Inception makes non-malleable codes shorter as well!

Divesh Aggarwal* Maciej Obremski†

April 16, 2019

Abstract

Non-malleable codes, introduced by Dziembowski, Pietrzak and Wichs in ICS 2010, have emerged in the last few years as a fundamental object at the intersection of cryptography and coding theory. Non-malleable codes provide a useful message integrity guarantee in situations where traditional error-correction (and even error-detection) is impossible; for example, when the attacker can completely overwrite the encoded message. Informally, a code is non-malleable if the message contained in a modified codeword is either the original message, or a completely "unrelated value". Although such codes do not exist if the family of "tampering functions" $\mathcal{F}$ allowed to modify the original codeword is completely unrestricted, they are known to exist for many broad tampering families $\mathcal{F}$.

The family which received the most attention is the family of tampering functions in the so called (2-part) *split-state* model: here the message x is encoded into two shares L and R, and the attacker is allowed to arbitrarily tamper with each L and R individually.

Dodis, Kazana, and the authors in STOC 2015 developed a generalization of non-malleable codes called the concept of non-malleable reduction, where a non-malleable code for a tampering family $\mathcal{F}$ can be seen as a non-malleable reduction from $\mathcal{F}$ to a family NM of functions comprising the identity function and constant functions. They also gave a constant-rate reduction from a split-state tampering family to a tampering family $\mathcal{G}$ containing so called 2-lookahead functions, and forgetful functions.

In this work, we give a constant rate non-malleable reduction from the family $\mathcal{G}$ to NM, thereby giving the first *constant rate non-malleable code in the split-state model.*

Central to our work is a technique called inception coding which was introduced by Aggarwal, Kazana and Obremski in TCC 2017, where a string that detects tampering on a part of the codeword is concatenated to the message that is being encoded.

*Department of Computer Science and Center for Quantum Technologies, National University of Singapore. Email: dcsdiva@nus.edu.sg.

†Center for Quantum Technologies, National University of Singapore. Email: obremski.math@gmail.com.

1 Introduction

Non-malleable codes, introduced by Dziembowski, Pietrzak and Wichs [DPW10], provide a useful message integrity guarantee in situations where traditional error-correction (and even error-detection) is impossible; for example, when the attacker can completely overwrite the encoded message. Informally, given a tampering family $\mathcal{F}$, a $\mathcal{F}$-non-malleable code (E, D) encodes a given message x into a codeword $y \leftarrow E(x)$ in a way that, if y is modified into $y' = f(y)$ by some $f \in \mathcal{F}$, then the message $x' = D(y')$ contained in the modified codeword y' is either the original message x, or a completely "unrelated value". In other words, non-malleable codes aim to handle a much larger class of tampering functions $\mathcal{F}$ than traditional error-correcting or error-detecting codes, at the expense of potentially allowing the attacker to replace a given message x by an unrelated message x' (and also necessarily allowing for a small "simulation error" ε). As shown by [DPW10], this relaxation still makes non-malleable codes quite useful in a variety of situations where (a) the tampering capabilities of the attacker might be too strong for error-detection, and, yet (b) changing x to unrelated x' is not useful for the attack. For example, imagine x being a secret key for a signature scheme. In this case, tampering which keeps x the same corresponds to the traditional chosen message attack (covered by the traditional definition of secure signatures), while tampering which changes x to an unrelated value x' will clearly not help in forging signatures under the original (un-tampered) verification key, as the attacker can produce such signatures under x' by himself.

Split-State Model. Although such codes do not exist if the family of "tampering functions" $\mathcal{F}$ is completely unrestricted [DPW10], they are known to exist for many broad tampering families $\mathcal{F}$. One such natural family is the family of tampering functions in the so called split-state model. Here the k-bit message x is encoded into 2 shares y_1, y_2 of length n each, and the attacker is allowed to *arbitrarily* tamper with each y_i *individually*. The rate of such an encoding is naturally defined as $\tau = \frac{k}{2n}$.

Non-malleable codes in this model could be interpreted as "non-malleable secret-sharing schemes": even if *all* the t message shares are independently tampered with, the recovered message is either x or is unrelated to x. Non-malleable codes in the split-state model have received a lot of attention so far so far [DPW10, LL12, DKO13, ADL14, CG14a, CG14b, Agg15, CGL16, Li17, Li18]. In addition, some of the recent results [GPR16, GK18a, GK18b, ADN+18, BS18, SV18] have shown application of non-malleable codes in the split-state model to other important problems like non-malleable commitments and non-malleable secret sharing.

The known results can be summarized as follows. The first non-malleable code in the split-state model against an information-theoretic adversary was constructed in [DKO13], who constructed a non-malleable code for 1-bit messages in the split-state model. Following that [ADL14, Agg15, AB16] gave the first information-theoretic construction supporting k-bit messages, but where the length of each share $n = O(k^5)$. There was a plausible conjecture stated in [ADL14] about the non-malleability of the inner product function under which one would get a 2-part split-state code with constant rate, i.e., $n = O(k)$.

In [CG14a], it was shown that the notion of non-malleable codes in the split-state model is closely related to the notion of non-malleable two-source extractors and using this insight, and the alternating extraction protocol from [DP07], recent results [CGL16, Li17, Li18] have obtained improved constructions of non-malleable codes in the split-state model. The most recent result [Li18] gives a construction with rate $\frac{c \cdot \log\log\log 1/\varepsilon}{\log\log 1/\varepsilon}$ for some constant c. This result has a constant rate if ε is a constant, but the rate approaches 0 if ε is negligible in n, as is required for cryptographic applications. In particular, if we choose $\varepsilon = 2^{-n^{\Omega(1)}}$, then the rate is $O(\frac{\log\log n}{\log n})$.

The authors, along with Dodis and Kazana [ADKO15a] introduced the concept of non-malleable reductions and, under a plausible conjecture, gave a series of reductions that results in constant rate non-malleable codes in the split-state model [1]

[1] A previous version of [ADKO15a] claimed a constant rate non-malleable codes in the split-state model. Unfortunately, Li [Li17] found a mistake in the proofs of one of the lemmas in the paper, and though the lemma is believable, currently the construction is secure only under a plausible conjecture.

However, until this work, the problem of unconditionally constructing constant rate non-malleable codes in the split-state model (with ε negligible in the size of the codeword) remains open.

In this paper we consider the strongest possible variant of non-malleable codes studied in the literature, super strong non-malleable codes. This variant ensures that every non-identity tampering will either be detected or the entire tampered codeword does not reveal any information about the secret message.

Our Result. In this work, we give a constant rate non-malleable code in the split-state model.

Theorem 1 (Main Result). *There exists an efficient, information-theoretically secure ε-non-malleable codes in the split-state model with shares of size $O(k)$, where k is the length of the message, and $\varepsilon = 2^{-k^{\Omega(1)}}$.*

Our result is achieved by giving a (super-strong) non-malleable code against the tampering family $\mathcal{G}$ containing 2-lookahead tampering functions and forgetful tampering functions. Combined with a non-malleable reduction from the 2-split tampering family to $\mathcal{G}$ gives a non-malleable code in the split-state model. For our construction, and a discussion of our proof techniques, we refer the reader to Section 4.

Other Related Work. If we relax the number of states to more than 2, or we restrict the adversary to be computationally bounded, then there are known efficient constructions of non-malleable codes. In particular, some recent results [CZ14, KOS17, KOS18, GMW18] obtain near optimal non-malleable codes in the t split-state model where t is a constant greater than 2, and [AAG$^+$16] gave a construction of a rate 1 non-malleable code against computationally bounded adversaries.

Other results that look at an (enhanced) split-state model are Faust et al. [FMNV14] which consider the model where the adversary can tamper continuously, and [ADKO15b], that considers the model where the adversary, in addition to performing split-state tampering, is also allowed some limited interaction between the two states.

There have been some results that have obtained non-malleable codes against continuous tampering in the split-state model [AKO17, ADN$^+$17]. In fact, some of our results rely on techniques developed in [AKO17].

In addition to the already-mentioned results, several recent works [CCFP11, CCP12, CKM11, FMVW14, AGM$^+$14, AGM$^+$15, BDSKM16, FHMV17, BDSKM18, BDSG$^+$18] either used or built non-malleable codes for various families $\mathcal{F}$, but did not concentrate on the split-state model, which is our focus here.

The notion of non-malleability was introduced by Dolev, Dwork and Naor [DDN00], and has found many applications in cryptography. Traditionally, non-malleability is defined in the computational setting, but recently non-malleability has been successfully defined and applied in the information-theoretic setting (generally resulting in somewhat simpler and cleaner definitions than their computational counter-parts). For example, in addition to non-malleable codes studied in this work, the work of Dodis and Wichs [DW09] defined the notion of non-malleable extractors as a tool for building round-efficient privacy amplification protocols.

Finally, the study of non-malleable codes falls into a much larger cryptographic framework of providing counter-measures against various classes of tampering attacks. This work was pioneered by the early works of [ISW03, GLM$^+$03, IPSW06], and has since led to many subsequent models. We do not list all such tampering models, but we refer to [KKS11, LL12] for an excellent discussion of various such models.

2 Preliminaries

For a set T, let U_T denote a uniform distribution over T, and, for an integer ℓ, let U_ℓ denote uniform distribution over ℓ bit strings. For any random variable A and any set $\mathcal{A}$, we denote $A|_{A\in\mathcal{A}}$ to be the random variable A' such that

$$\forall a,\ \Pr[A' = a] = \Pr[A = a \mid A \in \mathcal{A}] \ .$$

The *statistical distance* between two random variables A, B is defined by

$$\Delta(A\ ;\ B) = \frac{1}{2}\sum_v \left|\Pr[A=v] - \Pr[B=v]\right| \ .$$

We use $A \approx_\varepsilon B$ as shorthand for $\Delta(A,B) \le \varepsilon$.

Lemma 2. *For any function α, if $\Delta(A\ ;\ B) \le \varepsilon$, then $\Delta(\alpha(A)\ ;\ \alpha(B)) \le \varepsilon$.*

The following is a simple result from [ADL14].

