CONJURA · OPEN PROBLEM

The Linear-Test Bias of Multi-Disjoint Syndrome Decoding

c sparse error vectors with pairwise disjoint supports, one shared random parity-check matrix

Status. Statement: AI-written from Adrià Gascón, Yuval Ishai, Mahimna Kelkar, Baiyu Li, Yiping Ma and Mariana Raykova, *Computationally Secure Aggregation and Private Information Retrieval in the Shuffle Model*, IACR ePrint 2024/870. Not yet checked by a human. The source proves a (non-tight) reduction from standard syndrome decoding, and conjectures separately that a known generic attack is optimal; the linear-test bias below is the one quantity it states it cannot bound.

Category. *Information-Theoretic Cryptography.*

Conjecture (informal). *Sample one uniformly random $n \times m$ parity-check matrix H over a finite field and c error vectors of Hamming weight w whose supports are pairwise disjoint, and publish H together with the c syndromes $He_1, \ldots, He_c$. The source conjectures this is hard to distinguish from uniform, and it is the assumption its secure-aggregation and PIR protocols rest on. The standard first evidence for a distribution of this shape is a* linear test*: show that no fixed linear function of the syndromes has noticeable bias. The source identifies this as the natural route and states it does not know how to carry it out. The conjecture below is that the route closes — that every nonzero linear test against this distribution has bias exponentially small in the code's dual distance, exactly as for standard (and regular) LPN.*

1 The setting

Gascón et al. study the "split-and-mix" paradigm for secure aggregation in the shuffle model, where each of many clients splits its input into additive shares and the shares of all clients are delivered to the server in a uniformly random order. Their contribution is to replace the information-theoretic analysis of this paradigm [IKOS06, BBGN20, GMPV20] with a computational one,

which lets each client send far fewer shares. The computational problem the reduction lands on is new to that work: multi-disjoint syndrome decoding (MDSD, Definition 2), the syndrome-decoding problem in which several error vectors share one parity-check matrix and their supports are constrained to be pairwise disjoint — the disjointness being exactly what the shuffling of one client's shares produces.

The source gives two kinds of evidence for hardness. First, a reduction from standard decisional syndrome decoding (dual LPN), which it states is not tight and does not correspond to an efficient distinguisher. Second, a conjecture that the DOOM algorithm of Sendrier [Sen11], a generic information-set-decoding extension exploiting multiple syndromes, is the best attack; the concrete parameters of the source's protocols are set from that conjecture.

What is missing is the check that the community treats as the entry fee for a new LPN-style assumption: the *linear test* framework of Boyle et al. [BCG+20], under which one bounds, for every fixed nonzero linear functional applied to the samples, the bias of the resulting bit (or field element). For standard LPN and for regular LPN this bound follows from the dual distance of the underlying code, and it rules out the whole class of attacks that only take linear combinations of samples. The source names this route and records that it stalls: "We do not know how to bound such bias for the MDSD distribution, and we leave it as an open question."

The obstruction it points at is structural. In the MDSD distribution the c error vectors are *not* independent: conditioning on the support of e_1 removes those coordinates from the supports available to $e_2, \ldots, e_c$. A linear test may combine coordinates across different syndromes, and the standard dual-distance argument — which bounds the bias of a single sparse error vector against a single fixed test vector — gives nothing directly about a test that couples the syndromes whose errors are correlated in this way.

2 Notation and parameters

- $\mathbb{F}$ is a finite field; $q = |\mathbb{F}|$. The source's default is $\mathbb{F} = \mathbb{F}_2$ for its PIR protocol and a large prime field for aggregation.
- n is the number of parity-check rows (the syndrome length), m the code length, so $H \in \mathbb{F}^{n \times m}$.

- $c \geq 1$ is the number of error vectors sharing H; $w \geq 1$ is the Hamming weight of each. The source's regime is $w = O(\log m)$ and $m \approx cw + \lambda$.
- $\Delta(e)$ is the Hamming weight of $e \in \mathbb{F}^m$, and $\mathrm{supp}(e) \subseteq [m]$ its support.
- For a random variable X over $\mathbb{F}$, $\mathrm{bias}(X) := \max_{a \in \mathbb{F}} \Pr[X = a] - 1/q$.

3 The conjecture

Definition 1 (Disjoint error set, [GIKLMR24, Definition 7]). For $m, c, w > 0$, the (m, c, w)-*disjoint error set* over $\mathbb{F}$ is

$$\mathsf{DisError}_{m,c,w} = \left\{ E = (e_1 \mid \cdots \mid e_c) \in \mathbb{F}^{m \times c} \;:\; \mathrm{supp}(e_i) \cap \mathrm{supp}(e_j) = \emptyset \; \forall i \neq j, \right.$$

Definition 2 (Multi-disjoint syndrome decoding, [GIKLMR24, Definition 8]). The (n, m, c, w)-*MDSD distribution* over $\mathbb{F}$ is

$$\mathsf{MDSD}_{n,m,c,w} = \left\{ (H, H \cdot E) \;:\; H \leftarrow \mathbb{F}^{n \times m},\; E \leftarrow \mathsf{DisError}_{m,c,w} \right\},$$

both draws uniform. The decisional (n, m, c, w)-MDSD problem is to distinguish $\mathsf{MDSD}_{n,m,c,w}$ from the uniform distribution over $\mathbb{F}^{n \times m} \times \mathbb{F}^{n \times c}$.

