How Easy is Code Equivalence over $\mathbb{F}_q$?

Nicolas Sendrier[1] and Dimitris E. Simos[1,2]
[1] INRIA Paris-Rocquencourt
[2] SBA Research

8th International Worskhop on Coding Theory and Cryptography
WCC 2013, Bergen, Norway
April 15, 2013

Outline of the Talk

Introduction

Outline of the Talk

Introduction

Code Equivalence Problem

- Motivation
- Previous Work

Outline of the Talk

Introduction

Code Equivalence Problem
- Motivation
- Previous Work

Support Splitting Algorithm
- Mechanics
- Generalization

Outline of the Talk

secure
sba-research.org

Linear Codes

Linear Code

A linear $[n, k]$ code C of length n is a k-dimensional subspace of the finite vector space $\mathbb{F}_q^n$ and its n-bit elements are called codewords

Generator Matrix

- A $k \times n$ matrix G over $\mathbb{F}_q$, is called a generator matrix for C if the rows of G form a basis for C, so that $C = \{xG \mid x \in \mathbb{F}_q^k\}$

Hamming Space

- The Hamming distance (metric) on $\mathbb{F}_q^n$ is the following mapping,

$$d : \mathbb{F}_q^n \times \mathbb{F}_q^n \to \mathbb{N} : (x, y) \mapsto d(x, y) := \mid \{i \in \{1, 2, \dots, n\} \mid x_i \neq y_i\} \mid$$

- The pair $(\mathbb{F}_q^n, d)$ is a metric space, called the Hamming space of dimension n over $\mathbb{F}_q$, denoted by $H(n, q)$

secure sba-research.org

Equivalence of Linear Codes

Notion of Equivalence

What it means for codes to be essentially "different" but being of the same quality?

The Celebrated MacWilliams Theorem (1961)

1. Any (linear) mapping between linear codes preserving the weight of the codewords induces an equivalence for codes
2. Two codes C, C' are of the same quality if there exists a mapping $\iota : \mathbb{F}_q^n \mapsto \mathbb{F}_q^n$ with $\iota(C) = C'$ which preserves the Hamming distance, i.e. $d(\upsilon, \upsilon') = d(\iota(\upsilon), \iota(\upsilon'))$, for all $\upsilon, \upsilon' \in \mathbb{F}_q^n$
3. These distance-preserving mappings are called isometries and the codes C and C' will be called isometric

secure sba-research.org

Which are the Isometries of $H(n, q)$?

Permutation Equivalence: When $\mathbb{F}_q$, $q = 2$

- Permutation of codeword coordinates
- $C \overset{\text{PE}}{\sim} C'$, if $\exists\ \sigma \in \mathcal{S}_n$: $C' = \sigma(C) = \{\sigma(x) \mid x = (x_1, \ldots, x_n) \in C\}$ where $\sigma(x) = \sigma(x_1, \ldots, x_n) := (x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(n)})$

Monomial or Linear Equivalence: When $\mathbb{F}_q$, q is a prime

- Permutation of codeword coordinates and scaling of coordinate values
- $C \overset{\text{LE}}{\sim} C'$, if $\exists\ \iota = (\upsilon; \sigma) \in \mathbb{F}_q^{*\,n} \rtimes \mathcal{S}_n$:
  $C' = (\upsilon; \sigma)(C) = \{(\upsilon; \sigma)(x) \mid (x_1, \ldots, x_n) \in C\}$ where
  $(\upsilon; \sigma)(x_1, \ldots, x_n) := (\upsilon_1 x_{\sigma^{-1}(1)}, \ldots, \upsilon_n x_{\sigma^{-1}(n)})$

secure sba-research.org

Which are the Isometries of $H(n, q)$?

