Expander Graphs Are Non-Malleable Codes

Peter Michael Reichstein Rasmussen
Basic Algorithms Research Copenhagen, University of Copenhagen, Denmark
pmrr@di.ku.dk

Amit Sahai
UCLA, Los Angeles, CA, USA
sahai@cs.ucla.edu

Abstract

Any d-regular graph on n vertices with spectral expansion λ satisfying $n = \Omega(d^3 \log(d)/\lambda)$ yields a $O\left(\frac{\lambda^{3/2}}{d}\right)$-non-malleable code for single-bit messages in the split-state model.

2012 ACM Subject Classification Theory of computation → Cryptographic primitives; Mathematics of computing → Spectra of graphs

Keywords and phrases Non-Malleable Code, Expander Graph, Mixing Lemma

Digital Object Identifier 10.4230/LIPIcs.ITC.2020.6

Funding *Peter Michael Reichstein Rasmussen*: Supported in part by grant 16582, Basic Algorithms Research Copenhagen (BARC), from the VILLUM Foundation.
Amit Sahai: Supported in part from a DARPA/ARL SAFEWARE award, NSF Frontier Award 1413955, and NSF grant 1619348, BSF grant 2012378, a Xerox Faculty Research Award, a Google Faculty Research Award, an equipment grant rom Intel, and an Okawa Foundation Research Grant. This material is based upon work supported by the Defense Advanced Research Projects Agency through the ARL under Contract W911NF-15-C- 0205. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense, the National Science Foundation, or the U.S. Government.

Acknowledgements A significant effort was made to simplify our proof as much as possible, which eventually resulted in the approximately 2-page proof of our main result presented here; we thank Anders Aamand and Jakob Bæk Tejs Knudsen for suggestions and insights regarding the main theorem that helped simplify and improve the results presented. Furthermore, we thank Aayush Jain, Yuval Ishai, and Dakshita Khurana for early discussions regarding simple constructions of split-state non-malleable codes not based on expander graphs.

1 Introduction

A key goal in theoretical computer science is the identification of structures that exhibit resilience to adversarial tampering. The classical notion in this space is that of an error-detection or error-correction code, where we seek to ensure that tampering caused by an adversary that can modify a *bounded* number of symbols in a codeword can be detected or corrected.

But what if the number of errors that an adversary can introduce is unbounded? The objective of error detection or correction is clearly impossible to achieve in this setting – the adversary can simply replace the transmitted codeword with an encoding of some other fixed value. Thus, the main question of study in this context concerns the notion of *malleability*: informally speaking, our core goal must be to prevent the adversary from replacing an encoding of a value x with an encoding of some other related value $\tilde{x} \neq x$.

The central information-theoretic object in this setting is called a split-state non-malleable code [5]. Since their introduction in 2010 [5], split-state non-malleable codes have been the subject of intense study within theoretical computer science [5, 4, 1, 3, 2, 6]. Here,

1st Conference on Information-Theoretic Cryptography (ITC 2020).
Editors: Yael Tauman Kalai, Adam D. Smith, and Daniel Wichs; Article No. 6; pp. 6:1–6:10
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

we consider the most basic form of a split-state non-malleable code, namely a code for encoding a single bit. A split-state non-malleable code [5] for single-bit messages consists of randomized encoding and decoding algorithms $(\mathrm{enc}, \mathrm{dec})$. A message $m \in \{0,1\}$ is encoded as a pair of strings $(L, R) \in \{0,1\}^k \times \{0,1\}^k$, such that $\mathrm{dec}(L, R) = m$. An adversary then specifies an arbitrary pair of functions $g, h : \{0,1\}^k \to \{0,1\}^k$. The code is said to be non-malleable if, intuitively, the message obtained as $\mathrm{dec}(g(L), h(R))$ is "unrelated" to the original message m. In particular, to be ε-non-malleable, it is enough [4] to guarantee that when the message m is chosen uniformly at random and encoded into (L, R), the probability that $\mathrm{dec}(g(L), h(R)) = 1 - m$ is at most $\frac{1}{2} + \varepsilon$.

1.1 Previous Work

All known constructions and proofs of security for explicit split-state non-malleable codes have required complex mathematical proofs, and all known such proofs either directly or indirectly used the mathematics behind constructions of two-source extractors [4, 1, 3, 2, 6]. In fact, after constructing the first non-malleable code in the split-state model Dziembowski, Kazana, and Obremski wrote: "This brings a natural question if we could show some relationship between the extractors and the non-malleable codes in the split-state model. Unfortunately, there is no obvious way of formalizing the conjecture that non-malleable codes need to be based on extractors" [4].