Lemma 3. *Let $X_1, Y_1 \in \mathcal{A}_1$, and $Y_1, Y_2 \in \mathcal{A}_2$ be random variables such that $\Delta((X_1,X_2)\ ;\ (Y_1,Y_2)) \le \varepsilon$. Then, for any non-empty set $\mathcal{A}' \subseteq \mathcal{A}_1$, we have*

$$\Delta(X_2 \mid X_1 \in \mathcal{A}'\ ;\ Y_2 \mid Y_1 \in \mathcal{A}') \le \frac{2\varepsilon}{\Pr(X_1 \in \mathcal{A}')} \ .$$

The following is a slight variant of a similar simple lemma from [ADL14]. The proof is just a simple application of triangle inequality.

Lemma 4. *Let S be some random variable distributed over a set $\mathcal{S}$, and let $\mathcal{S}_1, \ldots, \mathcal{S}_j$ be a partition of $\mathcal{S}$. Let $\phi : \mathcal{S} \to \mathcal{T}$ be some function, and let $D_1, \ldots, D_j$ be some random variables over the set $\mathcal{T}$. Assume that for all $1 \le i \le j$,*

$$\Delta\left(\phi(S)|_{S\in\mathcal{S}_i}\ ;\ D_i\right) \le \varepsilon_i.$$

Then

$$\Delta\left(\phi(S)\ ;\ D\right) \le \sum \varepsilon_i \Pr[S \in S_i] \ ,$$

for some random variable $D \in \mathcal{T}$ such that for all d $\Pr[D=d] = \sum_i \Pr[S \in \mathcal{S}_i] \cdot \Pr[D_i = d]$.

The *min-entropy* of a random variable W is $\mathbf{H}_\infty(W) \stackrel{\text{def}}{=} -\log(\max_w \Pr[W=w])$, and the *conditional* min-entropy of W given Z is $\mathbf{H}_\infty(W|Z) \stackrel{\text{def}}{=} -\log\left(\mathbb{E}_{z\leftarrow Z} \max_w \Pr[W=w|Z=z]\right)$.

Definition 5. *We say that a function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ is an (n,k,m,ε)-*2*-source extractor if for all independent sources $X, Y \in \{0,1\}^n$ such that min-entropy $\mathbf{H}_\infty(X) + \mathbf{H}_\infty(Y) \ge k$, we have $(Y, \mathsf{Ext}(X,Y)) \approx_\varepsilon (Y, U_m)$, and $(X, \mathsf{Ext}(X,Y)) \approx_\varepsilon (X, U_m)$.*

For n being an integer multiple of m, and interpreting elements of $\{0,1\}^m$ as elements from $\mathbb{F}_{2^m}$ and those in $\{0,1\}^n$ to be from $(\mathbb{F}_{2^m})^{n/m}$, we have that the inner product function is a good 2-source extractor.

Lemma 6. *For all positive integers m, n such that n is a multiple of m, and for all $\varepsilon > 0$, the inner product function from two n-bit strings to an m-bit string is an efficient $\left(n, n+m+2\log\left(\frac{1}{\varepsilon}\right), m, \varepsilon\right)$ 2-source extractor.*

Definition 7. *A function $C : \{0,1\}^s \times \{0,1\}^n \to \{0,1\}^t$ is called an* ε-almost universal hash function *if for any $x, y \in \{0,1\}^n$ such that $x \ne y$,*

$$\Pr_{R\leftarrow\{0,1\}^s}(C(R,x) = C(R,y)) \le \varepsilon$$

The following is a standard construction of a polynomial evaluation ε-universal hash function. The parameters are from [DW09].

Lemma 8. *For any $n, t > 2\log n$, there exists an efficiently computable $2^{-t/2}$-almost univeral hash function $C : \{0,1\}^s \times \{0,1\}^n \to \{0,1\}^t$ with $s = 2t$.*

3 Non-malleable Codes and Reductions

DEFINITIONS. In [ADKO15a], the notion of non-malleable codes w.r.t. to a tampering family $\mathcal{F}$ [DPW10] was generalized to a more versatile notion of *non-malleable reductions* from $\mathcal{F}$ to $\mathcal{G}$. The following definitions are taken from [ADKO15a].

Definition 9 (non-malleable reduction). Let $\mathcal{F} \subset A^A$ and $\mathcal{G} \subset B^B$ be some classes of functions (which we call *manipulation* functions). We will write:

$$(\mathcal{F} \Rightarrow \mathcal{G}, \varepsilon)$$

and say $\mathcal{F}$ *reduces to* $\mathcal{G}$, if there exist an efficient randomized *encoding* function $E : B \to A$, and an efficient deterministic *decoding* function $D : A \to B$, such that (a) for all $x \in B$, we have $D(E(x)) = x$, and (b) for all $f \in \mathcal{F}$, there exists G such that for all $x \in B$,

$$\Delta\Big(D(f(E(x)))\ ;\ G(x)\Big) \leq \varepsilon, \tag{1}$$

where G is a *distribution* over $\mathcal{G}$, and $G(x)$ denotes the distribution $g(x)$, where $g \leftarrow G$.

The pair (E, D) is called $(\mathcal{F}, \mathcal{G}, \varepsilon)$-*non-malleable reduction*.

Intuitively, $(\mathcal{F}, \mathcal{G}, \varepsilon)$-non-malleable reduction allows one to encode a value x by $y \leftarrow E(x)$, so that tampering with y by $y' = f(y)$ for $f \in \mathcal{F}$ gets "reduced" (by the decoding function $D(y') = x'$) to tampering *with x itself* via some (distribution over) $g \in \mathcal{G}$.

In particular, the notion of *non-malleable code* w.r.t. $\mathcal{F}$, is simply a reduction from $\mathcal{F}$ to the family of "trivial manipulation functions" NM_k defined below.

Definition 10. Let NM_k denote the set of *trivial manipulation functions* on k-bit strings, which consists of the identity function $I(x) = x$ and all constant functions $f_c(x) = c$, where $c \in \{0,1\}^k$.

We say that a pair (E, D) defines an $(\mathcal{F}, k, \varepsilon)$-*non-malleable code*, if it defines a $(\mathcal{F}, \mathsf{NM}_k, \varepsilon)$-non-malleable reduction.

The utility of non-malleable reductions comes from the following natural composition theorem that was shown in [ADKO15a], which allows to gradually make our tampering families simpler and simpler, until we eventually end up with a non-malleable code (corresponding to the trivial family NM_k).

Theorem 11 (Composition). *If $(\mathcal{F} \Rightarrow \mathcal{G}, \varepsilon_1)$ and $(\mathcal{G} \Rightarrow \mathcal{H}, \varepsilon_2)$, then $(\mathcal{F} \Rightarrow \mathcal{H}, \varepsilon_1 + \varepsilon_2)$.*

We will also need the following trivial observation.

Observation 1 (Union). *Let (E, D) be an $(\mathcal{F}, \mathcal{H}, \varepsilon)$ and a $(\mathcal{G}, \mathcal{H}, \varepsilon')$ non-malleable reduction (resp. transformation). Then (E, D) is an $(\mathcal{F} \cup \mathcal{G}, \mathcal{H}, \max(\varepsilon, \varepsilon'))$ non-malleable reduction (resp. transformation).*

USEFUL TAMPERING FAMILIES. We define several natural tampering families we will use in this work. For this, we first introduce the following "direct product" operator on tampering families:

Definition 12. Given tampering families $\mathcal{F} \subset A^A$ and $\mathcal{G} \subset B^B$, let $\mathcal{F} \times \mathcal{G}$ denote the class of functions h from $(A \times B)^{A \times B}$ such that

$$h(x) = h_1(x_1) \| h_2(x_2)$$

for some $h_1 \in \mathcal{F}$ and $h_2 \in \mathcal{G}$ and $x = x_1 \| x_2$, where $x_1 \in A, x_2 \in B$.

We also let $\mathcal{F}^1 := \mathcal{F}$, and, for $t \geq 1$, $\mathcal{F}^{t+1} := \mathcal{F}^t \times \mathcal{F}$.

We can now define the following tampering families:

- $\mathcal{S}_n = (\{0,1\}^n)^{\{0,1\}^n}$ denote the class of *all* manipulation functions on n-bit strings.
- Given $t > 1$, $\mathcal{S}_n^t$ denotes the tampering family in the *t-split-state model*, where the attacker can apply t arbitrarily correlated functions $h_1, \ldots, h_t$ to t separate, n-bit parts of memory (but, of course, each h_i can only be applied to the i-th part individually).
- $\mathcal{FOR}^t_{n_1,n_2,\ldots,n_t}$ denotes *forgetful* family. It is applied to t parts of memory of length n_i but the output value can depend only on $(t-1)$ parts. More precisely: Let $x \in \{0,1\}^n$ be a bit vector and $x_i \in \{0,1\}^{n_i}$ denote i-th block of n bits. For any $h \in \mathcal{FOR}^t_{n_1,n_2,\ldots,n_t}$ there exist a subset $S \subset \{1,2,\ldots,t\}$ of size $(t-1)$ such that $h(x)$ can be evaluated from x_S. Besides that, it is not restricted in any way.
- Finally, $\mathcal{LA}^{\leftarrow t}_{n_1,\ldots,n_t}$, where $n = n_1 + \cdots + n_t$ denotes the class of *lookahead manipulation functions* l that can be rewritten as $l = (l_1, \ldots, l_t)$, for $l_i : \{0,1\}^{n_1+\cdots+n_i} \to \{0,1\}^{n_i}$, and where

  $$l(x) = l_1(x_1) \| \ldots \| l_t(x_1, \ldots, x_t)$$

  for $x_i \in \{0,1\}^{n_i}$. In other words, if $l(x_1, \ldots, x_t) = y_1, \ldots, y_t$, then y_1 depends on x_1, and y_2 depends on both x_1 and x_2, and in general, y_i depends on $x_1, \ldots, x_i$.

Super Strong Non-malleable Codes. The following is a definition of a stronger variant of non-malleable codes.