Semi-Linear Equivalence: When $\mathbb{F}_q$, $q = p^r$ is a prime power

- Permutation of codeword coordinates and scaling of coordinate values
- Application of field automorphisms in each coordinate position
- $C \overset{\text{SLE}}{\sim} C'$, if $(\upsilon; (\alpha, \sigma)) \in \mathbb{F}_q^{*\,n} \rtimes (\mathrm{Aut}(\mathbb{F}_q) \times \mathcal{S}_n)$:
  $C' = (\upsilon; (\alpha, \sigma))(C) = \{(\upsilon; (\alpha, \sigma))(x) \mid (x_i)_{i \in \mathcal{I}_n} \in C\}$ where
  $(\upsilon; (\alpha, \sigma))(x_1, \ldots, x_n) = (\upsilon_1 \alpha(x_{\sigma^{-1}(1)}), \ldots, \upsilon_n \alpha(x_{\sigma^{-1}(n)}))$

The LINEAR CODE EQUIVALENCE problem

- Parameters: n, k, q.
- Instance: two matrices $G, G' \in \mathbb{F}_q^{k \times n}$ such that $C = \langle G \rangle$, $C' = \langle G' \rangle$
- Decisional: are $\langle G \rangle \overset{\text{LE}}{\sim} \langle G' \rangle$?
- Computational: Find $(\upsilon; \sigma) \in \mathbb{F}_q^{*\,n} \rtimes \mathcal{S}_n$ such that $\langle G' \rangle = (\upsilon; \sigma)(\langle G \rangle)$

Importance of Code Equivalence

Relation to Error-Correcting Capability

Equivalent codes have the same error-correction properties
(i.e. decoding)

Relation of the Hardness of Code Equivalence in Cryptography

- The public key of the McEliece cryptosystem is a randomly permuted matrix G' of the generator matrix G of a binary Goppa code [McEliece, 1978]
- Identification schemes from error-correcting codes
  - Zero-knowledge protocols [Girault, 1990, Sendrier and Simos, 2013]

secure sba-research.org

What is known about Code Equivalence?

Complexity

PCE over $\mathbb{F}_2$ is difficult to decide in the worst case:

1. not NP-complete
2. at least as hard as GRAPH ISOMORPHISM [Petrank and Roth, 1997]
3. Recent result for $\mathbb{F}_q$: $\mathrm{GI} \preceq \mathrm{PCE}$ [Grochow, 2012]
4. PCE over $\mathbb{F}_q$ resists quantum Fourier sampling; Reduction of PCE to the HIDDEN SUBGROUP PROBLEM [Dinh, Moore and Russell, 2011]

Plan of this Talk

Exploit the average and worst-case hardness of the LINEAR CODE EQUIVALENCE problem over $\mathbb{F}_q$

What is known about Code Equivalence?

Recent Algorithms

- Mapping codes to graphs for PCE, LCE, SLCE over $\mathbb{F}_2, \mathbb{F}_3, \mathbb{F}_4$, respectively [Östergård, 2002]
- Classification of ELC orbits of bipartite graphs for PCE over $\mathbb{F}_2$ [Danielsen and Parker, 2008]
- Adaptation of Hypergraph Isomorphism algorithms for PCE over $\mathbb{F}_q$ [Babai, Codenotti and Grochow, 2011]
- Computation of canonical forms of linear codes for LCE over $\mathbb{F}_q$ [Feulner, 2009, 2011]
- Support splitting algorithm for PCE over $\mathbb{F}_q$ [Sendrier, 2000]
- No efficient algorithm for LCE or SLCE is known

Important

Can we develop a polynomial-time algorithm for settling the LINEAR CODE EQUIVALENCE problem on the average case?

The Support Splitting Algorithm (I)

$\mathcal{SSA}$

- Solves the PCE problem (decisional and computational versions)
- Partition the support $\mathcal{I}_n$ of a code $C \subseteq \mathbb{F}_2^n$ into small sets that are fixed under operations of $\mathrm{PAut}(C)$

Signatures and Invariants

- A mapping S is a signature if $S(\sigma(C), \sigma(i)) = S(C, i)$
- Property of the code and one of its positions (local property)
- S is called discriminant for C if there exist $i, j \in \mathcal{I}_n$ such that $S(C, i) \neq S(C, j)$ and fully discriminant if this holds $\forall\ i, j \in \mathcal{I}_n$
- A mapping $\mathcal{V}$ is an invariant if $C \sim C' \Rightarrow \mathcal{V}(C) = \mathcal{V}(C')$ (global property, "$\sim$" is w.r.t. to PCE but can be defined for LCE or SLCE)

secure sba-research.org

The Support Splitting Algorithm (II)