1.2 Our Contribution

In this work, we seek to establish new, simpler, foundations for the construction of single-bit split-state non-malleable codes. We do so by answering in the negative the implicit conjecture of [4]; we show that it is not necessary to base constructions of non-malleable codes on the theory of extractors.

Specifically, we show that expander graphs immediately give rise to split-state non-malleable codes for single-bit messages. We prove that any d-regular graph on $n = 2^k$ nodes with spectral expansion λ satisfying $n = \Omega(d^3 \log(d)/\lambda)$ yields a $O\left(\frac{\lambda^{3/2}}{d}\right)$-non-malleable code for single-bit messages in the split-state model. Our proof is elementary, requiring a little more than two pages to prove, having at its heart two nested applications of the Expander Mixing Lemma. Furthermore, we only need expanders of high degree (e.g., $d = n^{1/3}$), which can be constructed and analyzed easily (see, e.g., [7] or Appendix C), yielding $2^{-\Omega(k)}$-non-malleable codes. It is worth noting that the manner in which we construct a single-bit code from an expander graphs is similar to how [4] constructs a single-bit code from a two-source extractor. Thus, our main discovery is that expander graphs suffice for such a construction to succeed.

Our construction of non-malleable codes from expander graphs thus opens up a new line of attack in the study of split-state non-malleable codes. It is important to keep in mind that current constructions of non-malleable codes supporting messages of arbitrary length use many ideas pioneered in the construction of [4], in particular the use of extractors. While we do not yet know how to generalize our results beyond single-bit messages, we speculate that further investigation building upon our work will reveal a deeper connection and more powerful simple constructions based on expanders.

It should be noted that two-source extractors are well-known to exhibit expansion properties; however, in all previous proofs, much more than mere expansion was used to argue non-malleability. Indeed previous proofs apply extractors repeatedly; for instance the proof of [4] uses the extractor property multiple times (e.g., in equation (22) and using equation (43) in [4]). We also note that it is not surprising that 1-bit non-malleable codes will exhibit some sort of expansion properties. Our contribution is the converse: that good expansion is *sufficient* for the construction of non-malleable codes.

1.3 Parameters and a Comparison with DKO13

For completeness we include an analysis of the concrete parameters of our resulting code. Let $\gamma > 0$ be given. Our construction yields a 1-bit γ-non-malleable split-state code where each part of the message is a vertex of a d-regular graph G on n vertices. The graph G must have expansion λ and satisfy $n = \Omega(d^3 \log(d)/\lambda)$ and $\lambda^{3/2}/d = O(\gamma)$. Suppose that G has expansion $\Theta(\sqrt{d})$, which is the case for the instantiation in Appendix C. We may then set $d = \Theta((1/\gamma)^4)$ and $n = \tilde{\Theta}((1/\gamma)^{10})$. Thus, our code uses space $20 \log(1/\gamma) + O(\log\log(1/\gamma))$ to encode a single bit. The instantiation from Appendix C is not able to choose n as flexibly as suggested here and uses space $24 \log(1/\gamma)$. The time taken to encode and decode a message in this instantiation is $O(\log(1/\gamma))$. In comparison, the instantiation of [4] uses space around $90 \log(1/\gamma)$ and the time to encode and decode is $O(\log(1/\gamma) \log^2(\log(1/\gamma)))$. It should be noted, however, that the construction of [4] supports leakage as well, something that we are not considering in this paper.

1.4 Intuition behind our construction and analysis

Every graph, $G = (V, E)$ yields a single-bit split-state code in the following straightforward manner: To encode 1, pick an edge $(v, u) \in E$ uniformly at random, and set the left encoding to be v and the right encoding to be u. To encode a 0, do the same with a uniformly random non-edge in the graph.

Our analysis proceeds in two parts. First, in Proposition 6, using only elementary manipulations, we give an exact characterization of the success probability of any particular tampering split-state adversary against the code associated with any graph. The split-state adversary uses two functions, g and h, to tamper with the left and right encodings, respectively. The significant term of the probability to be analyzed is the quantity

$$\sum_{(v,u)\in E} \left(\frac{d\left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right|}{n} - \left|E(g^{-1}(v), h^{-1}(u))\right| \right).$$

To bound this expression, we make the following observations. First, sparsity of the graph allows us to bound many of the terms immediately. Second, the term in the parentheses above immediately suggests a bound using the Expander Mixing Lemma, applied to the number of edges from $g^{-1}(v)$ to $h^{-1}(u)$. Third, we observe that the sum itself is over edges $(v, u) \in E$, and furthermore, the remaining sum of problematic terms are a sum over edges of the form $(v, u) \in E \cap (T \times S)$ for some vertex subsets $S, T \subset V$. This allows us to apply the Expander Mixing Lemma a *second* time, effectively bounding the number of "error terms" that accumulate through the initial use of the Expander Mixing Lemma. The actual analysis of this bound is just over a page of calculation. The analysis follows the intuition above, modulo a partitioning of terms into sets of appropriate size for the analysis to work.