Definition 13. (Super Strong Non-Malleable Code.) *We say that an encoding scheme* $(\mathsf{Enc} : \mathcal{M} \to \mathcal{X}, \mathsf{Dec} : \mathcal{X} \to \mathcal{M} \cup \{\bot\})$ *is* ε-super strong non-malleable against the tampering family $\mathcal{F} \subseteq \mathcal{X}^{\mathcal{X}}$ *if for every functions $f \in \mathcal{F}$ and for every $m_0, m_1 \in \mathcal{M}$*

$$\mathsf{SupStrTamp}^{f,g}_{m_0} \approx_\varepsilon \mathsf{SupStrTamp}^{f,g}_{m_1}$$

where

$$\mathsf{SupStrTamp}^{f,g}_m = \left\{ \begin{array}{c} X \leftarrow \mathsf{Enc}(m), \\ \textit{output } \textbf{same} \textit{ if } X = f(X) \\ \textit{else if } \mathsf{Dec}(f(X)) = \bot \textit{ output } \bot \\ \textit{else output: } f(X) \end{array} \right\}$$

We will need an efficient construction of super strong non-malleable codes in the split-state model from [AKO17, ADL14, Agg15]. Recall that in the proof of [AKO17], and [ADL14], the ambient space is partitioned and it is shown that there are some partitions that are small, and for every other partition, it is shown that the tampered codeword either decodes to the same message, or to $\bot$, or the tampered codeword reveals no information about the original message and this implies super strong non-malleability.

Theorem 14. *There exists an efficient construction* $(\mathsf{Enc}, \mathsf{Dec})$ *of ε-super strong non-malleable codes in the split-state model from $\{0,1\}^{7t}$to[2] $\{0,1\}^n \times \{0,1\}^n$ with $\varepsilon = 2^{-n^{\Omega(1)}}$, and $n = O(t^5)$. Moreover, for any tampering functions f, g, the following hold*

1. $\mathsf{Dec}(x, y) = h(\mathsf{Ext}(x, y))$*, where* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \mathbb{Z}_p$ *is a $n + O(t) + 2\log 1/\varepsilon$ strong two-source extractor, where $p = 2^{O(t)}$ is a prime. With probability at least $1 - \varepsilon$, $h(U_{\mathbb{Z}_p}) = \bot$, and for every message $m \in \{0,1\}^{7t}$,* $\Pr[h(U_{\mathbb{Z}_p} = m \mid h(U_{\mathbb{Z}_p}) \neq \bot] = \frac{1}{2^{7t}}$.

[2]The constant 7 in this Theorem statement are chosen to match those required in our results. There is some freedom in the choice of parameters in [**?**], and so the result of this theorem follows for an appropriate choice of t.

2. Let $\mathcal{L}, \mathcal{R} \subset \{0,1\}^n$, such that for all $\ell \in \mathcal{L}$, $f(\ell) \neq \ell$,

$$|f^{-1}(f(\ell)) \cap \mathcal{L}| \times |g^{-1}(g(r)) \cap \mathcal{R}| \leq 2^{0.9n}$$

and $|\mathcal{L} \times \mathcal{R}| \geq 2^{2n-t}$. Then, for any $m \in \{0,1\}^{7t}$, $\Pr[\mathsf{Dec}(f(L), g(R)) \neq \perp \mid \mathsf{Dec}(L,R) = m] = O(\varepsilon)$.

Note that the "moreover" part of the statement above is not stated explicitly in [AKO17, ADL14], but is immediate from the construction and proof. In particular, the statement (1) follows from the fact that the first step of the decoding algorithm computes the inner product modulo a prime p, and an appropriately chosen affine-evasive set S modulo p (with $|S| \leq \varepsilon \cdot p$) is partitioned into 2^{7t} sets $S_1, \ldots, S_{2^{7t}}$, and the decoding algorithm outputs m if the inner product is in S_m and $\perp$, otherwise. The statement (2) follows from the fact that the proof proceeds by partitioning the ambient space $\{0,1\}^n \times \{0,1\}^n$ into sets of the form $\mathcal{L} \times \mathcal{R}$, and we show for each of these sets that either the codeword remains unchanged after tampering (which is not possible since we assume $f(\ell) \neq \ell$), or the tampered codeword decodes to $\perp$, or the tampered codeword is independent of the message. In the last case, the tampered codeword decodes to $\perp$ by (1), and the fact that Ext is a strong 2-source extractor.

4 Our constructions and the main result

It was shown in [ADKO15a] that

Theorem 15. *For any q, there is an $n = O(q)$ such that*

$$(S_n^2 \Rightarrow \mathcal{LA}_{q,q,q}^{\leftarrow 3} \times \mathcal{LA}_{q,q,q}^{\leftarrow 3} \cup \mathcal{FOR}_{q,q,q,q,q,q}^{6}, 2^{-\Omega(q)}) \,.$$

So, now we construct (super-strong) non-malleable codes for the tampering family $\mathcal{LA}_{q,q,q}^{\leftarrow 3} \times \mathcal{LA}_{q,q,q}^{\leftarrow 3} \cup \mathcal{FOR}_{q,q,q,q,q,q}^{6}$. In Section 4.1, we give a super-strong non-malleable code against 2-lookahead tampering family, and in Section 4.2, we show how to extend it to include the forgetful tampering family.

4.1 A super-strong non-malleable code against 2-lookahead tampering

Theorem 16. *There exists a $2^{-k^{\Omega(1)}}$-super strong non-malleable code for k-bit messages against the tampering family $\mathcal{LA}_{100k,25k,5k}^{\leftarrow 3} \times \mathcal{LA}_{100k,25k,5k}^{\leftarrow 3}$.*

Construction. Our construction (E, D) depicted in Figure 4.1 that achieves the above result is as follows.

Encoding : Given $m \in \{0,1\}^k$, we do the following.

- Let Ext_3 be the inner product function from $\mathbb{F}_{2^k}^5 \times \mathbb{F}_{2^k}^5 \to \mathbb{F}_{2^k}$. Let A, B be chosen uniformly at random from $\{0,1\}^{5k}$.
- Let Ext_2 be the inner product function from $\mathbb{F}_{2^k}^{25} \times \mathbb{F}_{2^k}^{25} \to \mathbb{F}_{2^k}$. Sample $X, Y \in \{0,1\}^{25k}$ uniformly at random, conditioned on $z := \mathsf{Ext}_2(X,Y) = m \oplus \mathsf{Ext}_3(A,B)$.
- Let σ_1, σ_2 be $2t$-bit strings sampled uniformly at random for an appropriately chosen $t = \Theta(k^{1/5})$.
- Let $C : \{0,1\}^{2t} \times \{0,1\}^{60k} \to \{0,1\}^t$ be a $2^{-t/2}$-almost universal hash function as defined in Lemma 8. Also, let $z = z_1 \| z_2$ where $|z_1| = 2t$.
- Let $s = \sigma_1, \sigma_2, c_1 := C(\sigma_1, X \| Y \| A \| B), c_2 := z_1 \oplus \sigma_2$.
- Let $L, R := \mathsf{Enc}(s)$, where $(\mathsf{Enc}, \mathsf{Dec})$ be a super strong non-malleable code in the split state model given by Theorem 14 from $\{0,1\}^{7t}$ to $\{0,1\}^n \times \{0,1\}^n$ where $n = 100k$.

- Output (L, X, A) as the first part of the codeword, and (R, Y, B) as the second part.

Decoding : Given $(L, X, A), (R, Y, B)$ we do the following.

- Compute $s = \mathsf{Dec}(L, R)$, and $z = Ext_2(X, Y)$.
- If $s = \perp$, output $\perp$, else let $s = \sigma_1, \sigma_2, c_1, c_2$.
- If $z_1 \neq c_2 \oplus \sigma_2$, where z_1 is the first $2t$ bits of z, or $c_1 \neq C(\sigma_1, X\|Y\|A\|B)$, output $\perp$.
- Else output $z \oplus Ext_3(A, B)$.

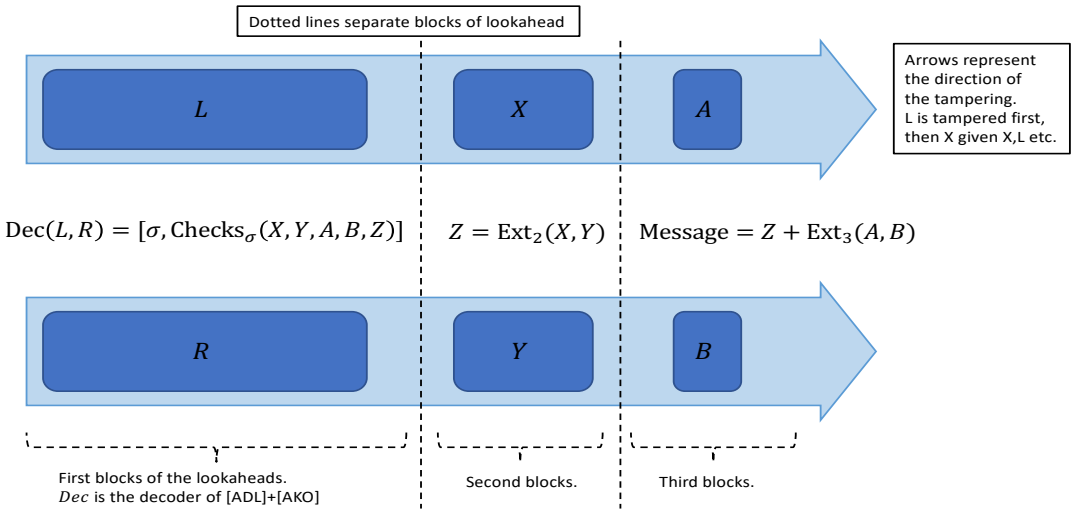

Figure 1: The decoding algorithm D.

Intuition behind the construction. Before giving an overview of the proof, we look at a few tampering scenarios to give the intuition behind the construction.

Scenario 1: Adversary leaves L, R unchanged.
Then we can retrieve the original checks for X, Y, Z, A, B, and we know that adversary won't be able to come up with choice of X', Y', Z', A', B' such that the checks remain fulfilled. There is a technicality: adversary tampers with X, A after seeing L and with Y, B after seeing R, but we choose the lengths of the elements appropriately so we can model everything as a small leakage from L and R and the secrecy of checks is preserved i.e. $X, X', Y, Y', Z, Z', A, A', B, B'$ together do not reveal any information about the random seeds σ_1, σ_2, and thus chance that checks of original and tampered parts of the codeword will collide is negligible.