The Procedure [Sendrier, 2000]

- From given signature S and code C, we wish to build a sequence $S_0 = S, S_1, \ldots, S_r$ of signatures of increasing "discriminancy" such that S_r is fully discriminant for C (by succesive refinements of S)
- The idea is to label positions with different signature values; what remains in the end reveals a matching between codeword coordinates

Fundamental Properties of $\mathcal{SSA}$

1. $\mathcal{SSA}(C)$ **returns** a labeled partition $\mathcal{P}(S, C)$ of $\mathcal{I}_n = \{1, \ldots, n\}$
2. Assuming the existence of a fully discriminant signature, $\mathcal{SSA}(C)$ recovers the desired permutation σ of $C' = \sigma(C)$ ($\forall\ i\ \in\ \mathcal{I}_n\ \exists$ unique $j\ \in\ \mathcal{I}_n$ such that $S(C, i) = S(C', j)$ and $\sigma(i) = j$)
3. If $C' = \sigma(C)$ then $\mathcal{P}'(S, C') = \sigma(\mathcal{P}(S, C))$
4. The **output** of $\mathcal{SSA}(C)$ where $C =< G >$ is independent of G

secure sba-research.org

The Support Splitting Algorithm (III)

Dual Code

$C^\perp = \{x \in \mathbb{F}_q^n \mid \langle x, y \rangle = 0 \text{ for all } y \in C\}$ where:

1. $\langle x, y \rangle_\mathrm{E} = \sum_{i=1}^n \langle x_i, y_i \rangle_\mathrm{E} = \sum_{i=1}^n x_i y_i = x_1 y_1 + \ldots + x_n y_n \in \mathbb{F}_q$
2. $\langle x, y \rangle_\mathrm{H} = \sum_{i=1}^n \langle x_i, y_i \rangle_\mathrm{H} = \sum_{i=1}^n x_i \overline{y}_i = x_1 y_1^2 + \ldots + x_n y_n^2 \in \mathbb{F}_4$

A Good Signature

The mapping $(C, i) \mapsto \mathcal{W}_{\mathcal{H}(C_i)}(X)$, where $\mathcal{H}(C) = C \cap C^\perp$ is the hull of a code, is a signature which for random codes,

- commutes with permutations $\sigma(\mathcal{H}(C)) = \mathcal{H}(\sigma(C))$; Hence, any invariant applied to $\mathcal{H}(C)$ still remains an invariant
- easy to compute because of the small dimension [Sendrier, 1997]
- discriminant, i.e. $\mathcal{W}_{\mathcal{H}(C_i)}(X)$ and $\mathcal{W}_{\mathcal{H}(C_j)}(X)$ are "often" different

Heuristic Complexity of $\mathcal{SSA}$

Complexity of Auxiliary Functions

- Gaussian Elimination for computing $k \times n$ generator matrices: $\mathcal{O}(n^3)$
- Cost for computing $\mathcal{W}_C(X)$ for $[n, h]$ code C: $\mathcal{O}(n2^h)$

Algorithmic Cost of $\mathcal{SSA}$

Let C be a binary code of length n, and let $h = \dim(\mathcal{H}(C))$:

- First step: $\mathcal{O}(n^3) + \mathcal{O}(n2^h)$
- Each refinement: $\mathcal{O}(hn^2) + \mathcal{O}(n2^h)$
- Number of refinements: $\approx \log n$

Total (heuristic) complexity: $\mathcal{O}(n^3 + 2^h n^2 \log n)$

- When $h = \mathcal{O}(1) \Longrightarrow \mathcal{SSA}$ runs in polynomial time

41.94% of codes over $\mathbb{F}_2$ have $h = 0$, 41.94% have $h = 1$, 13.98% have $h = 2$, 0.02% have $h = 3$ and so on..

Computational vs. Decisional Code Equivalence

How are Computational and Decisional Problems Related?