2 Preliminaries

We shall assume familiarity with the basics of codes and non-malleable codes. A cursory review of relevant definitions can be found in the appendix.

▶ **Notation 1** (Graphs). *A graph $G = (V, E)$ consists of vertices V and edges $E \subset V \times V$. In this exposition every graph is undirected and $n = |V|$ always denotes the number of vertices of the graph in question.*

- *For any $v \in V$ we denote by $N(v)$ the set of neighbors of v in G.*
- *For any two subsets $S, T \subseteq V$ we denote by $E(S, T)$ the set of (directed) edges from S to T in G. I.e. $E(S, T) = \{(v, u) \in S \times T \mid (v, u) \in E\}$.*

▶ **Definition 2** (Spectral Expander). *Let $G = (V, E)$ be a d-regular graph, A_G be its adjacency matrix, and $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of A_G. We say that G is a λ spectral expander if $\lambda \geq \max\{|\lambda_2|, \ldots, |\lambda_n|\}$.*

▶ **Theorem 3** (Expander Mixing Lemma). *Suppose that $G = (V, E)$ is a λ spectral expander. Then for every pair of subsets $S, T \subset V$ we have*

$$\left| |E(S,T)| - \frac{d \cdot |S| \cdot |T|}{n} \right| \leq \lambda \sqrt{|S| \cdot |T|}.$$

Our results will rely on the following characterization of 1-bit non-malleable codes by Dziembowski, Kazana, and Obremski found in [4].

▶ **Theorem 4.** *Let $(\mathrm{enc}, \mathrm{dec})$ be a coding scheme with $\mathrm{enc}\colon \{0,1\} \to \mathcal{X}$ and $\mathrm{dec}\colon \mathcal{X} \to \{0,1\}$. Further, let $\mathcal{F}$ be a set of functions $f\colon \mathcal{X} \to \mathcal{X}$. Then $(\mathrm{enc}, \mathrm{dec})$ is ε-non-malleable with respect to $\mathcal{F}$ if and only if for every $f \in \mathcal{F}$,*

$$\Pr_{b \overset{u}{\leftarrow} \{0,1\}} (\mathrm{dec}(f(\mathrm{enc}(b))) = 1 - b) \leq \frac{1}{2} + \varepsilon,$$

where the probability is over the uniform choice of b and the randomness of enc.

3 Results

We first formally introduce our candidate code and then prove that it is a non-malleable code.

3.1 Candidate Code

From a graph we can very naturally construct a coding scheme as follows.

▶ **Definition 5** (Graph Code). *Let $G = (V, E)$ be a graph. The associated* graph code, *$(\mathrm{enc}_G, \mathrm{dec}_G)$, consists of the functions*

$$\mathrm{enc}_G\colon \{0,1\} \to V \times V, \qquad\qquad \mathrm{dec}_G\colon V \times V \to \{0,1\}$$

which are randomized and deterministic, respectively, and given by

$$\mathrm{enc}_G(b) = \begin{cases} (u,v) \overset{u}{\leftarrow} (V \times V) \setminus E, & b = 0, \\ (u,v) \overset{u}{\leftarrow} E, & b = 1, \end{cases}$$

$$\mathrm{dec}_G(v_1, v_2) = \begin{cases} 0, & (v_1, v_2) \notin E, \\ 1, & (v_1, v_2) \in E. \end{cases}$$

3.2 Non-Malleability of Expander Graph Codes

Finally, arriving at the core of the matter, we first establish the following lemma casting the expression of Theorem 4 in terms of graph properties.

