Detecting General Algebraic Manipulation Attacks

Kim Ramchen

Department of Computing and Information Systems
The University of Melbourne
kim.ramchen@unimelb.edu.au

Abstract

Algebraic manipulation detection codes are a class of error detecting codes which have found numerous applications in cryptography. In this paper we extend these codes to defeat general algebraic attacks - we call such codes general algebraic manipulation detection (GAMD) codes. Positive results are shown for the existence of GAMDs for the families of tampering functions corresponding to point additions and polynomial functions over a finite field. Compared to non-malleable codes, we demonstrate both positive and negative results regarding the existence of GAMDs for arbitrary families of tampering functions.

1 Introduction

Fault injection attacks are a class of attacks involve the deliberate introduction of errors into the circuity or memory modules of a cryptographic device in attempt to deduce some secret state. Algebraic manipulation detection codes [CDF+08] are a class of error detecting codes that can thwart such attacks when the class of induced faults corresponds to additions on code-words over a finite space. More precisely let s be a message supplied by an adversary, and suppose c, an element of an abelian group $\mathcal{G}$, is the corresponding code-word. If for any $\Delta \in \mathcal{G}$ it holds that $c + \Delta$ decodes to s' for any $s' \neq s$, with probability bounded by ϵ, the scheme is said to be an AMD code with error probability ϵ.

Even though AMD codes provide an elegant, keyless alternative to the widely used message authentication codes for robust transmission over an error-prone channel, they cannot defeat some types of powerful adversaries. Suppose that an AMD code is used to protect the output of a one time pad scheme. Let $\mathcal{E}(K \oplus M)$ be the output on ciphertext $c = K \oplus M$. If it happens that $\mathcal{E}$ possesses a linear homomorphism ϕ, then we have $\Delta M \circ_\phi \mathcal{E}(c) = \Delta M \circ_\phi \mathcal{E}(K \oplus M) = \mathcal{E}(K \oplus (M \oplus \Delta M)) = \mathcal{E}(K \oplus M')$, where M' is the message to be substituted. It is therefore desirable to consider a more powerful adversarial model in which an attacker can choose, in addition to the source message, a tampering

function F from a rich class of tampering functions $\mathcal{F}$. In this work, we consider precisely this model, when the class $\mathcal{F}$ corresponds to algebraic functions over some finite field or the rationals corresponding to the co-domain of the AMD code. We call such a code a generalised algebraic manipulation detection code (GAMD code). Following previous works on algebraic manipulation detection, we distinguish the case when the source message is assumed to be uniformly distributed over the message space, from the usual (which provides tampering detection with bounded error probability for any message). These are called weak generalised algebraic manipulation detection (weak GAMD) and generalised algebraic manipulation detection (GAMD) respectively.

1.1 Our Contributions

We formally introduce the model of generalised algebraic manipulation detection, in which tamperings corresponding to algebraic functions over the ambient field of the encoding function. In this model we review the previous constructions for manipulation detection against point additions. We show that such constructions translate directly to our new model, leading to direct instantiations of weak GAMDs and GAMDs for this class. Additionally we present a new construction for weak GAMDs in the case of encoding over $\mathbb{F}_2$ based upon the probabilistic method, leading to the following result (we actually construct a GAMD for a more general class of tampering functions, this is discussed in Section 3.1.1)

Theorem 1 (Probabilistic construction of addition evasive GAMDs - Informal)**.** *Let n be a power of two. There exists a n^{c-1}-GAMD against the class of point additions on $\mathbb{F}_n$ with rate $c - o(1)$, for any constant $0 < c < 1$.*

We also consider attacks corresponding to the class of polynomial functions. Such attacks in the affine case have been considered in the context of non-malleable cryptography by [ADL14, KLT16]. We demonstrate an explicit construction of a GAMD secure against the class of polynomial functions of bounded degree.