- If one can explicitly solve the computational problems of code equivalence then one can also solve its corresponding decisional versions
- The other direction is also possible (Sendrier and Simos, 2012)
- Provided that for the PCE problem we have access to an oracle; An abstract version of $\mathcal{SSA}$ denoted by $\textbf{Or}_{\textbf{PCE}}(G, G') \in \{\textrm{TRUE}, \textrm{FALSE}\}$

Computational and Decisional PCE are equally hard

- Let G and G' span two $[n, k]$ linear codes C and C' over $\mathbb{F}_q$
- If $\textbf{Or}_{\textbf{PCE}}(G, G')$ is $\textrm{TRUE}$ and $\textbf{Or}_{\textbf{PCE}}(G_i, G'_j)$ is $\textrm{TRUE}$ for some $i, j \in \mathcal{I}_n$ then there exists $\sigma \in \mathcal{S}_n$ such that $C' = \sigma(C)$ and $j = \sigma(i)$
- Building block of an algorithm that retrieves the permutational part of a (semi)-linear isometry for the computational (S)LCE problems

The Closure of a Linear Code (I)

Approach for the Generalization of $\mathcal{SSA}$

- Reduce LCE or SLCE to PCE (similar approach by [Skersys, 1999])
- Recall that $\mathcal{SSA}$ solves PCE in $\mathcal{O}(n^3)$ (for "several" instances)

Closure of a Code

Let $\mathbb{F}_q = \{a_0, a_1, \ldots, a_{q-1}\}$, with $a_0 = 0$, and a linear code $C \subseteq \mathbb{F}_q^n$. Define $\mathcal{I}_{q-1}^{(n)}$ as the cartesian product of $\mathcal{I}_{q-1} \times \mathcal{I}_n$ where $\mathcal{I}_n = \{1, \ldots, n\}$. The closure $\widetilde{C}$ of the code C is a code of length $(q-1)n$ over $\mathbb{F}_q$ where,

$$\widetilde{C} = \{(a_k x_i)_{(k,i) \in \mathcal{I}_{q-1}^{(n)}} \mid (x_i)_{i \in \mathcal{I}_n} \in C\}$$

- $\widetilde{C}$ contains every possible multiplication of the coordinate x_i of a codeword $x = (x_i)_{i \in \mathcal{I}_n} \in C$ with all nonzero elements of $\mathbb{F}_q$

The Closure of a Linear Code (II)

Dependance from a Lexicographical Ordering on $\mathbb{F}_q^* = \mathbb{F}_q \backslash \{0\}$

For example, the ordering $(a_1, 1) < \ldots < (a_1, n) < (a_2, 1) < \ldots (a_2, n) < \ldots < (a_{q-1}, 1) < \ldots < (a_{q-1}, n)$ gives a total order for $\mathcal{I}_{q-1} \times \mathcal{I}_n$, and gives rise to the following closure,

$$\widetilde{C} = \{(a_1x_1, \ldots, a_1x_n, \ldots, a_{q-1}x_1, \ldots, a_{q-1}x_n) \mid (x_1, \ldots, x_n) \in C\}$$

Systematic Form of the Closure (Sendrier and Simos, 2012)

- Let p a primitive element of $\mathbb{F}_q = \{0, p, p^2, \ldots, p^{q-2}, p^{q-1} = 1\}$
- Define an ordering according to a cyclic shift of a power of p
- $\widetilde{C}_{\mathbf{sys}} = \{(x_1, px_1 \ldots, p^{q-2}x_1, \ldots, x_n, px_n \ldots, p^{q-2}x_n) \mid (x_i)_{i \in \mathcal{I}_n} \in C\}$
- Systematic form is **unique**
- Let $C, C' \subseteq \mathbb{F}_q^n$. Then $C \overset{\mathbf{LE}}{\sim} C'$, if and only if $\widetilde{C} \overset{\mathbf{PE}}{\sim} \widetilde{C'}$

secure sba-research.org

The Closure of a Linear Code (III)

The Closure is a Weakly Self-Dual Code ($C \subset C^\perp$)

$\forall\ \widetilde{x}, \widetilde{y} \in \widetilde{C}$ the Euclidean inner product is

$$\widetilde{x} \cdot \widetilde{y} = \underbrace{\left(\sum_{j=1}^{q-1} p^{2j} \right)}_{=0 \text{ over } \mathbb{F}_q,\ q \geq 5} \left(\sum_i x_i y_i \right) = 0$$