▶ **Proposition 6.** *Let $G = (V, E)$ be a graph, functions $g, h \colon V \to V$ be given, and $f = (g, h) \colon V \times V \to V \times V$ satisfy $f(u, v) = (g(u), h(v))$. For the probability that f flips a random bit encoded by enc_G, write*

$$T = \Pr_{b \overset{u}{\leftarrow} \{0,1\}} (\mathrm{dec}_G(f(\mathrm{enc}_G(b))) = 1 - b)$$

where the probability is taken over the randomness of enc_G and the sampling of b. Then

$$T = \frac{1}{2} + \frac{1}{2d(n-d)} \sum_{(v,u)\in E} \left(\frac{d\left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right|}{n} - \left|E(g^{-1}(v), h^{-1}(u))\right| \right). \tag{1}$$

Proof. For $b \in \{0, 1\}$ denote by Q_b the probability

$$Q_b = \Pr(\mathrm{dec}_G(f(\mathrm{enc}_G(b))) = 1 - b)$$

taken over the randomness of enc_G. It is clear that $T = \frac{Q_0+Q_1}{2}$ and that by definition

$$Q_0 = \Pr_{(v,u) \overset{u}{\leftarrow} V\times V \setminus E} [(g(v), h(u)) \in E], \qquad Q_1 = \Pr_{(v,u) \overset{u}{\leftarrow} E} [(g(v), h(u)) \notin E].$$

First, for $b = 0$ we see that the number of non-edges that are mapped by f to any given $(v, u) \in E$ is given by $\left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right| - \left|E(g^{-1}(v), h^{-1}(u))\right|$. There are $n(n-d)$ non-edges in G so it follows that

$$Q_0 = \frac{\sum_{(v,u)\in E} \left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right| - \left|E(g^{-1}(v), h^{-1}(u))\right|}{n(n-d)}.$$

Second, for $b = 1$ the number of edges of G that are mapped to non-edges by f is given by $\sum_{(v,u)\notin E} \left|E(g^{-1}(v), h^{-1}(u))\right|$. Since there are dn edges of G to choose from when encoding the bit $b = 1$,

$$Q_1 = \frac{\sum_{(v,u)\notin E} \left|E(g^{-1}(v), h^{-1}(u))\right|}{dn}.$$

Now, observing that the number of (directed) edges in the graph is dn and that $\{g^{-1}(v)\}_{v\in V}$ and $\{h^{-1}(u)\}_{u\in V}$ are both partitions of V, we get

$$Q_1 = \frac{dn - \sum_{(v,u)\in E} \left|E(g^{-1}(v), h^{-1}(u))\right|}{dn} = 1 - \frac{\sum_{(v,u)\in E} \left|E(g^{-1}(v), h^{-1}(u))\right|}{dn}.$$

Putting it all together,

$$\begin{aligned} T &= \frac{\sum_{(v,u)\in E} \left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right| - \left|E(g^{-1}(v), h^{-1}(u))\right|}{2n(n-d)} + \frac{1}{2} - \frac{\sum_{(v,u)\in E} \left|E(g^{-1}(v), h^{-1}(u))\right|}{2dn} \\ &= \frac{1}{2} + \frac{1}{2d(n-d)} \sum_{(v,u)\in E} \left(\frac{d\left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right|}{n} - \left|E(g^{-1}(v), h^{-1}(u))\right| \right). \end{aligned}$$ ◀

We proceed immediately with the main theorem, which concludes the exposition. In order to keep this presentation short and to the point, more elaborate calculations, which avoid the log-factors, have been placed in the appendix as Theorem 10.

▶ **Theorem 7.** *Let $G = (V, E)$ be d-regular with spectral expansion λ satisfying $n = \Omega(d^3 \log(d)^4/\lambda)$. Then $(\mathrm{enc}_G, \mathrm{dec}_G)$ is an $\tilde{O}\left(\frac{\lambda^{3/2}}{d}\right)$-non-malleable code in the split-state model.*

Proof. Let $f = (g, h)\colon V \times V \to V \times V$ be given. By Theorem 4 and Proposition 6 we just need to show that

$$R = \frac{1}{2d(n-d)} \cdot \sum_{(v,u)\in E} \left(\frac{d\left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right|}{n} - \left|E(g^{-1}(v), h^{-1}(u))\right| \right)$$

is bounded by $\tilde{O}\left(\frac{\lambda^{3/2}}{d}\right)$. Define the sets

$$G_1 = \left\{ v \in V \mid \left|g^{-1}(v)\right| > \frac{n}{d^2} \right\}, \qquad H_1 = \left\{ u \in V \mid \left|h^{-1}(u)\right| > \frac{n}{d^2} \right\},$$
$$G_2 = \left\{ v \in V \mid \left|g^{-1}(v)\right| \le \frac{n}{d^2} \right\}, \qquad H_2 = \left\{ u \in V \mid \left|h^{-1}(u)\right| \le \frac{n}{d^2} \right\},$$

for $i, j \in \{1, 2\}$ write

$$R_{i,j} = \frac{1}{2d(n-d)} \sum_{(v,u)\in E\cap(G_i \times H_j)} \left(\frac{d\left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right|}{n} - \left|E(g^{-1}(v), h^{-1}(u))\right| \right),$$

and observe that $R = \sum_{1\le i,j\le 2} R_{i,j}$.

