Relative generalized matrix weights of matrix codes for universal security on wire-tap networks

Ryutaroh MATSUMOTO[1] (speaker)
(jont work with Umberto Martínez-Peñas[2])

[1]Tokyo Institute of Technology, Japan [2]Aalborg University, Denmark

8 March 2017
Institute of Network Coding Seminar
Chinese University of Hong Kong
(Please ask your question at any time.)

Secure network coding

Assumptions:

- single source multicast, and
- an adversary (Eve) can eavesdrop her chosen μ links in the network.

Goal: The legitimate users want to hide transmitted data from Eve.
The above problem and its solution were proposed as "secure network coding" by Cai and Yeung (2002).

Relation to other areas:

- Secure network coding is the network coding counterpart of the wiretap channel coding initiated by Wyner (1975) and Csiszár-Körner (1978).
- Secure network coding is a generalization of (threshold-type linear) secret sharing proposed by Shamir and Blakley (1979).

Nested coset coding

$C_2 \subsetneqq C_1 \subseteq GF(q^m)^n$: $GF(q^m)$-linear codes
A message is a coset $\vec{a} + C_2 = \{\vec{a} + \vec{x} : \vec{x} \in C_2\} \in C_1/C_2$, for $\vec{a} \in C_1$.
$|\vec{a} + C_2| = |C_2|$ for any $\vec{a}$.

The number of messages is

$$= \frac{|C_1|}{|C_2|} = \frac{q^{m \dim C_1}}{q^{m \dim C_2}} = q^{m(\dim C_1 - \dim C_2)}.$$

Generation of a packet from a given message

single source multicast (acyclic, delay-free)
n: # outgoing links from the source
m: time slots in a packet, m must be $\geq n$ for existence of MRD codes.
One $GF(q)$ symbol is carried on a link per time slot
$GF(q)$-linear coding at all intermediate nodes
$m \times n$ $GF(q)$ symbols in a packet.

$C_2 \subsetneqq C_1 \subseteq GF(q^m)^n$: $GF(q^m)$-linear (MRD) codes
$S \in C_1/C_2$: Given message

1. Randomly choose a vector $\vec{x} = (x_1, \ldots, x_n) \in S \subsetneqq GF(q^m)^n$.
2. Expand $x_i \in GF(q^m)$ into $(x_i^{(1)}, \ldots, x_i^{(m)}) \in GF(q)^m$ by some fixed $GF(q)$-linear basis of $GF(q^m)$,
3. Send $x_i^{(j)}$ on link i at time j.

Generation of $\vec{x}$ from S is called the nested coset coding.

q-th power of subspaces (Stichtenoth (1990))

$\vec{x} = (x_1, \ldots, x_n) \in GF(q^m)^n$,
$\vec{x}^q = (x_1^q, \ldots, x_n^q)$.
$W^q = \{\vec{x}^q : \vec{x} \in W\}$ for an $GF(q^m)$-linear subspace W of $GF(q^m)^n$.
W^q is again an $GF(q^m)$-linear subspace despite $\vec{x} \mapsto \vec{x}^q$ is $GF(q^m)$-nonlinear.
$W^* = W + W^q + W^{q^2} + W^{q^3} + \cdots + W^{q^{m-1}}$.

- For an $GF(q^m)$-subspace $W \subseteq GF(q^m)^n$, $W^q = W$ iff W has an $GF(q^m)$-basis written in $GF(q)^n$,

The above were given by Stichtenoth (1990) for studying subfield subcodes.

j-th Relative Generalized Rank Weight (RGRW) (Kurihara et al. 2015)

For $D_2 \subsetneqq D_1 \subseteq GF(q^m)^n$,

$$R_j(D_1, D_2) \quad = \quad \min\{\dim W^* : W \subseteq D_1, \dim W = j, W \cap D_2 = \{0\}\}$$

Eve creates a network of arbitrary shape and choose arbitrary μ links to observe.
Z: observed information, S: secret message (uniform distribution)

Relation between RGRW and eavesdropped information

$\max I(S;Z)$ in $\log_{q^m} \geq j \Leftrightarrow \mu \geq R_j(D_2^\perp, D_1^\perp)$

The maximum is taken over all shapes of network and all choices of μ links.