- Clearly $\dim(\mathcal{H}(\widetilde{C})) = \dim(\widetilde{C})$ and $\mathcal{SSA}$ runs in $\mathcal{O}(2^{\dim(\mathcal{H}(\widetilde{C}))})$
- The closure reduces LCE to the hard instances of $\mathcal{SSA}$ for PCE
- Exceptions are for $q = 3$ and $q = 4$ with the Hermitian inner product

Building Efficient Invariants from the Closure

- For any invariant $\mathcal{V}$ the mapping $C \longmapsto \mathcal{V}(\mathcal{H}(\widetilde{C}))$ is an invariant
- The dimension of the hull over $\mathbb{F}_q$ is on average a small constant

The Reduction of LCE to PCE

Illustration of the Reduction

- Ψ a linear isometry of the Hamming space $H(n, q)$
- τ a block-wise permutation of the generalized symmetric group $\mathcal{S}(q-1, n) := \mathcal{C}_{q-1} \wr_n \mathcal{S}_n$ (The semidirect product of n copies of $\mathcal{C}_{q-1}$ and $\mathcal{S}_n$)

$$\begin{array}{ccc} C & \xrightarrow{\Psi} & C' \\ \downarrow & & \downarrow \\ \widetilde{C} & \xrightarrow{\tau} & \widetilde{C}' \\ \downarrow & & \downarrow \\ \mathcal{H}(\widetilde{C}) & \xrightarrow{\tau} & \mathcal{H}(\widetilde{C}') \end{array}$$

- The LINEAR CODE EQUIVALENCE problem can be solved if we can retrieve Ψ from τ

An Extension of $\mathcal{SSA}$

A Good Signature for $\mathbb{F}_3$ and $\mathbb{F}_4$

- $\widetilde{\mathcal{H}(C)} = \mathcal{H}(\widetilde{C})$ (valid only for these fields)
- $S(\widetilde{C}, i) = \mathcal{W}_{\mathcal{H}(\widetilde{C}_i)}(X)$

An Efficient Algorithm for Solving LCE

- **Input**: C, C', S
  1. Compute $\widetilde{C}$ and $\widetilde{C}'$
  2. $\mathcal{P}(S, \widetilde{C}) \longleftarrow \mathcal{SSA}(\widetilde{C})$ and $\mathcal{P}'(S, \widetilde{C}') \longleftarrow \mathcal{SSA}(\widetilde{C}')$
  3. If $\mathcal{P}'(S, \widetilde{C}') = \tau(\mathcal{P}(S, \widetilde{C}))$ return τ; else $C \nsim C'$ w.r.t. LCE
  4. $\widetilde{C}' = \tau(\widetilde{C})$ and a Gaussian elimination (GE) on the permuted generator matrices of the closures will reveal the scaling coefficients
- **Note:** For SLCE we only have to consider an additional GE

secure sba-research.org

Heuristic Complexity for $\mathcal{SSA}$ and its Extension

Polynomial extension of $\mathcal{SSA}$

- For $\mathbb{F}_3$ and $\mathbb{F}_4$ but still exponential for all other cases..

Algorithm	Field (alphabet)	Random codes (average-case)	Weakly self-dual codes (worst-case)
$\mathcal{SSA}$	$\mathbb{F}_2$	$\mathcal{O}(n^3)$	$\mathcal{O}(2^k n^2 \log n)$
$\mathcal{SSA}$ extension	$\mathbb{F}_3$	$\mathcal{O}(n^3)$	$\mathcal{O}(3^k n^2 \log n)$
$\mathcal{SSA}$ extension	$\mathbb{F}_4$	$\mathcal{O}(n^3)$	$\mathcal{O}(2^{2k} n^2 \log n)$
$\mathcal{SSA}$ extension	$\mathbb{F}_q,\ q \geq 5$	$\mathcal{O}(q^k n^2 \log n)$	$\mathcal{O}(q^k n^2 \log n)$

Remark

The **hardness** of LINEAR CODE EQUIVALENCE arises from the absence of an easy computable invariant not the inexistence of an algorithm!