Definition 3 (Linear test against MDSD). A *linear test* is a nonzero matrix $V \in \mathbb{F}^{n \times c}$, applied to a sample $(H, Y) \leftarrow \mathsf{MDSD}_{n,m,c,w}$ as the field element $\langle V, Y \rangle := \sum_{i \in [n]} \sum_{j \in [c]} V_{i,j} Y_{i,j}$. Its *bias on* H is

$$\mathrm{bias}_V(H) := \mathrm{bias}\big(\langle V, H \cdot E \rangle\big), \qquad E \leftarrow \mathsf{DisError}_{m,c,w},$$

the probability being over E alone, with H fixed.

Conjecture 1 (Linear-test bias of MDSD). *There exist constants $\alpha, \beta > 0$ such that the following holds for every finite field $\mathbb{F}$ and all n, m, c, w with $cw \leq m$ and $w \leq \alpha n / \log m$. With probability at*

least $1 - 2^{-\beta n}$ *over the choice of* $H \leftarrow \mathbb{F}^{n \times m}$,

$$\max_{V \in \mathbb{F}^{n \times c} \setminus \{0\}} \mathrm{bias}_V(H) \;\leq\; 2^{-\beta w}.$$

Remark 1 (What the conjecture is and is not). Conjecture 1 is not the hardness of MDSD, and proving it would not establish that hardness: a linear-test bound rules out one attack class (distinguishers that read only a fixed linear function of the samples) and says nothing about the rest. It is the step the source names as the natural route to studying the assumption and reports it cannot take. Refuting it — exhibiting a fixed V with noticeable bias for most H, in the source's parameter regime — would in contrast be a genuine distinguishing attack, and would bear directly on the concrete parameters the source's protocols are instantiated with.

Remark 2 (The exponent is the object of interest). The two constants are stated as a single pair because the source gives no target rate: it names the framework and the quantity, not a bound to aim at. A proof that matches what is known for regular LPN [HOSS22] — bias exponentially small in the error weight w, for all but an exponentially small fraction of H — is the natural target, and is what Conjecture 1 asserts. A weaker quantitative bound, e.g. bias $\leq 2^{-\beta w/\log m}$ under the same quantifiers, would already be a substantive resolution of the source's open question, and any such result should be reported with its own exponent rather than folded into this statement.

Remark 3 (The binary-error variant). The source also defines beMDSD, in which the error entries are restricted to $\{0,1\}$ inside a large field, and states that over $\mathbb{F}_2$ the two problems coincide. Conjecture 1 is stated for MDSD; the corresponding question for beMDSD over a large field is a separate statement, since the error alphabet changes which linear combinations can cancel.

4 Bibliography

[GIKLMR24] Adrià Gascón, Yuval Ishai, Mahimna Kelkar, Baiyu Li, Yiping Ma and Mariana Raykova. *Computationally secure aggregation and private information retrieval in the shuffle model.* IACR ePrint 2024/870.

The source. MDSD is Definitions 7 and 8, page 12; the linear-test open question is in Section 3.3.1, page 17.

[BCG+20] Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl and Peter Scholl. *Correlated pseudorandom functions from variable-density LPN.* 61st Annual Symposium on Foundations of Computer Science (FOCS), 2020, pages 1069–1080. The linear test framework the source names.

[Sen11] Nicolas Sendrier. *Decoding one out of many.* PQCrypto 2011. The DOOM algorithm the source conjectures to be optimal against MDSD.

[HOSS22] Carmit Hazay, Emmanuela Orsini, Peter Scholl and Eduardo Soria-Vazquez. *TinyKeys: a new approach to efficient multi-party computation.* Journal of Cryptology, 35(2):13, April 2022. The regular-LPN analysis the source appeals to for the analogy.

[IKOS06] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky and Amit Sahai. *Cryptography from anonymity.* FOCS 2006. Introduces the split-and-mix paradigm the source makes computational.

[BBGN20] Borja Balle, James Bell, Adrià Gascón and Kobbi Nissim. *Private summation in the multi-message shuffle model.* ACM CCS 2020, pages 657–676.

[GMPV20] Badih Ghazi, Pasin Manurangsi, Rasmus Pagh and Ameya Velingker. *Private aggregation from fewer anonymous messages.* EUROCRYPT 2020, Part II, LNCS 12106, pages 798–827.