Expander Graphs are Non-Malleable Codes

Peter M. R. Rasmussen* Amit Sahai†

September 28, 2018

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 in the split-state model.

1 Introduction

A split-state non-malleable code [DPW10] consists of randomized encoding and decoding algorithms (enc, 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 $\text{dec}(L, R) = m$. 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 $\text{dec}(g(L), h(R))$ is "unrelated" to the original message m. In particular, to be ε-non-malleable, it is enough [DKO13] to guarantee that when the message m is chosen uniformly at random and encoded into (L, R), the probability that $\text{dec}(g(L), h(R)) = 1 - m$ is at most $\frac{1}{2} + \varepsilon$. Since their introduction in 2010 [DPW10], split-state non-malleable codes have been the subject of intense study within theoretical computer science [DPW10, DKO13, ADL14, CZ14, CGL16, Li17].

Until our work, all known 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 [DKO13, ADL14, CZ14, CGL16, Li17].

In this work, we show that expander graphs immediately give rise to split-state non-malleable codes. Specifically, we show that any d-regular graph on n 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 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., $d = n^{\varepsilon}$), which can be constructed and analyzed easily, yielding $2^{-\Omega(k)}$-non-malleable codes (see, e.g., [Tre] or the appendix).

2 Preliminaries

We shall assume familiarity with the basics of codes and non-malleable codes. A cursory introduction to the most relevant definitions and intuition 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.*

*University of Copenhagen and Basic Algorithms Research Copenhagen. nmq584@alumni.ku.dk.
†UCLA. sahai@cs.ucla.edu

- *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 [DKO13].

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 $f \in \mathcal{F}$,*

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

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*

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

which are randomized and deterministic, respectively, and given by

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

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

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*

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

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

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

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

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

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,

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

□

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 save a few log-factors, have been placed in the appendix as Theorem 12.

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

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

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

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

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

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

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

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

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

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

□

References

[ADL14] Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In *Symposium on Theory of Computing, STOC*, 2014.

[CGL16] 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.

[CZ14] Eshan Chattopadhyay and David Zuckerman. Non-malleable codes against constant split-state tampering. In *Foundations of Computer Science, FOCS*, 2014.

[DKO13] Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Non-malleable codes from two-source extractors. In *CRYPTO*, 2013.

[DPW10] Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. In *ICS*, 2010.

[Li17] Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Symposium on Theory of Computing, STOC*, 2017.

[Tre] 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 Non-Malleable Codes

The following section will outline the definition and basic results regarding non-malleable codes. We shall start with an informal overview.

Consider the following very general scenario: A sender A wants to encode a value m, obtaining an encoding, $\text{enc}(m) = \tilde{m}$, and send it to a recipient B through a channel C such that B can then decode the received message to recover the original message, $m = \text{dec}(\tilde{m})$. Restricting the size of $\tilde{m}$ or letting C be a noisy channel that alters the message in some randomized way leads to the field of coding theory, whereas for instance restricting the amount of information that an adversary E with limited computational resources can glean from observing the traffic through C leads us to the field of cryptography.

A.1 Definition

Working with non-malleable codes, we ask the following information theoretic question. Consider some publicly known encoding and decoding functions (the encoding function may be randomized) and suppose that an adversary can pick a *tampering* function $f\colon \{0,1\}^k \to \{0,1\}^k$ from a family $\mathcal{F}$ of functions and apply that function to whatever message passes through C such that B in fact receives not $\tilde{m}$ but $f(\tilde{m})$. What restrictions must then apply to $\mathcal{F}$ to guarantee that no $f \in \mathcal{F}$ results in $\text{dec}(f(\tilde{m}))$ encoding a message with some relation to m that is not the identity? By this we mean that f with some non-negligible probability transforms the original message m into a related message, which is not just m itself but different and depending on m.

Since this explanation is rather vague, let us define the notion of a non-malleable code more formally.

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

$$P(\text{dec}(\text{enc}(s)) = s) = 1$$

where the probability is taken over the randomness of enc*.*

Note that dec can return $\perp$ meaning that the encoding it was given could not be decoded.