Scenario 2: Adversary tampers with L, R.
In this case by super-strong nmc properties we get that after the decoding both random seeds σ_1', σ_2' and corresponding checks values c_1', c_2' are independent of original values, but we can not exclude that adversary knows both σ_1', σ_2' and c_1', c_2' (e.g., if he completely overwrote L, R by something unrelated). Now we have two sub-scenarios:

Scenario 2.1**:** Adversary lost some information about X or Y.
What we mean is that X can not be fully recovered from L', X' or Y from R', Y' then by Ext_2 extractor properties adversary has lost all information about Z and, as a consequence, lost all information about the secret message.

Scenario 2.2**:** Adversary preserved information about X and Y.
We know that σ'_1, σ'_2 and c'_1, c'_2 are controlled by adversary but completely independent of original checks. We also know that X' and Y' have to have high min-entropy else they wouldn't carry information about X, Y. We would like to say that adversary can not produce X' that has the same check as X but there are few issues with this reasoning. First of all the new seeds σ'_1, σ'_2 are not random but controlled by adversary. Although the adversary commits to those checks before seeing and tampering with X and Y, we can only guarantee that X' has high min-entropy not that it is uniform, and thus the adversary can pick the distribution of X' such that it always fulfils fixed checks. This is where the check on Z comes into play. We can argue that X', Y' are high-entropic even given L', R' thus $Z' = \mathsf{Ext}_2(X', Y')$ is close to uniform. Notice that the check for Z (or Z') has the following property, for any fixing of σ'_2 and c'_2, the probability that for U uniform $U + \sigma'_2 = c'_2$ is negligible. Since, as we just discussed, Z' is close to uniform and independent of σ'_2 and c'_2 the probability that $Z' + \sigma'_2 = c'_2$ is negligible[3].

Scenario 2.2′**:** Imagine we are in scenario 2.2 but we do not have parts A, B i.e. message is simply Z instead of $Z + \mathsf{Ext}_3(A, B)$.
Notice that in the previous scenario we couldn't guarantee independence of Z and Z' we only knew that Z' is close to uniform. Now, however, if message is simply Z we do have a problem. Remember that in a tampering experiment adversary picks two messages m_0, m_1 and has to distinguish which one of them was encoded. If $Z = m_b$ and Z' is not independent of Z (indeed it is possible that $Z' = Z$) then even the check on Z will not save us. Adversary could overwrite L, R with L', R' encoding checks σ'_1, σ'_2, and $c'_1, c'_2 = m_0 + \sigma'_2$ and tamper X, Y in such a way that X', Y' fulfil the first check while preserving $\mathsf{Ext}_2(X', Y') = \mathsf{Ext}_2(X, Y)$ (such tampering can not be excluded without showing very strong non-malleable properties of inner-product). Now the output of tampering experiment will be *same* if encoded message was m_0 and $\perp$ if encoded message was m_1. To summarise A, B are used only to make Z' independent of the message and uniform even given checks, so that we can detect this tampering.

Proof overview. Given a message $m \in \{0,1\}^k$, let $E(z) = (L, X, A), (R, Y, B)$. Let $f_1, g_1 : \{0,1\}^{100k} \to \{0,1\}^{100k}$, $f_2, g_2 : \{0,1\}^{125k} \to \{0,1\}^{25k}$, and $f_3, g_3 : \{0,1\}^{130k} \to \{0,1\}^{5k}$ be arbitrarily chosen functions, and let

$$L' = f_1(L),\ R' = g_1(R),\ X' = f_2(L, X),\ Y' = g_2(R, Y),\ A' = f_3(L, X, A),\ B' = g_3(R, Y, B)\ .$$

Also, let $z', z'_1, z'_2, \sigma'_1, \sigma'_2, c'_1, c'_2$ be the corresponding tampered values.

As is the case with almost all proofs for non-malleable code constructions, our proof proceeds by first partitioning the ambient space $\{0,1\}^{130k} \times \{0,1\}^{130k}$ depending on the functions $f_1, g_1, f_2, g_2, f_3, g_3$. We then argue that for each partition, as long as the partition is large enough, conditioned on the random variables L, X, A, R, Y, B being restricted to be in that partition, we can show that either the codeword remains unchanged after tampering, or $D((L', X', A'), (R', Y', B')) = \perp$ with high probability, or the tampered codeword is almost independent of the message m, (i.e., it reveals no information about the message m).

We first consider the partition where $L', R' = L, R$. In this case, notice that if X, Y, A, B are changed then with high probability, $C(\sigma_1, X \| Y \| A \| B) \neq C(\sigma_1, X \| Y \| A \| B)$, and so the decoding algorithm outputs $\perp$ with high probability. On the other hand, if X, Y, A, B are unchanged, then the decoder outputs same.

[3]To be precise, we are not checking if $Z' = c'_2 - \sigma'_2$ but only if some short prefix of Z' is equal to $c'_2 - \sigma'_2$, which is still unlikely given that Z' and thus also its prefix are uniform.

For the formal proof, we need to deal with the dependence between various random variables, and the detailed proof can be found in Lemma 19.

We next consider the partition where $\mathbf{H}_\infty(L') + \mathbf{H}_\infty(R') \gg n$, and $L', R' \neq L, R$. In this case, by Theorem 14, we have that $\mathsf{Dec}(L', R') = \perp$ with high probability.

This leaves us with the partitions where one of $\mathbf{H}_\infty(L|L')$ or $\mathbf{H}_\infty(R|R')$ (say $\mathbf{H}_\infty(L|L')$) is at least $0.45n$. Notice that here we are using the fact that for an appropriate choice of partitions, we have that $\mathbf{H}_\infty(L') + \mathbf{H}_\infty(L|L') \approx n$ for L chosen uniformly from that partition. This in particular means that $\mathbf{H}_\infty(L|L', X', A') \geq 45k - 25k - 5k > 0.15n$. Thus, again using the observation that $\mathsf{Dec}(L, R)$ is a deterministic function of a strong two-source extractor $h(\mathsf{Ext}(L, R))$, we have that $\mathsf{Dec}(L, R)$ is independent of $L', R', X', Y', A', B', X, Y, A, B$. At this point, we can fix L, R, thereby fixing $L' = \ell', R' = r'$.

Thus, X', Y' are deterministic functions of X, Y, respectively. Now we further partition the space $\{0,1\}^{25k} \times \{0,1\}^{25k}$ based on the functions f_2, g_2. First we consider the case where $\mathbf{H}_\infty(X') + \mathbf{H}_\infty(Y') \gg 26k$. In this case, by using the fact that inner product is a strong 2-source extractor, and noting that X, Y, and hence X', Y' is independent of the message m, we have that z' (and hence z'_1 is close to uniform and independent of the message m, and ℓ', r'. Thus, the probability that $\sigma'_2 = c'_2 \oplus z'_1$ is negligible, and hence the decoding algorithm outputs $\perp$ with high probability.

The only remaining case is when one of $\mathbf{H}_\infty(X|X')$ or $\mathbf{H}_\infty(Y|Y')$ (say $\mathbf{H}_\infty(X|X')$) is at least $10k$, in which case $\mathbf{H}_\infty(X|X', A') \geq 5k$, and hence by the strong extractor property of the inner product, we have that z is independent of X', Y', A', B' and hence is independent of the tampered codeword (since we already fixed L', R'). The tampered codeword is thus independent of the message.

4.2 A super strong non-malleable code secure against 2-lookahead and forgetful tampering

Theorem 17. *There is an $2^{-k^{\Omega(1)}}$-super-strong non-malleable code for $k - O(k^{1/5})$-bit messages against the tampering family $\mathcal{LA}^{\leftarrow 3}_{125k,25k,5k} \times \mathcal{LA}^{\leftarrow 3}_{125k,25k,5k} \cup \mathcal{FOR}^{6}_{125k,25k,5k,125k,25k,5k}$.*

Construction. Our construction (E^*, D^*) depicted in Figure 4.2 that achieves the above result is as

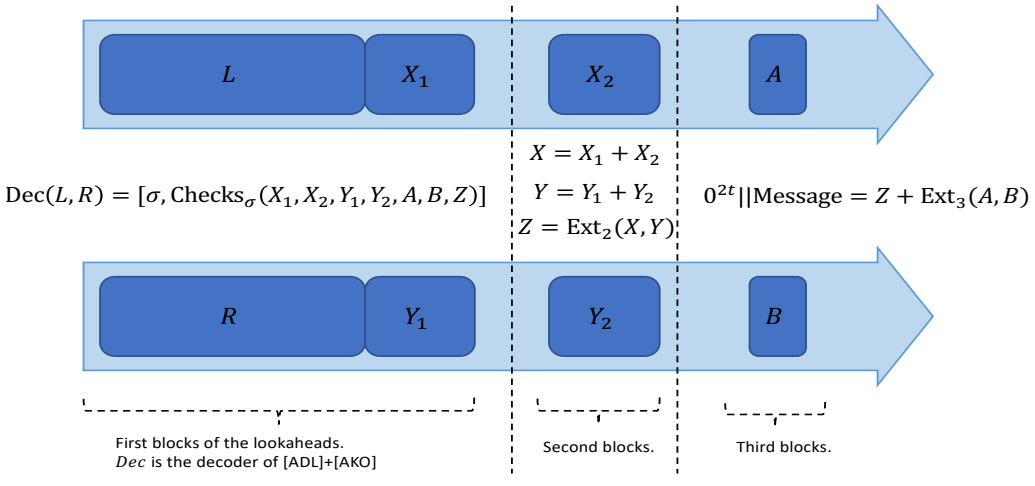

Figure 2: The decoding algorithm D^*.