Consider the case when $i = 2$. Simply bounding the terms of the form $\left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right|$ by using that each vertex has only d neighbours, we get

$$\begin{aligned} R_{2,1} + R_{2,2} &\le \frac{1}{2n(n-d)} \sum_{(v,u)\in E\cap(G_2\times V)} \left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right| \\ &\le \frac{1}{2n(n-d)} \cdot d \cdot \sum_{u\in V} \frac{n}{d^2} \cdot \left|h^{-1}(u)\right| \\ &= \frac{n}{2(n-d)d}. \end{aligned}$$

Thus, $R_{2,1} + R_{2,2} = O\left(d^{-1}\right)$. By symmetry, $R_{1,2} = O\left(d^{-1}\right)$. It only remains to show that $R_{1,1} = \tilde{O}\left(\frac{\lambda^{3/2}}{d}\right)$. To this end, partition G_1 and H_1, respectively, as

$$G_1^k = \left\{ v \in G_1 \mid \frac{n}{2^{k-1}} \ge \left|g^{-1}(v)\right| > \frac{n}{2^k} \right\}, \qquad H_1^l = \left\{ v \in H_1 \mid \frac{n}{2^{l-1}} \ge \left|h^{-1}(u)\right| > \frac{n}{2^l} \right\}$$

for $1 \le k, l \le \left\lceil \log_2\left(d^2\right) \right\rceil$. Now, focusing on each pair G_1^k and H_1^l, we write

$$S_{k,l} = \frac{1}{2d(n-d)} \sum_{(v,u)\in E\cap(G_1^k \times H_1^l)} \left(\frac{d\left|g^{-1}(v)\right| \cdot \left|h^{-1}(u)\right|}{n} - \left|E(g^{-1}(v), h^{-1}(u))\right| \right)$$

and apply first the mixing lemma then the Cauchy-Schwartz inequality to get

$$\begin{aligned} 2d(n-d)S_{k,l} &= \sum_{v\in G_1^k} \left(\frac{d\left|g^{-1}(v)\right| \cdot \sum_{u\in N(v)\cap H_1^l} \left|h^{-1}(u)\right|}{n} - \left| E\left(g^{-1}(v), \bigcup_{u\in N(v)\cap H_1^l} h^{-1}(u) \right) \right| \right) \\ &\le \sum_{v\in G_1^k} \lambda \sqrt{|g^{-1}(v)| \cdot \sum_{u\in N(v)\cap H_1^l} |h^{-1}(u)|} \\ &\le \lambda \sqrt{\frac{n}{2^{k-1}} \cdot \frac{n}{2^{l-1}}} \cdot \sum_{v\in G_1^k} \sqrt{\left|N(v) \cap H_1^l\right|} \\ &\le 2\lambda n \cdot 2^{-\frac{l+k}{2}} \cdot \sqrt{\left|G_1^k\right|} \cdot \sqrt{\left|E(G_1^k, H_1^l)\right|}. \end{aligned}$$

We use the fact that $\left|G_1^k\right| \leq 2^k, \left|H_1^l\right| \leq 2^l$, apply the mixing lemma to the last factor, and wield Jensen's inequality on the arising square root to obtain

$$\begin{aligned} d(n-d)S_{k,l} &\leq \lambda n \cdot 2^{-\frac{l+k}{2}} \cdot \sqrt{\left|G_1^k\right|} \cdot \sqrt{\frac{d \cdot \left|G_1^k\right| \cdot \left|H_1^l\right|}{n} + \lambda\sqrt{\left|G_1^k\right| \cdot \left|H_1^l\right|}} \\ &\leq \lambda\sqrt{2^k dn} + 2^{\frac{k-l}{4}}\lambda^{3/2}n \leq \lambda \cdot \sqrt{d^3 n} + 2^{\frac{k-l}{4}}\lambda^{3/2}n. \end{aligned}$$

By symmetry of k and l, $d(n-d)S_{k,l} \leq \lambda \cdot \sqrt{d^3 n} + 2^{\frac{l-k}{4}}\lambda^{3/2}n$. Thus,

$$\begin{aligned} R_{1,1} &= \sum_{1 \leq k,l \leq \lceil \log_2(d^2) \rceil} S_{k,l} \\ &\leq O\left(\frac{\lambda \log(d)^2 \cdot \sqrt{d}}{\sqrt{n}}\right) + O\left(\frac{\lambda^{3/2}}{d}\right) \cdot \sum_{1 \leq k,l \leq \lceil \log_2(d^2) \rceil} 2^{-\frac{|k-l|}{4}} \\ &= O\left(\frac{\log(d)\lambda^{3/2}}{d}\right). \end{aligned}$$

◀

References

1 Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In *Symposium on Theory of Computing, STOC*, 2014.