Corollary

If $\mu < R_1(D_2^\perp, D_1^\perp)$ then there is no information leakage.

Kurihara et al.'s RGRW is insufficient

$GF(q^m)^n \supseteq D_1 \supsetneqq D_2$ have to be $GF(q^m)$-linear in Kurihara et al. (2015).

Guruswami et al. (2016) proposed a code $C_1^{(\mathrm{GWX})} \subset GF(q^m)^n$ that is

- capable of list decoding more errors than known codes, and
- $GF(q)$-linear but **NOT** $GF(q^m)$-linear.

To find a subcode $C_2^{(\mathrm{GWX})} \subsetneqq C_1^{(\mathrm{GWX})}$ providing secrecy of transmitted messages, we need to extend Kurihara et al. (2015).

Codewords as matrices

$\vec{x} = (x_1, \ldots, x_n) \in GF(q^m)^n$
Expand $x_i \in GF(q^m)$ into $(x_i^{(1)}, \ldots, x_i^{(m)}) \in GF(q)^m$ by some fixed $GF(q)$-linear basis of $GF(q^m)$

There is a one-to-one correspondence between

$$\vec{x} = (x_1, \ldots, x_n) \text{ and } \begin{pmatrix} x_1^{(1)} & \cdots & x_n^{(1)} \\ \vdots & & \vdots \\ x_1^{(m)} & \cdots & x_n^{(m)} \end{pmatrix}.$$

From now, we will consider nested coset coding by $C_2 \subsetneqq C_1 \subseteq \underline{GF(q)^{m\times n}}$.

$C_2 \subsetneqq C_1$ are assumed to be $GF(q)$-linear, and will be called "matrix spaces".

Rank support of a matrix space

$GF(q)^{m\times n} \supseteq C$: matrix space
$C \ni M$: an $m \times n$ matrix
$\mathrm{Row}(M)$: row space of M

$\mathrm{Rsupp}(C) = \sum_{M\in C} \mathrm{Row}(M)$ = vector space spanned by row vectors of all $M \in C$.

The rank weight of the matrix space C is $\dim \mathrm{Rsupp}(C)$. Note that

- $0 \le \dim C \le mn$, and
- $0 \le \dim \mathrm{Rsupp}(C) \le n$.

j-th relative generalized matrix weight (RGMW)

For $C_2 \subsetneqq C_1 \subseteq GF(q)^{m\times n}$,

$$M_j(C_1, C_2) \quad = \quad \min\{\dim \mathrm{Rsupp}(V) : V \subseteq C_1, \dim V = j, V \cap C_2 = \{0\}\}$$

Eve creates a network of arbitrary shape and choose arbitrary μ links to observe.
Z: observed information, S: secret message (uniform distribution)

Relation between RGMW and eavesdropped information

$\max I(S;Z)$ in $\underline{\log_q} \geq j \Leftrightarrow \mu \geq M_j(C_2^\perp, C_1^\perp)$

The maximum is taken over all shapes of network and all choices of μ links.

Corollary

If $\mu < M_1(C_2^\perp, C_1^\perp)$ then there is no information leakage.

Relation between RGRW and RGMW

$$M_j(C_1, C_2) = \min\{\dim \mathrm{Rsupp}(V) : V \subseteq C_1, \dim V = j, V \cap C_2 = \{0\}\} \text{(RGMW)}$$
$$R_j(D_1, D_2) = \min\{\dim W^* : W \subseteq C_1, \dim W = j, W \cap C_2 = \{0\}\} \text{(RGRW)}$$

If $C_2 \subsetneqq C_1 \subseteq GF(q)^{m\times n}$ are matrix versions of $D_2 \subsetneqq D_1 \subseteq GF(q^m)^n$ then

$$M_j(C_1, C_2) = R_{mj}(D_1, D_2).$$

When $D_2 \subsetneqq D_1$ are not $GF(q^m)$-linear but $GF(q)$-linear,
$M_j(C_1, C_2) = R_{mj}(D_1, D_2)$ may be false and $R_j(D_1^\perp, D_2^\perp)$ can be meaningless to study the security performance. Even in such a case, $M_j(C_1^\perp, C_2^\perp) - 1$ gives the maximum number of wire-tapped links over which leaked information is $< j$.