Encoding : Given a message $m^* \in \{0,1\}^{k-2t}$, let $m = 0^{2t}\|m^*$. Let $X, A, Y, B, \sigma_1, \sigma_2, z, z_1, z_2, c_2$ be as in the encoding of $E(m)$, where E is the encoding algorithm from Section 4.1. Choose X_1, Y_1 uniformly at random from $\{0,1\}^{25k}$, and let $X_2 = X \oplus X_1$, $Y_2 = Y \oplus Y_1$. Let $C : \{0,1\}^{2t} \times \{0,1\}^{110k} \to \{0,1\}^t$ be a $2^{-t/2}$-almost universal hash function as defined in Lemma 8. Let $s = \sigma_1, \sigma_2, c_1 := C(\sigma_1, X_1\|X_2\|Y_1\|Y_2\|A\|B), c_2$, and let $\mathsf{Enc}(s) = L, R$. Output the three parts of the first lookahead as $((L, X_1), X_2, A$, and the three parts of the second lookahead as $(R, Y_1), Y_2, B$.

Decoding : Given $((L, X_1), X_2, A), ((R, Y_1), Y_2, B)$, compute $\mathsf{Dec}(L, R) = s$, and $\mathsf{Ext}_2(X_1 \oplus X_2, Y_1 \oplus Y_2) = s$. Output $\perp$ if $s = \perp$, else let $s = \sigma_1, \sigma_2, c_1, c_2$. If $c_1 \neq C(\sigma_1, X_1\|X_2\|Y_1\|Y_2\|A\|B)$ or $c_2 \oplus z_1 \neq c_2$, output $\perp$, else output $z \oplus \mathsf{Ext}_3(A, B)$.

We now give a simple argument that shows that this construction is secure against the tampering family $\mathcal{LA}^{\leftarrow 3}_{125k,25k,5k} \times \mathcal{LA}^{\leftarrow 3}_{125k,25k,5k} \cup \mathcal{FOR}^{6}_{125k,25k,5k,125k,25k,5k}$ if the construction given in Theorem 16 is secure against the tampering family $\mathcal{LA}^{\leftarrow 3}_{100k,25k,5k} \times \mathcal{LA}^{\leftarrow 3}_{100k,25k,5k}$.

We first argue security against lookahead tampering. Let the tampering functions be $f_1, g_1 : \{0,1\}^{125k} \to \{0,1\}^{100k}$, $f_2, g_2 : \{0,1\}^{125k} \to \{0,1\}^{25k}$, $f_3, g_3 : \{0,1\}^{150k} \to \{0,1\}^{25k}$, $f_4, g_4 : \{0,1\}^{155k} \to \{0,1\}^{5k}$, such that

$$L'_1 = f_1(L, X_1),\ X'_1 = f_2(L, X_1),\ X'_2 = f_3(L, X_1, X_2),\ A' = f_4(L, X_1, X_2, A)\ ,$$

and

$$R'_1 = g_1(R, Y_1),\ Y'_1 = g_2(R, Y_1),\ Y'_2 = g_3(R, Y_1, Y_2),\ B' = g_4(R, Y_1, Y_2, B)\ ,$$

We condition on $X_1 = x, Y_1 = y$, and then we define the functions f^*_1, f^*_2, f^*_3 as

$$f^*_1(L) = f_1(L, x),\ f^*_2(L, X) = f_2(L, x) \oplus f_3(L, x, X \oplus x),\ f^*_3(L, X, A) = f_4(L, x, X \oplus x, A)\ ,$$

and similarly define g^*_1, g^*_2, g^*_3, which is an attack in $\mathcal{LA}^{\leftarrow 3}_{100k,25k,k} \times \mathcal{LA}^{\leftarrow 3}_{100k,25k,5k}$ against the construction from Theorem 16. With this change, the proof is almost identical to that of Theorem 16. The only slight difference from the proof of Theorem 16 is that here the tampering experiment does not output same if $L' = L, R' = R, X' = X, Y' = Y, A' = A, B' = B$, but $X'_1 \neq X_1$ or $Y'_1 \neq Y_1$. However, in this case, the decoder outputs $\perp$ with high probability.

The non-malleability against the forgetful family is immediate from the fact that $\mathsf{Ext}_2, \mathsf{Ext}_3$ are strong 2-source extractors and losing one of A, B, X_1, X_2, Y_1, Y_2 loses information about the message m. The only subtlety here is that if the adversary loses information about, say X_2, the adversary still knows z_1 given L, R, but since $m = 0^{2t}\|m^*$, learning z_1 does not reveal any information about m^*.

The detailed proof is in Section 6.

4.3 Final result via a non-malleable reduction from [ADKO15a]

Setting $q = 125k$ in Theorem 18, and padding the required number of 0's as a prefix to each part of the codeword, we obtain the following

Theorem 18. *There is an $2^{-q^{\Omega(1)}}$-super-strong non-malleable code for $k - O(k^{1/5})$-bit messages against the tampering family $\mathcal{LA}^{\leftarrow 3}_{q,q,q} \times \mathcal{LA}^{\leftarrow 3}_{q,q,q} \cup \mathcal{FOR}^{6}_{q,q,q,q,q,q}$.*

Theorem 1 then follows from Theorem 11 and Theorem 15.

5 Proof of Theorem 16

We now prove Theorem 16. Given a message $z \in \{0,1\}^k$, let $E(z) = (L, X, A), (R, Y, B)$. Let $f_1, g_1 : \{0,1\}^{100k} \to \{0,1\}^{100k}$, $f_2, g_2 : \{0,1\}^{125k} \to \{0,1\}^{25k}$, $f_3, g_3 : \{0,1\}^{125k} \to \{0,1\}^{5k}$ be arbitrarily chosen functions, and let

$$L' = f_1(L),\ R' = g_1(R),\ X' = f_2(L, X),\ Y' = g_2(R, Y),\ A' = f_3(L, X, A),\ B' = g_3(R, Y, B)\ .$$

Also, let $z', z'_1, \sigma'_1, \sigma'_2, c'_1, c'_2$ be the corresponding tampered values.

5.1 f_1 and g_1 are identity functions

The following lemma considers the case when $f_1(L) = L$, and $g_1(R) = R$.

Lemma 19. *Let $\mathcal{L}$ be the set of all $\ell \in \{0,1\}^n$ such that $f_1(\ell) = \ell$, let $\mathcal{R}$ be the set of all $r \in \{0,1\}^n$ such that $g_1(r) = r$. Then, if $\mathcal{L} \times \mathcal{R} \geq 2^{2n-t}$, then*

$$\mathsf{SupStrTamp}^{f,g}_{m_0}|_{L_0 \in \mathcal{L}, R_0 \in \mathcal{R}} \approx_{2^{-t/3}} \mathsf{SupStrTamp}^{f,g}_{m_1}|_{L_0 \in \mathcal{L}, R_0 \in \mathcal{R}} ,$$

where $E(m_b) = L_b, X_b, A, R_b, Y_b, B$ for $b = 0, 1$, in the first step of the corresponding SupStrTamp *experiment.*

Proof. For $b \in \{0,1\}$, let $L'_b = f_1(L_b), R'_b = g_1(R_b), X'_b = f_2(L_b, X_b), Y'_b = g_2(R_b, Y_b), A'_b = f_3(L_b, X_b, A), B'_b = g_3(R_b, Y_b, B)$ be the corresponding tampered codeword of $\mathsf{Enc}(m_b)$. Let σ_1, σ_2 be sampled uniformly and independently of everything else from $\{0,1\}^{2t}$. Also, let $\widetilde{L}, \widetilde{R}$ be sampled uniformly from $\mathcal{L}, \mathcal{R}$ respectively. Then,

$$\mathbf{H}_\infty(\widetilde{L}|f_2(\widetilde{L}, X_b), f_3(\widetilde{L}, X_b, A)) + \mathbf{H}_\infty(\widetilde{R}) \geq 2n - t - 30k .$$

Thus,

$$\Delta\left(Ext(\widetilde{L}, \widetilde{R}) \; ; \; U_{\mathbb{Z}_p} \mid X_b, Y_b, A, B, f_2(\widetilde{L}, X_b), g_2(\widetilde{R}, Y_b), f_3(\widetilde{L}, X_b, A), g_3(\widetilde{R}, Y_b, B)\sigma_1, \sigma_2\right) \leq 2^{-34k} .$$

Conditioning on $D(\mathsf{Ext}(\widetilde{L}, \widetilde{R})) = \sigma_1, \sigma_2, C(\sigma_1, X_b \| Y_b \| A \| B), \sigma_2 \oplus z_1$, where z_1 is the first $2t$ bits of $\mathsf{Ext}_2(X_b, Y_b)$ (respectively, $D(U_{\mathbb{Z}_p}) = \sigma_1, \sigma_2, C(\sigma_1, X_b \| Y_b \| A \| B), \sigma_2 \oplus z_1$) and using Lemma 3, we have that

$$\begin{aligned} &X_b, Y_b, A, B, f_2(L_b, X_b), g_2(R_b, Y_b), f_3(L_b, X_b, A), g_3(R_b, Y_b, B) \qquad (2) \\ \approx_{2^{-33k}} \; &X_b, Y_b, A, B, f_2(\widetilde{L}, X_b), g_2(\widetilde{R}, Y_b), f_3(\widetilde{L}, X_b, A), g_3(\widetilde{R}, Y_b, B) . \end{aligned}$$

Now we assume that we fix $A = \alpha$, and $B = \beta$. Let $\phi(\ell, x)$ be a binary function such that $\phi(\ell, x) = 1$ if $f_2(\ell, x) = x$ and $f_3(\ell, x, \alpha) = \alpha$, and 0, otherwise. Similarly, let $\psi(r, y)$ be a binary function such that $\psi(r, y) = 1$ if $g_2(r, y) = y$ and $g_3(r, y, \beta) = \beta$, and 0, otherwise.