2 Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. In *Symposium on Theory of Computing, STOC*, 2016.

3 Eshan Chattopadhyay and David Zuckerman. Non-malleable codes against constant split-state tampering. In *Foundations of Computer Science, FOCS*, 2014.

4 Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Non-malleable codes from two-source extractors. In *CRYPTO*, 2013.

5 Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. In *ICS*, 2010.

6 Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Symposium on Theory of Computing, STOC*, 2017.

7 Luca Trevisan. Luca trevisan's 'in theory' blog. `https://lucatrevisan.wordpress.com/2011/02/28/cs359g-lecture-16-constructions-of-expanders/`. Accessed: 2018-09-27.

A Definitions for Split-State Non-Malleable Codes

Here, we recall the basic definition of a split-state non-malleable code due to [5].

▶ **Definition 8** (Coding scheme). *We define a* coding scheme *to be a pair of functions* $(\mathrm{enc}, \mathrm{dec})$*. The encoding function* $\mathrm{enc}\colon \mathcal{M} \to \mathcal{X}$ *is randomized while the decoding function* $\mathrm{dec}\colon \mathcal{X} \to \mathcal{M} \cup \{\perp\}$ *is deterministic. Further, for all* $s \in \mathcal{M}$ *the pair satisfies*

$$\Pr[\mathrm{dec}(\mathrm{enc}(s)) = s] = 1$$

where the probability is taken over the randomness of enc.

▶ **Definition 9** (Split State Non-Malleable Code). *A coding scheme* (enc, dec)*,* $\mathrm{enc}\colon \mathcal{M} \to \mathcal{L} \times \mathcal{R}$ *and* $\mathrm{dec}\colon \mathcal{L} \times \mathcal{R} \to \mathcal{M} \cup \{\perp\}$*, is* ε*-non-malleable in the split state model if for every pair of functions* $g\colon \mathcal{L} \to \mathcal{L}, h\colon \mathcal{R} \to \mathcal{R}$ *and writing* $f = (g, h)$ *there exists a distribution* D_f

supported on $\mathcal{M} \cup \{*, \perp\}$ such that for every $s \in \mathcal{M}$ the two random variables defined by the experiments

$$A_f^s = \left\{ \substack{(L,R)\leftarrow \mathrm{enc}(s);\\ \textit{Output } \mathrm{dec}(g(L),h(R))} \right\}$$

$$B_f^s = \left\{ \substack{\tilde{s}\leftarrow D_f;\\ \textit{If } \tilde{s}=* \textit{ output } s \textit{ else output } \tilde{s}} \right\}$$

have statistical distance at most ε.

B Deliver Us from Log Factors

A more thorough analysis of the sums in the proof of Theorem 7 allows us to get slightly better bounds. The technicalities are of little interest to the big picture and were hence omitted in the body of the paper. The addition consists of an alternative ending to the proof of Theorem 7.

▶ **Theorem 10.** *Let $G = (V, E)$ be d-regular with spectral expansion λ satisfying $n = \Omega(d^3 \log(d)/\lambda)$. Then $(\mathrm{enc}_G, \mathrm{dec}_G)$ is an $O\left(\frac{\lambda^{3/2}}{d}\right)$-non-malleable code in the split-state model.*

Proof. At the very end of the proof of Theorem 7, we arrived at

$$d(n-d)S_{k,l} \leq 2^{-\frac{l+k}{2}} \lambda n \cdot \sqrt{|G_1^k|} \cdot \sqrt{\frac{d \cdot |G_1^k| \cdot |H_1^l|}{n} + \lambda \cdot \sqrt{|G_1^k| \cdot |H_1^l|}}.$$

Applying Jensen's inequality, we get

$$S_{k,l} \leq O\left(\frac{\lambda}{\sqrt{dn}}\right) \cdot 2^{-\frac{l+k}{2}} \cdot |G_1^k| \cdot \sqrt{|H_1^l|} + O\left(\frac{\lambda^{3/2}}{d}\right) \cdot 2^{-\frac{l+k}{2}} \cdot \sqrt[4]{|G_1^k|^3 \cdot |H_1^l|} \tag{2}$$

with the functions hidden by the O-notation being independent of k, l.