Definition 9 (Non-malleable code)**.** *A coding scheme* $(\mathrm{enc}, \mathrm{dec})$*,* $\mathrm{enc}\colon \mathcal{M} \to \mathcal{X}$ *and* $\mathrm{dec}\colon \mathcal{X} \to \mathcal{M} \cup \{\perp\}$*, is said to be* ε*-non-malleable with respect to a family of tampering functions* $\mathcal{F}$*, each* $f \in \mathcal{F}$ *being a function* $f\colon \mathcal{X} \to \mathcal{X}$*, if the following holds. For every* $f \in \mathcal{F}$ *there exists a distribution* D_f *supported on* $\mathcal{M} \cup \{*, \perp\}$ *such that for every* $s \in \mathcal{M}$ *the random variables defined by the experiments*

$$A_f^s = \left\{ \begin{smallmatrix} \tilde{s} \leftarrow \mathrm{enc}(s); \\ \mathit{Output}\ \mathrm{dec}(f(\tilde{s})) \end{smallmatrix} \right\}, \qquad B_f^s = \left\{ \begin{smallmatrix} \tilde{s} \leftarrow D_f; \\ \mathit{If}\ \tilde{s} = * \ \mathit{output}\ s\ \mathit{else}\ \mathit{output}\ \tilde{s} \end{smallmatrix} \right\}$$

satisfy $A_f^s \approx_\varepsilon B_f^s$*.*

The intuition behind this definition is that we would like a code to be non-malleable if the only possible "attacks" against it either copy the message or outputs something that is not related to the message. The distribution B in the definition satisfies just that. It can either sample some constant not depending on s or simply copy s.

A.2 The Necessity of Restricting $\mathcal{F}$

A natural initial question to ask, is whether we need any restrictions on $\mathcal{F}$ at all. We are quickly assured that this is indeed the case. If we let $\mathcal{F}$ be the set of all functions, it contains a function that decodes the message, changes it to a related message, and then encodes it again. It is thus intuitively clear that there is no defense against such tampering. The following proposition proves this formally.

Proposition 10. *Suppose that the coding scheme* $(\mathrm{enc}, \mathrm{dec})$ *is* ε*-non-malleable with respect to the family of all functions from* $\mathcal{X}$ *to* $\mathcal{X}$*,* $\mathcal{F} = \mathcal{X}^{\mathcal{X}}$*. Then* $\varepsilon \geq 1 - \frac{1}{|\mathcal{M}|}$*.*

Proof. If $|\mathcal{M}| = 1$ then the proposition is trivially true. So assume $|\mathcal{M}| > 1$, let $\varphi\colon \mathcal{M} \to \mathcal{M}$ be a permutation of $\mathcal{M}$ with no fixed points, and let $f \in \mathcal{F}$ be the function $f = \mathrm{enc} \circ \varphi \circ \mathrm{dec}$ where enc is computed using some fixed randomness. Further, let D_f be the distribution corresponding to ε and f in the definition of non-malleability. Then for every $m \in \mathcal{M}$, we have $\mathrm{dec}(f(\mathrm{enc}(m))) = \varphi(m) \neq m$. Thus, for every $m \in \mathcal{M}$ the fact that $A_f^m \approx_\varepsilon B_f^m$ implies $P(D_f = \varphi(m)) \geq 1 - \varepsilon$ since $P(A_f^m = \varphi(m)) = 1$ and

$$P(B_f^m = \varphi(m) \mid D_f \in (\mathcal{M} \setminus \{\varphi(m)\}) \cup \{*, \perp\}) = 0$$

as $m \neq \varphi(m)$. Now, since φ is a permutation of m, we have

$$1 \geq P(D_f \in \mathcal{M}) = \sum_{m \in \mathcal{M}} P(D_f = m) = \sum_{m \in \mathcal{M}} P(D_f = \varphi(m)) \geq |\mathcal{M}| \cdot (1 - \varepsilon),$$

which yields the desired inequality, $\varepsilon \geq 1 - \frac{1}{|\mathcal{M}|}$. □