Notice that by the inequality 3, by introducing an additional statistical distance of 2^{-33k}, we may assume that the probability that the corresponding SupStrTamp experiment outputs same with probability

$$\Pr[f_2(\widetilde{L}, X_b) = X_b \; \wedge g_2(\widetilde{R}, Y_b) = Y_b \; \wedge \; f_3(\widetilde{L}, X_b, \alpha) = \alpha \; \wedge \; g_3(\widetilde{L}, Y_b, \beta) = \beta] ,$$

and by the almost universality of C, we have that the corresponding SupStrTamp experiment outputs $\perp$ with probability at least

$$\Pr[f_2(\widetilde{L}, X_b) \neq X_b \; \vee \; g_2(\widetilde{R}, Y_b) \neq Y_b \; \vee \; f_3(\widetilde{L}, X_b, \alpha) \neq \alpha \; \vee \; g_3(\widetilde{L}, Y_b, \beta) \neq \beta] - 2^{-t/2} .$$

Thus, upto a statistical distance of at most $2^{-33k} + 2^{-t/2}$, the output of the SupStrTamp experiment is determined by the functions $\phi(\widetilde{L}, X_b)$, $\psi(\widetilde{R}, Y_b)$.

Now, let $\widetilde{X}, \widetilde{Y}$ be uniform in $\{0,1\}^{25k}$ independent of everything else. Then, by the strong 2-source extractor property of the inner product, we have that

$$\mathsf{Ext}_2(\widetilde{X}, \widetilde{Y}), \phi(\widetilde{L}, \widetilde{X}), \psi(\widetilde{R}, \widetilde{Y}) \approx_{2^{-11k}} U_k, \phi(\widetilde{L}, \widetilde{X}), \psi(\widetilde{R}, \widetilde{Y}) .$$

Conditioning $\mathsf{Ext}_2(\widetilde{X}, \widetilde{Y}) = m_b$ (respectively, $U_k = m_b$) and applying Lemma 3, we get that

$$\phi(\widetilde{L}, X_0), \psi(\widetilde{R}, Y_0) \approx_{2^{-10k}} \phi(\widetilde{L}, X_1), \psi(\widetilde{R}, Y_1) .$$

Since $2^{-t/2} + 2^{-33k} + 2^{-10k} < 2^{-t/3}$, we get the desired result. □

5.2 f_1 is far from being bijective

Lemma 20. *Let $\mathcal{L}$ be the set of all $\ell \in \{0,1\}^n$ such that $f_1(\ell) \neq \mathcal{L}$, and $|f_1^{-1} f_1(\ell) \cap \mathcal{L}| \geq 2^{45k}$, and $\mathcal{R}$ be a subset of $\{0,1\}^n$ such that $\mathcal{L} \times \mathcal{R} \geq 2^{2n-t}$. Then*

$$\mathsf{SupStrTamp}^{f,g}_{m_0}|_{L_0 \in \mathcal{L}, R_0 \in \mathcal{R}} \approx_{2^{-t}} \mathsf{SupStrTamp}^{f,g}_{m_1}|_{L_0 \in \mathcal{L}, R_0 \in \mathcal{R}} ,$$

where $E(m_b) = L_b, X_b, A, R_b, Y_b, B$ for $b = 0, 1$, in the first step of the corresponding SupStrTamp *experiment.*

Proof. Let $\widetilde{L}, \widetilde{R}, \widetilde{X}, \widetilde{Y}$ be sampled uniformly from $\mathcal{L}, \mathcal{R}, \{0,1\}^{25k}, \{0,1\}^{25k}$ respectively. Let σ_1, σ_2 be sampled uniformly and independently of everything else from $\{0,1\}^{2t}$. Also, let $M = Ext_3(A, B) \oplus Ext_2(\widetilde{X}, \widetilde{Y})$. Let

$$\widetilde{L}' = f_1(\widetilde{L}), \widetilde{R}' = g_1(\widetilde{R}), \widetilde{X}' = f_2(\widetilde{L}, \widetilde{X}), \widetilde{Y}' = g_2(\widetilde{R}, \widetilde{Y}), A' = f_3(\widetilde{L}, \widetilde{X}, A), B' = g_3(\widetilde{R}, \widetilde{Y}, B) .$$

Now, let $\mathsf{Dec}(\widetilde{L}', \widetilde{R}') = \sigma_1' \| \sigma_2' \| c_1' \| c_2'$ if $\mathsf{Dec}(\widetilde{L}', \widetilde{R}') \neq \perp$. Also, let $\widetilde{z}' = \mathsf{Ext}_2(\widetilde{X}', \widetilde{Y}')$, and let $\widetilde{z}_1'$ be the first $2t$ bits of $\widetilde{z}'$. For any message m, T_m to be a random variable that depends on $\widetilde{L}', \widetilde{R}', \widetilde{X}', \widetilde{Y}', A', B'$ conditioned on $\mathsf{Ext}_2(\widetilde{X}, \widetilde{Y}) \oplus A \oplus B = m$ and is $\perp$ if $\mathsf{Dec}(\widetilde{L}', \widetilde{R}') = \perp$, or one of $c_1' = C(\sigma_1', \widetilde{X}' \| \widetilde{Y}' \| A' \| B')$, or $c_2' = \widetilde{z}_1' \oplus \sigma_2'$ does not hold. Otherwise, $T_m = \widetilde{L}', \widetilde{R}', \widetilde{X}', \widetilde{Y}', A', B'$.

Since

$$\mathbf{H}_\infty(\widetilde{L} | \widetilde{L}', \widetilde{X}', A') + \mathbf{H}_\infty(\widetilde{R}) \geq 45k - 25k - 5k + n - t = 15k + n - t .$$

Thus,

$$\Delta\left(Ext(\widetilde{L}, \widetilde{R}) \; ; \; U_{\mathbb{Z}_p} \mid \widetilde{X}, \widetilde{Y}, A, B, \widetilde{L}', \widetilde{R}', \widetilde{X}', \widetilde{Y}', A', B', \sigma_1, \sigma_2\right) \leq 2^{-7k} .$$

Conditioning on $h(\mathsf{Ext}(\widetilde{L}, \widetilde{R})) = \sigma_1, \sigma_2, C(\sigma_1, \widetilde{X} \| \widetilde{Y} \| A \| B), \sigma_2 \oplus z_1$, where z_1 is the first $2t$ bits of $\mathsf{Ext}_2(\widetilde{X}, \widetilde{Y})$ (respectively, $h(U_{\mathbb{Z}_p}) = \sigma_1, \sigma_2, C(\sigma_1, \widetilde{X} \| \widetilde{Y} \| A \| B), \sigma_2 \oplus z_1$) and $M = m$, and using Lemma 3, we have that, for any message m_0,

$$\Delta(\mathsf{SupStrTamp}^{f,g}_{m_0}|_{L_0 \in \mathcal{L}, R_0 \in \mathcal{R}} \; ; \; T_{m_0}) \leq 2^{-6k} . \tag{3}$$

Thus, it is sufficient to show that $\Delta := \Delta(T_{m_0} \; ; \; T_{m_1})$ is small. For this, we bound Δ for every choice of $\widetilde{L} = \ell$ and $\widetilde{R} = r$. We denote this as $\Delta_{\ell,r}$. Since $\widetilde{L}' = f_1(\widetilde{L}), \widetilde{R}' = g_1(\widetilde{R})$ is a deterministic function of $\widetilde{L}, \widetilde{R}$, this fixes $\mathsf{Dec}(\widetilde{L}', \widetilde{R}') = \mathsf{Dec}(f_1(\ell), g_1(r))$. If $\mathsf{Dec}(\widetilde{L}', \widetilde{R}') = \perp$, then we have that

$$\Delta_{\ell,r} = 0 .$$

We now consider the case when $\mathsf{Dec}(\widetilde{L}', \widetilde{R}') \neq \perp$. Using a slight abuse of notation, we let $f_2(\widetilde{X}) = f_2(\ell, \widetilde{X})$, $g_2(\widetilde{Y}) = g_2(r, \widetilde{Y})$. We partition $\{0,1\}^{25k} \times \{0,1\}^{25k}$ into $\mathcal{X}_0 \times \mathcal{Y}_0$, $\mathcal{X}_0 \times \mathcal{Y}_1$, $\mathcal{X}_1 \times \mathcal{Y}_0$, and $\mathcal{X}_1 \times \mathcal{Y}_1$, where

$$\mathcal{X}_0 = \{x \in \{0,1\}^{25k} \mid |f_2^{-1}(f_2(x))| \geq 2^{10k}\} ,$$

$$\mathcal{Y}_0 = \{y \in \{0,1\}^{25k} \mid |g_2^{-1}(g_2(y))| \geq 2^{10k}\} ,$$

$\mathcal{X}_1 = \{0,1\}^{20k} \setminus \mathcal{X}_0$, and $\mathcal{Y}_1 = \{0,1\}^{20k} \setminus \mathcal{Y}_0$. Let us introduce two claims needed to finish the proof.

Claim 21. *If $|\mathcal{X}_0 \times \mathcal{Y}_0| \geq 2^{49k}$, then*

$$\Delta\left(T_{m_0}|_{\widetilde{X} \in \mathcal{X}_0, \widetilde{Y} \in \mathcal{Y}_0} \; ; \; T_{m_1}|_{\widetilde{X} \in \mathcal{X}_0, \widetilde{Y} \in \mathcal{Y}_0}\right) \leq 2^{-0.5k+1} .$$

Proof. Let X^*, Y^* be uniform in $\mathcal{X}_0, \mathcal{Y}_0$, respectively. Then, $\mathbf{H}_\infty(X^* | f_2(\ell, X^*), f_3(\ell, X^*, A)) \geq 5k$, and also $\mathbf{H}_\infty(Y^*) \geq 24k$. Thus, by the strong 2-source extractor property of the inner product,

$$\Delta(\mathsf{Ext}_2(X^*, Y^*) \; ; \; U_k \mid A, B, f_2(\ell, X^*), g_2(r, Y^*), f_3(\ell, X^*, A), g_3(r, Y^*, B)) \leq 2^{-1.5k} .$$

Conditioning on $Ext_2(X^*, Y^*) = m_b \oplus \mathsf{Ext}_3(A, B)$ (respectively $U_k = m_b$), and using Lemma 3, and noting that T_{m_b} is a deterministic function of $f_2(\ell, X^*), g_2(r, Y^*), f_3(\ell, X^*, A), g_3(r, Y^*, B)$ conditioned on $Ext_2(X^*, Y^*) = m_b \oplus \mathsf{Ext}_3(A, B)$, we obtain the desired result. □

Similarly, since we only used that one of f_2, g_2 has a large preimage, we have that if $|\mathcal{X}_0 \times \mathcal{Y}_1| \geq 2^{49k}$, then

$$\Delta\left(T_{m_0}|_{\widetilde{X}\in\mathcal{X}_0,\widetilde{Y}\in\mathcal{Y}_1}\ ;\ T_{m_1}|_{\widetilde{X}\in\mathcal{X}_0,\widetilde{Y}\in\mathcal{Y}_1}\right) \leq 2^{-0.5k+1}\ ,$$

and If $|\mathcal{X}_1 \times \mathcal{Y}_0| \geq 2^{49k}$, then

$$\Delta\left(T_{m_0}|_{\widetilde{X}\in\mathcal{X}_1,\widetilde{Y}\in\mathcal{Y}_0}\ ;\ T_{m_1}|_{\widetilde{X}\in\mathcal{X}_1,\widetilde{Y}\in\mathcal{Y}_0}\right) \leq 2^{-0.5k+1}\ .$$

We now show a similar result for $\mathcal{X}_1 \times \mathcal{Y}_1$.