Review of the Gabidulin code

A Gabidulin code $D \subset GF(q^m)^n$ with $\dim D = k$ is defined by
$\{(f(\alpha_1), \ldots, f(\alpha_n)) \mid f(x) = a_0 x + a_1 x^q + a_2 x^{q^2} + \cdots + a_{k-1} x^{q^{k-1}}, a_i \in GF(q^m)\}$
$\alpha_1, \ldots, \alpha_n$ belong to $GF(q^m)$ and linearly independent over $GF(q)$.
We must have $m \geq n$.

If the received matrix without error is X and additive errors at η links change X to Y, then $\mathrm{rank}(X - Y) \leq \eta$.
The Gabidulin code can correct up to $\lfloor (n-k)/2 \rfloor$ rank errors.

The rank distance of two matrices X, Y is $\mathrm{rank}(X - Y)$.

List decoding with respect to the rank metric

A Gabidulin code of dimension k cannot (uniquely) correct errors whose ranks $> \lfloor (n-k)/2 \rfloor$.
A list decoding algorithm with radius τ finds all codewords within rank distances $\leq \tau$ from a given received word.

Raviv and Wachter-Zeh (2016) showed that there exist exponentially many (of n) codewords within rank distances $1 + \lfloor (n-k)/2 \rfloor$ in Gabidulin codes for some $m = n$, q and k.
$\Rightarrow$ Polynomial-time list decoding is impossible for the Gabidulin codes.

List decodable rank-metric codes by Guruswami-Xing-Wang (2016)

Guruswami et al. proposed an $GF(q)$-linear code $\subset GF(q)^{m\times n}$:

- It can correct roughly twice as many errors as the Gabidulin code with the same k, n.
- It cannot be expressed as a $GF(q^m)$-linear code $\subset GF(q^m)^n$.
- It does NOT provide secrecy of encoded messages.

We will propose a list decodable secure network coding based on Guruswami et al.'s research.

Sketch of the list decodable code

The Gabidulin code of length n and dimension k was $\{(f(\alpha_1), \ldots, f(\alpha_n)) \mid f(x) = a_0 x + a_1 x^q + a_2 x^{q^2} + \cdots + a_{k-1} x^{q^{k-1}}, a_i \in GF(q^m)\}$.
$\alpha_i \in GF(q^m)$. It encodes km symbols in $GF(q)$.

$H_0, \ldots, H_{k-1}$: suitably chosen $GF(q)$-subspace of $GF(q^m)$.
Guruswami et al.'s code is defined as $\{(f(\alpha_1), \ldots, f(\alpha_n)) \mid f(x) = a_0 x + a_1 x^q + a_2 x^{q^2} + \cdots + a_{k-1} x^{q^{k-1}}, a_i \in H_i\}$.
$\alpha_i \in GF(q^n) \subset GF(q^m)$. We assume n divides m.

$a_0, \ldots, a_{k-1}$ are used as a message to be transmitted. It encodes $\sum_{i=0}^{k-1} \dim H_i$ symbols in $GF(q)$.

Parameters of the list decoding

If

- $\sum_{i=0}^{k-1} \dim H_i = km - 2mn\epsilon$,
- the list size is $O(q^{s^2/\epsilon^2})$ $(2 \leq s \leq m/n)$,

then $(n-k)s/(s+1)$ rank errors can be list-decoded.
Information rate is decreased only by 2ϵ. List size is independent of m and n.
Decoding radius $(n-k)s/(s+1)$ is almost twice as large as $\lfloor (n-k)/2 \rfloor$.