A.3 Split State Model

Having established that the class $\mathcal{F}$ of allowed functions must be restricted, let us specify the particular restriction we shall work with. A very common setting to consider is that of the split state model. Let sets $\mathcal{L}$ and $\mathcal{R}$ be given. In the split state model, the class $\mathcal{F}$ consists of all functions $f\colon \mathcal{L} \times \mathcal{R} \to \mathcal{L} \times \mathcal{R}$ which can be written as $f = (g, h)$ for functions $g\colon \mathcal{L} \to \mathcal{L}$ and $h\colon \mathcal{R} \to \mathcal{R}$, i.e. such that $f(x) = (g(x), h(x))$. For clarity, we repeat the definition of non-malleability in the split state model.

Definition 11 (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 random variables defined by the experiments*

$$A_f^s = \left\{ \begin{smallmatrix} (L,R) \leftarrow \mathrm{enc}(s); \\ \mathit{Output}\ \mathrm{dec}(g(L), h(R)) \end{smallmatrix} \right\}$$

$$B_f^s = \left\{ \begin{smallmatrix} \tilde{s} \leftarrow D_f; \\ \mathit{If}\ \tilde{s} = * \ \mathit{output}\ s\ \mathit{else}\ \mathit{output}\ \tilde{s} \end{smallmatrix} \right\}$$

satisfy $A_f^s \approx_\varepsilon B_f^s$.

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 12. *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

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

Applying Jensen's inequality, we get

$$S_{k,l} \leq O\left(\frac{\lambda}{\sqrt{dn}}\right) \cdot 2^{-\frac{l+k}{2}} \cdot \left|G_1^k\right| \cdot \sqrt{\left|H_1^l\right|} + O\left(\frac{\lambda^{3/2}}{d}\right) \cdot 2^{-\frac{l+k}{2}} \cdot \sqrt[4]{\left|G_1^k\right|^3 \cdot \left|H_1^l\right|} \tag{1}$$

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

Now, note that

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

and for all $k \leq \lceil \log_2(d^2) \rceil$ we have $\frac{\left|G_1^k\right|}{2^{k/2}} \leq 2d$. We shall bound each of the terms of (1) separately.

First, using the Cauchy-Schwartz inequality in the second inequality,

$$\begin{aligned}
\sum_{1 \leq k,l \leq \lceil \log_2(d^2) \rceil} \left(2^{-\frac{l+k}{2}} \cdot \left|G_1^k\right| \cdot \sqrt{\left|H_1^l\right|}\right) &\leq 2d \cdot \sum_{1 \leq l \leq \lceil \log_2(d^2) \rceil} \sqrt{2^{-l} \left|H_1^l\right|} \\
&\leq O\left(d \cdot \sqrt{\log(d)}\right) \cdot \sqrt{\sum_{1 \leq l \leq \lceil \log_2(d^2) \rceil} 2^{-l} \cdot \left|H_1^l\right|} \\
&\leq O\left(d \cdot \sqrt{\log(d)}\right) \cdot \sqrt{\sum_{1 \leq l \leq \lceil \log_2(d^2) \rceil} \frac{\left|h^{-1}(H_1^l)\right|}{n}} \\
&= O\left(d \cdot \sqrt{\log(d)}\right)
\end{aligned}$$

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

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

Second, let $k \leq l$ and write $t = l - k$. We now bound the sum using (2).

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

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

It now follows that

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

By symmetry of k and l,

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

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 $(u, v) \xleftarrow{u} E_k$ and sampling a non-edge $(u, v) \xleftarrow{u} (V_k \times V_k) \setminus E_k$ 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 Cayley Graphs

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

Definition 13. *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 $\mathrm{LD}_{p,t}$ 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 14 (Trevisan [Tre])**.** *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 $(u, v) \xleftarrow{u} E(\mathrm{LD}_{p_k,t})$ 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 $(1, a, a^2, \ldots, a^t)$.

Acknowledgements

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.

Research 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 from 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.