Claim 22. *If $|\mathcal{X}_1 \times \mathcal{Y}_1| \geq 2^{49k}$, then*

$$\Delta\left(T_{m_0}|_{\widetilde{X}\in\mathcal{X}_1,\widetilde{Y}\in\mathcal{Y}_1}\ ;\ T_{m_1}|_{\widetilde{X}\in\mathcal{X}_1,\widetilde{Y}\in\mathcal{Y}_1}\right) \leq 2^{-2t+2}\ .$$

Proof. Let X^*, Y^* be uniform in $\mathcal{X}_1, \mathcal{Y}_1$, respectively. In this case, $\mathbf{H}_\infty(f_2(X^*)) + \mathbf{H}_\infty(g_2(Y^*)) \geq 49k - 10k - 10k = 29k$. Thus,

$$\Delta(\mathsf{Ext}_2(f_2(X^*), g_2(Y^*))\ ;\ U_k) \leq 2^{-1.5k}\ .$$

Let z_1^* be the first $2t$ bits of $\mathsf{Ext}_2(f_2(X^*), g_2(Y^*))$. Notice that (X^*, Y^*) and $\mathsf{Ext}_3(A, B) \oplus \mathsf{Ext}_2(X^*, Y^*)$ are independently distributed. Thus, independent of the message, the probability that $z_1^* \oplus \sigma_2' = c_2'$ is at most $\frac{1}{2^{2t}} + \frac{1}{2^{2k}}$. This implies that the probability that $T_{m_b} \neq \perp$ is at most $\frac{1}{2^{2t}} + \frac{1}{2^{1.5k}}$. The result follows. □

Thus, for any $i, j \in \{0,1\}$, we have that

$$\Pr[\widetilde{X} \in \mathcal{X}_i, \widetilde{Y} \in \mathcal{Y}_j] \cdot \Delta\left(T_{m_0}|_{\widetilde{X}\in\mathcal{X}_i,\widetilde{Y}\in\mathcal{Y}_j}\ ;\ T_{m_1}|_{\widetilde{X}\in\mathcal{X}_i,\widetilde{Y}\in\mathcal{Y}_j}\right) \leq 2^{-2t+2}\ ,$$

and by Lemma 4, this implies that the statistical distance is at most $4 \cdot 2^{-2t+2}) \leq 2^{-2t+4}$. Combining with inequality 3, we get the desired result. □

5.3 Partitioning the space and finishing the proof

We now prove Theorem 16. For this, we partition $\{0,1\}^n \times \{0,1\}^n$ depending on the functions f_1, g_1. Define $\mathcal{L}_{id}$ to be the set of all $\ell \in \{0,1\}^n$ such that $\mathcal{L}_{bij}$ to be the set of all ℓ in $\mathcal{L}' = \{0,1\}^n \setminus \mathcal{L}_{id}$ such that $|f_1^{-1}(f_1(\ell)) \cap \mathcal{L}_{id}|$ is at most $2^{0.45n}$ (i.e., the function has few preimages, and is close to being a bijective function), and $\mathcal{L}_{ffb}$ to be the remaining set of all ℓ in $\mathcal{L}' = \{0,1\}^n \setminus \mathcal{L}_{id}$ such that $|f_1^{-1}(f_1(\ell)) \cap \mathcal{L}_{id}|$ is greater than $2^{0.45n}$ (i.e., the function has many preimages, and is far being a bijective function). We similarly define the partitions $\mathcal{R}_{id}, \mathcal{R}_{bij}, \mathcal{R}_{ffb}$ based on the function g_1.

Together, the above partitions define 9 partitions for the space $\{0,1\}^n \times \{0,1\}^n$. By Lemma **??**, we have that if $\mathcal{L}_{id} \times \mathcal{R}_{id}$ is at least 2^{2n-t} then the desired statistical distance for L, R restricted to be in $\mathcal{L}_{id} \times \mathcal{R}_{id}$ is at most $2^{-t/3}$. Also, by Theorem 14, if $\mathcal{L} \times \mathcal{R}$ (where $\mathcal{L} \times \mathcal{R}$ is one of $\mathcal{L}_{id} \times \mathcal{R}_{bij}$, $\mathcal{L}_{bij} \times \mathcal{R}_{id}$, or $\mathcal{L}_{bij} \times \mathcal{R}_{bij}$) is at least 2^{2n-t} then the desired statistical distance for L, R restricted to be in $\mathcal{L}_{id} \times \mathcal{R}_{id}$ is at most $2^{-k^{\Omega(1)}}$. Finally, by Lemma 20, we have that if $\mathcal{L} \times \mathcal{R}$ (where one or both of $\mathcal{L}$ and $\mathcal{R}$ are $\mathcal{L}_{ffb}$ and $\mathcal{R}_{ffb}$, respectively) is at least 2^{2n-t} then the desired statistical distance for L, R restricted to be in $\mathcal{L}_{id} \times \mathcal{R}_{id}$ is at most $2^{-k^{\Omega(1)}}$.

The result then follows from Lemma 4.

6 Proof of Theorem 18

We now give a simple argument that shows that this construction is secure against the tampering family $\mathcal{LA}^{\leftarrow 3}_{125k,25k,5k} \times \mathcal{LA}^{\leftarrow 3}_{125k,25k,5k} \cup \mathcal{FOR}^{6}_{125k,25k,5k,125k,25k,5k}$ if the construction given in Theorem 16 is secure against the tampering family $\mathcal{LA}^{\leftarrow 3}_{125k,25k,5k} \times \mathcal{LA}^{\leftarrow 3}_{125k,25k,5k}$.

We first argue security against lookahead tampering. The proof for this is essentially identical to that of Theorem 16, as explained below. Let the tampering functions be $f_1, g_1 : \{0,1\}^{125k} \to \{0,1\}^{100k}$, $f_2, g_2 : \{0,1\}^{125k} \to \{0,1\}^{25k}$, $f_3, g_3 : \{0,1\}^{150k} \to \{0,1\}^{25k}$, $f_4, g_4 : \{0,1\}^{155k} \to \{0,1\}^{5k}$, such that

$$L'_1 = f_1(L, X_1),\ X'_1 = f_2(L, X_1),\ X'_2 = f_3(L, X_1, X_2),\ A' = f_4(L, X_1, X_2, A)\ ,$$

and

$$R'_1 = g_1(R, Y_1),\ Y'_1 = g_2(R, Y_1),\ Y'_2 = g_3(R, Y_1, Y_2),\ B' = g_4(R, Y_1, Y_2, B)\ ,$$

We condition on $X_1 = x_1, Y_1 = y_1$, and we will argue that the statistical distance between the $\mathsf{SupStrTamp}$ for the message m_0 and that for the message m_1 conditioned on the fixing of X_1, Y_1 is small. We define the functions $f^* = (f^*_1, f^*_2, f^*_3)$ as

$$f^*_1(L) = f_1(L, x_1),\ f^*_2(L, X) = f_2(L, x_1) \oplus f_3(L, x_1, X \oplus x_1),\ f^*_3(L, X, A) = f_4(L, x_1, X \oplus x_1, A)\ ,$$

and similarly, we define $g^* = (g^*_1, g^*_2, g^*_3)$. Notice that in the proof of Theorem 16, we partitioned the space and for each space, we showed that the super strong tampering experiment either outputs same, or decodes to $\perp$, or the tampered codeword is independent of the original message. Notice that if the super-strong tampering experiment for the coding scheme of Theorem 16 for the tampering functions f^*, g^* outputs $\perp$, or is independent of the message, then so is the case for the corresponding tampering experiment in the current construction for the functions $f = (f_1, f_2, f_3, f_4)$, $g = (g_1, g_2, g_3, g_4)$. Thus, the only case where the tampering experiment differs is the case when $f^*_1(L) = g^*_1(R)$, in which case from the proof of Lemma 19, with high probability, the output is same or $\perp$, depending on the boolean random variables $\phi(L, X)$, and $\phi(R, Y)$, where $\phi(\ell, x) = 1$ if $f_2(\ell, x) = x$ and $f_3(\ell, x, \alpha) = \alpha$, and 0, otherwise and $\psi(r, y) = 1$ if $g_2(r, y) = y$ and $g_3(r, y, \beta) = \beta$, and 0, otherwise.

In order for the proof to go through, we need to change the definition to $\phi(\ell, x) = 1$, if $f^*_2(\ell, x) = x$, $f^*_3(\ell, x, \alpha) = \alpha$, and $f_2(\ell, x_1) = x_1$, and similarly redefine $\psi(r, y)$. The rest of the proof remains the same.

