction for 1-round 1-RMT and (0, δ)-SMT", *ACNS*, pp. 344–362, 2012.

[9] R. Hamming, "Error Detecting and Error Correcting Codes", *Bell System Technical Journal*, vol. 29 (2), pp. 147-160, 1950.