Now, note that

$$\left|g^{-1}(G_1^k)\right| \geq \frac{n \cdot |G_1^k|}{2^k} \qquad\qquad \left|h^{-1}(H_1^l)\right| \geq \frac{n \cdot |H_1^l|}{2^l} \tag{3}$$

and for all $k \leq \lceil \log_2(d^2) \rceil$ we have $\frac{|G_1^k|}{2^{k/2}} \leq 2d$. We shall bound each of the terms of (2) separately.

First, write

$$L = \sum_{1 \leq k,l \leq \lceil \log_2(d^2) \rceil} \left(2^{-\frac{l+k}{2}} \cdot |G_1^k| \cdot \sqrt{|H_1^l|} \right).$$

Using the Cauchy-Schwartz inequality in the second inequality,

$$\begin{aligned}
L &\leq 2d \cdot \sum_{1 \leq l \leq \lceil \log_2(d^2) \rceil} \sqrt{2^{-l} \, |H_1^l|} \\
&\leq O\left(d \cdot \sqrt{\log(d)}\right) \cdot \sqrt{\sum_{1 \leq l \leq \lceil \log_2(d^2) \rceil} 2^{-l} \cdot |H_1^l|} \\
&\leq O\left(d \cdot \sqrt{\log(d)}\right) \cdot \sqrt{\sum_{1 \leq l \leq \lceil \log_2(d^2) \rceil} \frac{|h^{-1}(H_1^l)|}{n}} \\
&= O\left(d \cdot \sqrt{\log(d)}\right)
\end{aligned}$$

since the H_1^l are disjoint subsets of V. In conclusion,

$$O\left(\frac{\lambda}{\sqrt{dn}}\right) \cdot \sum_{1 \le k,l \le \lceil \log_2(d^2) \rceil} 2^{-\frac{l+k}{2}} \cdot \left|G_1^k\right| \cdot \sqrt{\left|H_1^l\right|} = O\left(\frac{\lambda \cdot \sqrt{d \cdot log(d)}}{\sqrt{n}}\right)$$
$$= O\left(\frac{\lambda^{3/2}}{d}\right).$$

Second, let $k \le l$ and write $t = l - k$. We now bound the sum using (3). Write

$$K = \sum_{1 \le k < l \le \lceil \log_2(d^2) \rceil} 2^{-\frac{l+k}{2}} \cdot \sqrt[4]{\left|G_1^k\right|^3 \cdot \left|H_1^l\right|}.$$

Then

$$\begin{aligned} K &\le \sum_{1 \le k < l \le \lceil \log_2(d^2) \rceil} \left(\frac{2^{\frac{k-l}{4}}}{n} \cdot \sqrt[4]{\left|g^{-1}(G_1^k)\right|^3 \cdot \left|h^{-1}(H_1^l)\right|}\right) \\ &\le \sum_{t=0}^{\lceil \log_2(d^2) \rceil} \left(\frac{2^{-\frac{t}{4}}}{n} \sum_{l=t}^{\lceil \log_2(d^2) \rceil} \sqrt[4]{\left|g^{-1}(G_1^{l-t})\right|^3 \cdot \left|h^{-1}(H_1^l)\right|}\right) \\ &\le \sum_{t=0}^{\lceil \log_2(d^2) \rceil} \left(\frac{2^{-\frac{t}{4}}}{n} \left(\sum_{l=t}^{\lceil \log_2(d^2) \rceil} \left|g^{-1}(G_1^{l-t})\right|\right)^{3/4} \cdot \left(\sum_{l=t}^{\lceil \log_2(d^2) \rceil} \left|h^{-1}(H_1^l)\right|\right)^{1/4}\right) \\ &\le \sum_{t=0}^{\lceil \log_2(d^2) \rceil} 2^{-\frac{t}{4}} = O(1), \end{aligned}$$

where the third inequality is established using Hölder's inequalty.