Sketch of the list decoding algorithm

$(y_1, \ldots, y_n) \in GF(q^m)^n$: a received word.
Find a nonzero polynomial $Q(x, z_1, \ldots, z_s)$ with $GF(q^m)$ coefficients s.t.
each monomial in Q has the form x^{q^i} or $z_j^{q^i}$,
$Q(\alpha_i, y_i, y_i^{q^n}, y_i^{q^{2n}}, \ldots, y_i^{q^{(s-1)n}}) = 0 \ (i = 1, \ldots, n)$

$f(x) = a_0 x + a_1 x^q + a_2 x^{q^2} + \cdots + a_{k-1} x^{q^{k-1}}$
$f^{[q^{in}]}(x) = a_0^{q^{in}} x + a_1^{q^{in}} x^q + a_2^{q^{in}} x^{q^2} + \cdots + a_{k-1}^{q^{in}} x^{q^{k-1}}$

The rank distance between $(y_1, \ldots, y_n)$ and $(f(\alpha_1), \ldots, f(\alpha_n))$
$\leq (n-k)s/(s+1)$

$\Downarrow$

$Q(x, f(x), f^{[q^n]}(x), \ldots, f^{[q^{(s-1)n}]}(x)) = 0$, which is a system of $GF(q)$-linear equations in $a_0, \ldots, a_{k-1}$.
The above algorithm works with Gabidulin codes, i.e., $a_i \in GF(q^m)$ for $i = 0, \ldots, k-1$, but the list size $q^{(s-1)k}$ can be exponential of n.

Trick for the polynomial size list

Limitation of $a_i \in H_i$ makes the list size $q^{(2m/n-2)s/\epsilon}$.
m/n can be set constant, and Guruswami et al. suggested $m/n = O(s/\epsilon)$.

Securing the list decodable codes

Goal: No information is leaked up to μ eavesdropped links.

Encoding: 1. Choose $a_i \in GF(q^m)$ randomly for $i = 0, \ldots, \mu - 1$.
2. $a_\mu \in H_\mu, \ldots, a_{k-1} \in H_{k-1}$ are chosen according to the message.
3. The codeword is $(f(\alpha_1), \ldots, f(\alpha_n))$ for
$f(x) = a_0 x + a_1 x^q + a_2 x^{q^2} + \cdots + a_{k-1} x^{q^{k-1}}$.

Decoding: $a_\mu \in H_\mu, \ldots, a_{k-1} \in H_{k-1}$ are decoded by the same procedure as Guruswami et al. $a_0, \ldots, a_{\mu-1}$ are ignored. The list size is smaller than Guruswami et al.

Why the proposal is secure

As a nested coset coding $D_2 \subsetneqq D_1 \subset GF(q^m)^n$ we have

$$
\begin{aligned}
D_2 &= \{(f(\alpha_1), \ldots, f(\alpha_n)) \mid a_0, \ldots, a_{\mu-1} \in GF(q^m), a_\mu = \cdots = a_{k-1} = 0\}, \\
D_1 &= \{(f(\alpha_1), \ldots, f(\alpha_n)) \mid a_0, \ldots, a_{\mu-1} \in GF(q^m), a_i \in H_i (i = \mu, \ldots, k-1)\}, \\
f(x) &= a_0 x + a_1 x^q + a_2 x^{q^2} + \cdots + a_{k-1} x^{q^{k-1}}
\end{aligned}
$$

Corollary (repeated)

If $\mu < M_1(D_2^\perp, D_1^\perp)$ then there is no information leakage.

$M_1(D_2^\perp, D_1^\perp) \geq M_1(D_2^\perp, \{0\})$
D_2 is a Gabidulin code, so is $D_2^\perp$
$D_2^\perp$ is MRD, thus $R_1(D_2^\perp, \{0\}) = \mu + 1$. We have $R_1(D_2^\perp, \{0\}) = M_1(D_2^\perp, \{0\})$.

References

V. Guruswami, C. Wang and C. Xing, “Explicit list-decodable rank-metric and subspace codes via subspace designs,” IEEE Trans. Inform. Theory, vol.62, pp.2707–2718, 2016.
N. Raviv and A. Wachter-Zeh, “Some Gabidulin codes cannot be list decoded efficiently at any radius,” IEEE Trans. Inform. Theory, vol.62, pp.1605–1615, 2016.
J. Kurihara, R. Matsumoto and T. Uyematsu, “Relative generalized rank weight of linear codes and its applications to network coding,” IEEE Trans. Inform. Theory, vol.61, pp.3912–3936, 2015.
U. Martínez-Peñas and R. Matsumoto, “Relative generalized matrix weights of matrix codes for universal security on wire-tap networks,” arXiv:1612.01888, 2016.