In order to argue security against forgetful tampering, consider the case where the adversary loses information about one of A or B (say A), but knows $L, R, X_1, X_2, Y_1, Y_2, B$. We assume that A, B, X_1, X_2, Y_1, Y_2 are uniformly distributed and L, R is computed as in the E^* given A, B, X_1, X_2, Y_1, Y_2. In this case, since $\mathbf{H}_\infty(A | C(\sigma_2, X_1 \| X_2 \| Y_1 \| Y_2 \| A \| B)) \geq 5k - t$, and thus we have that

$$\Delta(\mathsf{Ext}_3(A, B)\ ;\ U_k \mid B, X_1, X_2, Y_1, Y_2, L, R) \leq 2^{-1.5k}\ .$$

For any message m^*, we have that $\mathsf{Ext}_3(A, B) \oplus \mathsf{Ext}_2(X_1, X_2) = m^*$ (respectively $U_k \oplus \mathsf{Ext}_2(X_1, X_2) = m$), and using Lemma 3, we have that upto statistical distance $2^{-0.5k}$, $B, X_1, X_2, Y_1, Y_2, L, R$ are independent of the message m.

Similarly, if the adversary loses information about one of X_2 or Y_2 (say X_2), then a similar argument shows that z_2 is uniform and independent of $A, B, X_1, Y_1, Y_2, L, R$, and hence conditioning on $(z_1 \| z_2) \oplus \mathsf{Ext}_3(A, B) = 0^{2t} \| m^*$, which implies that upto statistical distance $2^{-\Omega(k)}$, m^* is independent of $A, B, X_1, Y_1, Y_2, L, R$.

Losing one of (L, X_1) or (R, Y_1) (say (L, X_1)) is clearly worse for the adversary, and so the adversary cannot distinguish between the tampered codeword of any two messages. The result follows. □

References

[AAG+16] Divesh Aggarwal, Shashank Agrawal, Divya Gupta, Hemanta K Maji, Omkant Pandey, and Manoj Prabhakaran. Optimal computational split-state non-malleable codes. In *Theory of Cryptography Conference*, pages 393–417. Springer, 2016.

[AB16] Divesh Aggarwal and Jop Briët. Revisiting the sanders-bogolyubov-ruzsa theorem in f p n and its application to non-malleable codes. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 1322–1326. Ieee, 2016.

[ADKO15a] Divesh Aggarwal, Yevgeniy Dodis, Tomasz Kazana, and Maciej Obremski. Non-malleable reductions and applications. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 459–468, 2015.

[ADKO15b] Divesh Aggarwal, Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Leakage-resilient non-malleable codes, 2015.

[ADL14] Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In *STOC*. ACM, 2014.

[ADN+17] Divesh Aggarwal, Nico Dottling, Jesper Buus Nielsen, Maciej Obremski, and Erick Purwanto. Continuous non-malleable codes in the 8-split-state model. Technical report, Cryptology ePrint Archive, Report 2017/357, 2017.

[ADN+18] Divesh Aggarwal, Ivan Damgård, Jesper Buus Nielsen, Maciej Obremski, Erick Purwanto, Joao Ribeiro, and Mark Simkin. Stronger leakage-resilient and non-malleable secret-sharing schemes for general access structures. *IACR Cryptology ePrint Archive*, 2018:1147, 2018.

[Agg15] Divesh Aggarwal. Affine-evasive sets modulo a prime. *Information Processing Letters*, 115(2):382–385, 2015.

[AGM+14] Shashank Agrawal, Divya Gupta, Hemanta K. Maji, Omkant Pandey, and Manoj Prabhakaran. Explicit non-malleable codes resistant to permutations and perturbations. *IACR Cryptology ePrint Archive*, 2014:841, 2014.

[AGM+15] Shashank Agrawal, Divya Gupta, Hemanta K. Maji, Omkant Pandey, and Manoj Prabhakaran. A rate-optimizing compiler for non-malleable codes against bit-wise tampering and permutations. In *Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part I*, pages 375–397, 2015.

[AKO17] Divesh Aggarwal, Tomasz Kazana, and Maciej Obremski. Inception makes non-malleable codes stronger. In *Theory of Cryptography Conference*, pages 319–343. Springer, 2017.

[BDSG+18] Marshall Ball, Dana Dachman-Soled, Siyao Guo, Tal Malkin, and Li-Yang Tan. Non-malleable codes for small-depth circuits. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 826–837. IEEE, 2018.

[BDSKM16] Marshall Ball, Dana Dachman-Soled, Mukul Kulkarni, and Tal Malkin. Non-malleable codes for bounded depth, bounded fan-in circuits. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 881–908. Springer, 2016.

[BDSKM18] Marshall Ball, Dana Dachman-Soled, Mukul Kulkarni, and Tal Malkin. Non-malleable codes from average-case hardness: Decision trees, and streaming space-bounded tampering. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 618–650. Springer, 2018.

[BS18] Saikrishna Badrinarayanan and Akshayaram Srinivasan. Revisiting non-malleable secret sharing. *IACR Cryptology ePrint Archive*, 2018:1144, 2018.

[CCFP11] Hervé Chabanne, Gérard Cohen, J Flori, and Alain Patey. Non-malleable codes from the wire-tap channel. In *Information Theory Workshop (ITW), 2011 IEEE*, pages 55–59. IEEE, 2011.

[CCP12] Herve Chabanne, Gerard Cohen, and Alain Patey. Secure network coding and non-malleable codes: Protection against linear tampering. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2546–2550. IEEE, 2012.

[CG14a] Mahdi Cheraghchi and Venkatesan Guruswami. Capacity of non-malleable codes. In *ITCS*, 2014.

[CG14b] Mahdi Cheraghchi and Venkatesan Guruswami. Non-malleable coding against bit-wise and split-state tampering. In *TCC*, 2014.

[CGL16] Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 285–298. ACM, 2016.

[CKM11] Seung Geol Choi, Aggelos Kiayias, and Tal Malkin. Bitr: built-in tamper resilience. In *Advances in Cryptology–ASIACRYPT 2011*, pages 740–758. Springer, 2011.

[CZ14] Eshan Chattopadhyay and David Zuckerman. Non-malleable codes in the constant split-state model. *To appear in FOCS*, 2014.

[DDN00] D. Dolev, C. Dwork, and M. Naor. Nonmalleable cryptography. *SIAM*, 30:391–437, 2000.

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

[DP07] Stefan Dziembowski and Krzysztof Pietrzak. Intrusion-resilient secret sharing. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 227–237. IEEE, 2007.

[DPW10] Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. In *ICS*, pages 434–452. Tsinghua University Press, 2010.

[DW09] Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In Michael Mitzenmacher, editor, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, pages 601–610, Bethesda, MD, USA, 2009. ACM.

[FHMV17] Sebastian Faust, Kristina Hostáková, Pratyay Mukherjee, and Daniele Venturi. Non-malleable codes for space-bounded tampering. In *Annual International Cryptology Conference*, pages 95–126. Springer, 2017.

[FMNV14] S. Faust, P. Mukherjee, J. Nielsen, and D. Venturi. Continuous non-malleable codes. In *Theory of Cryptography Conference - TCC*. Springer, 2014.

[FMVW14] S. Faust, P. Mukherjee, D. Venturi, and D. Wichs. Efficient non-malleable codes and key-derivation for poly-size tampering circuits. In *Eurocrypt*. Springer, 2014. To appear.

[GK18a] Vipul Goyal and Ashutosh Kumar. Non-malleable secret sharing. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 685–698. ACM, 2018.

[GK18b] Vipul Goyal and Ashutosh Kumar. Non-malleable secret sharing for general access structures. In Hovav Shacham and Alexandra Boldyreva, editors, *Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part I*, volume 10991 of *Lecture Notes in Computer Science*, pages 501–530. Springer, 2018.

[GLM+03] Rosario Gennaro, Anna Lysyanskaya, Tal Malkin, Silvio Micali, and Tal Rabin. Algorithmic Tamper-Proof (ATP) security: Theoretical foundations for security against hardware tampering. In Moni Naor, editor, *First Theory of Cryptography Conference — TCC 2004*, volume 2951 of *LNCS*, pages 258–277. Springer-Verlag, February 19–21 2003.

[GMW18] Divya Gupta, Hemanta K Maji, and Mingyuan Wang. Constant-rate non-malleable codes in the split-state model. Technical report, Technical Report Report 2017/1048, Cryptology ePrint Archive, 2018.

[GPR16] Vipul Goyal, Omkant Pandey, and Silas Richelson. Textbook non-malleable commitments. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 1128–1141. ACM, 2016.

[IPSW06] Yuval Ishai, Manoj Prabhakaran, Amit Sahai, and David Wagner. Private circuits II: Keeping secrets in tamperable circuits. In Serge Vaudenay, editor, *Advances in Cryptology—EUROCRYPT 2006*, volume 4004 of *LNCS*, pages 308–327. Springer-Verlag, 2006.

[ISW03] Yuval Ishai, Amit Sahai, and David Wagner. Private circuits: Securing hardware against probing attacks. In Dan Boneh, editor, *Advances in Cryptology—CRYPTO 2003*, volume 2729 of *LNCS*. Springer-Verlag, 2003.

[KKS11] Yael Tauman Kalai, Bhavana Kanukurthi, and Amit Sahai. Cryptography with tamperable and leaky memory. In *Advances in Cryptology–CRYPTO 2011*, pages 373–390. Springer, 2011.

[KOS17] Bhavana Kanukurthi, Sai Lakshmi Bhavana Obbattu, and Sruthi Sekar. Four-state non-malleable codes with explicit constant rate. In *Theory of Cryptography Conference*, pages 344–375. Springer, 2017.

[KOS18] Bhavana Kanukurthi, Sai Lakshmi Bhavana Obbattu, and Sruthi Sekar. Non-malleable randomness encoders and their applications. In *EUROCRYPT*, pages 589–617. Springer, 2018.

[Li17] Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1144–1156. ACM, 2017.

[Li18] Xin Li. Non-malleable extractors and non-malleable codes: Partially optimal constructions. *arXiv preprint arXiv:1804.04005*, 2018.

[LL12] Feng-Hao Liu and Anna Lysyanskaya. Tamper and leakage resilience in the split-state model. In *Advances in Cryptology–CRYPTO 2012*, pages 517–532. Springer, 2012.

[SV18] Akshayaram Srinivasan and Prashant Nalini Vasudevan. Leakage resilient secret sharing and applications. *IACR Cryptology ePrint Archive*, 2018:1154, 2018.