It now follows that

$$\sum_{1 \le k \le l \le \lceil \log_2(d^2) \rceil} S_{k,l} = O\left(\frac{\lambda^{3/2}}{d}\right).$$

By symmetry of k and l,

$$R_{1,1} = \sum_{1 \le k,l \le \lceil \log_2(d^2) \rceil} S_{k,l} = O\left(\frac{\lambda^{3/2}}{d}\right),$$

which completes the proof. ◀

C Instantiating Our Construction

Using our results to instantiate an efficient, secure split-state non-malleable code, we require a family of graphs $\{G_k\}_{k \in \mathbb{N}}$, where each $G_k = (V_k, E_k)$ is d_k-regular with spectral expansion λ_k, satisfying the following:

1. The function $\varepsilon(k) = \frac{\lambda_k^{3/2}}{d_k}$ is negligible.
2. We have $n_k = |V(G_k)| = \Omega(d_k^3 \log(d_k)/\lambda_k)$
3. Both sampling an edge $(u, v) \overset{u}{\leftarrow} E_k$ and sampling a non-edge $(u, v) \overset{u}{\leftarrow} (V_k \times V_k) \setminus E_k$ can be done in time polynomial in k.

4. Determining membership of a pair $(u, v) \in V \times V$ in $E(G_k)$ can be done deterministically in time polynomial in k.

Given such a family of graphs it is clear that the corresponding graph code $(\mathrm{enc}_{G_k}, \mathrm{dec}_{G_k})$ is an efficiently computable non-malleable code.

C.1 Instantiation with High-Degree Cayley Graphs

Explicit constructions of such families of graphs do indeed exist. We shall here give an example from [7] from the class of graphs known as Cayley graphs. The construction is as follows.

▶ **Definition 11.** *For p a prime and $1 \leq t < p$ let the graph $\mathrm{LD}_{p,t}$ have vertex set $\mathbb{F}_p^{t+1}$ and edge set*

$$E(\mathrm{LD}_{p,t}) = \left\{(x, x + (b, ab, a^2b, \ldots, a^tb)) \mid x \in \mathbb{F}_p^{t+1}, a, b \in \mathbb{F}_p\right\},$$

i.e. $x, y \in V(\mathrm{LD}_{p,T})$ are connected by an edge if and only if there exists $a, b \in \mathbb{F}_p$ such that $y = x + (b, ab, a^2b, \ldots, a^tb)$.

It is worth nothing that the graph $\mathrm{LD}_{p,t}$ is p^2-regular and that it is undirected as x is connected to y if and only if y is connected to x.

Now, let $t = 5$ and for each $k \in \mathbb{N}$ let p_k be some k-bit prime. We consider the family of graphs $\{\mathrm{LD}_{p_k,5}\}_{k\in\mathbb{N}}$ for our instantiation. In the following, we shall check the criteria from the beginning of the section point by point.

1. The family of graphs $\mathrm{LD}_{p,t}$ has great expander properties.
   ▶ **Theorem 12** (explicit in Trevisan [7])**.** *For $1 < t < p$, the graph $\mathrm{LD}_{p,t}$ is a pt-spectral expander.*
   This fact allows us to note that for our particular choice of graphs, $\varepsilon(k) = \frac{(p_k t)^{3/2}}{p_k^2} < \frac{12}{\sqrt{p_k}}$, which in fact is $2^{-\Omega(k)}$ and the representation size is $O(k)$ bits.
2. We have $\Omega\left(\frac{d_k^3 \log(d_k)}{\lambda_k}\right) = \Omega(p^5 \log(p))$ such that indeed,
   $$n_k = |V(\mathrm{LD}_{p_k,5})| = p^6 = \Omega\left(\frac{d_k^3 \log(d_k)}{\lambda_k}\right).$$
3. Sampling an edge $(u, v) \overset{u}{\leftarrow} E(\mathrm{LD}_{p_k,t})$ is simply a question of picking $x \in \mathbb{F}_{p_k}^{t+1}, a, b \in \mathbb{F}_{p_k}$ uniformly at random and then outputting the edge $(x, x + (b, ab, a^2b, \ldots, a^tb))$.
   To pick a non-edge, simply sample two random vertices $u, v \in \mathbb{F}_{p_k}^{t+1}$ uniformly at random and check (with the procedure to be specified below) whether $(u, v) \in E(\mathrm{LD}_{p_k,t})$. Since for $t > 1$ the probability of hitting an edge with such a random choice is $\leq 1/p_k$, the expected number of repetitions is constant and hence the procedure takes expected polynomial time.
4. To test membership of some $(u, v) \in \left(\mathbb{F}_{p_k}^{t+1}\right)^2$ in $E(\mathrm{LD}_{p_k,t})$, perform the following operation: Compute $x = u - v$ and write $x = (x_0, \ldots, x_t)$. It is now trival to check whether $\left(1, \frac{x_1}{x_0}, \ldots, \frac{x_t}{x_0}\right)$ is of the form $(1, a, a^2, \ldots, a^t)$.