Can we Do Better?

What about $\mathbb{F}_q$, $q \geq 5$?

- If $C \sim C'$ w.r.t. LCE or $\mathrm{SLCE} \Longrightarrow \mathcal{H}(C) \sim \mathcal{H}(C')$ w.r.t. LCE or SLCE is **not** true
- The hull is not an invariant for LCE or SLCE over $\mathbb{F}_q$, $q \geq 5$
- (The weight enumerator) of the hull of the closure is **not** an easy computable invariant over $\mathbb{F}_q$, $q \geq 5$ (closure is weakly self-dual)

Conjecture (Sendrier and Simos, 2012)

The LINEAR CODE EQUIVALENCE problem seems to be hard for all instances over $\mathbb{F}_q$, $q \geq 5$

- Supported by some impossibility results on the Tutte polynomial of a graph which corresponds to the weight enumerator of a code
- Evaluation of weight enumerator is always hard except for a handful of points which correspond to $\mathbb{F}_q$ for $q \in \{2, 3, 4\}$ (Vertigan, 1998)

Research Problems

Related to Invariants

- Are all invariants related to the weight enumerator of a code?
- Do we already know all easy computable invariants?

Related to the Closure

- Other reductions of LCE or SLCE to PCE?

Related to Code-based Cryptography

- LCE or SLCE seems to be hard over $\mathbb{F}_q$, $q \geq 5$
- Can we build zero-knowledge protocols or other cryptographic schemes based on the hardness of LCE or SLCE?

Summary

Highlights

1. We defined the closure of a linear code
2. We presented a generalization of the support splitting algorithm for solving the LINEAR CODE EQUIVALENCE problem for $\mathbb{F}_3$ and $\mathbb{F}_4$
3. We conjectured that the LINEAR CODE EQUIVALENCE problem over $\mathbb{F}_q$, $q \geq 5$ is hard for almost all instances

secure
sba-research.org

Summary

Highlights

1. We defined the closure of a linear code
2. We presented a generalization of the support splitting algorithm for solving the LINEAR CODE EQUIVALENCE problem for $\mathbb{F}_3$ and $\mathbb{F}_4$
3. We conjectured that the LINEAR CODE EQUIVALENCE problem over $\mathbb{F}_q$, $q \geq 5$ is hard for almost all instances

Future Work

Solve (some) of the research problems..!

References

László Babai, Paolo Codenotti, Joshua Grochow and Youming Qiao, "Code equivalence and group isomorphism," In Proc. 22nd Ann. Symp. on Discrete Algorithms (SODA 2011), pages 1395-1408. ACM-SIAM, 2011.

E. Petrank and R. M. Roth, "Is code equivalence easy to decide?," IEEE Trans. Inf. Theory, vol. 43, pp. 1602–1604, 1997.

N. Sendrier, "On the dimension of the hull," SIAM J. Discete Math., vol. 10, pp. 282–293, 1997.

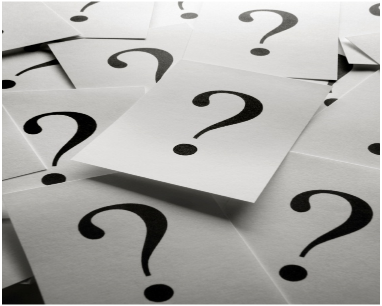
N. Sendrier, "Finding the permutation between equivalent codes: the support splitting algorithm," IEEE Trans. Inf. Theory, vol. 46, pp. 1193–1203, 2000.

N. Sendrier and D. E. Simos, "The Hardness of Code Equivalence over $\mathbb{F}_q$ and its Application to Code-based Cryptography," to appear in Proceedings of the 5th International Conference on Post-Quantum Cryptography (PQCrypto 2013),Preprint, 2013.

D. Vertigan, "Bicycle dimension and special points of the Tutte polynomial," Journal of Comb. Theory, Series B, vol. 74, pp. 378–396, 1998

Questions - Comments

Thanks for your Attention!



{nicolas.sendrier,dimitrios.simos}@inria.fr
dsimos@sba-research.org