Theorem 2 (Construction of GAMDS for bounded degree polynomials - Informal)**.** *Fix a positive integer d. There exists an explicit weak ϵ-GAMD secure against the class of polynomials of degree bounded by d of rate $2/\Theta(d^2)$ and error probability $\frac{O(k)}{d} \cdot 2^{-\frac{k}{\Theta(d^2)}}$ where k is the prime bit-length.*

We show that exact constructions imply corresponding weak GAMD codes with inverse polynomial rate and low error-probability. We present a black-box transformation of any weak GAMD to a GAMD. This construction is quite efficient, implying in view of the above results, the existence of GAMDs with constant rate and low error probability for the classes of point additions and polynomial functions respectively. Compared to the celebrated non-malleable codes [DPW10] we also establish some separations. Our first result is negative and states that there exists a class of tampering functions for which non-malleable codes but not GAMD codes exist. This may be summarised by

Theorem 3 (Non-existence of GAMDs for all functions - Informal)**.** *There exists a family of tampering functions for which non-malleable codes exist with constant rate and negligible simulation error but ϵ-GAMD codes with constant rate do not exist, for any choice of non-negligible ϵ.*

Our second result is a positive one and states that for any non-malleable code there exists a class of tampering functions which violates non-malleability, but for which an efficient GAMD code exists, leading to

Theorem 4 (Existence of GAMDs breaking non-malleability - Informal)**.** *For any non-malleable code $\mathcal{C}$ there exists a family of tampering functions such that $\mathcal{C}$ is malleable with respect to this family but there exists a GAMD for this family with constant rate and negligible error probability.*

We also show how to extend the construction of non-malleable codes for the class of bounded degree polynomials to *super non-malleable codes* in the split-state model [DPW10]. The core observation behind this construction is that super non-malleable codes of Faust et al. [FMVW14] can be de-randomised by embedding t-wise independent hash functions inside a plain non-malleable code which is appended to the resulting codeword.

Theorem 5 (Super non-malleability in two-state model - Informal)**.** *In the two-state model there exists, for any $0 < \epsilon < 1$, an explicit ϵ-super non malleable code for the class of polynomials of degree bounded by d. The rate is $\frac{1}{\Theta(d^2)}$.*

A significant limitation of our results is that they only apply for tampering functions in one variable, while achieving corresponding deterministic results for multi-variate tampering classes seems considerably more challenging.

1.2 Related Work

Cabello et al. constructed AMD codes in the context of robust secret sharing [CPS02]. The notion was made explicit by the works of [DKRS06, CDF$^+$08] and some further applications provided including robust fuzzy extraction and message authentication codes with key manipulation security. In the former one wishes to guarantee recovery of a uniformly random key from biometric or other noisy data with the property that correctness is maintained under addition of errors up to some prior fixed bound even if the public parameters are compromised. In a similar vein the goal of the latter is to prevent forgery of message authentication tags even in the case that the adversary has algebraic manipulation access to the device storing the key. Other applications include robust information dispersal and anonymous message transmission [CDF$^+$08]. Dziembowksi et al. introduced the notion of non-malleable coding schemes and gave existential constructions for arbitrary tampering classes as well as efficient constructions in the random oracle [DPW10]. Liu et al. constructed computationally secure non-malleable codes for split-state tampering in the CRS model [LL12]. Dziembowksi et al. initiated the study of non-malleable codes from two-source extractors [DKO13]. Aggarwal et al. [ADL14] and Chattopadhyay et al. [CZ14] constructed explicit

efficient non-malleable codes in the split-state model. Faust et al. constructed asymptotically optimal non-malleable codes for sufficiently small tampering classes in the CRS model [FMVW14]. Faust et al. constructed non-malleable codes secure against continual leakage [FMNV14].

Although non-malleable cryptography is not the major focus of this work we show how to construct non-malleable codes from polynomial evasive GAMDs as well super non-malleable codes [FMVW14] for this class in the two-state model.

2 Preliminaries

We describe the preliminary tools and definitions to be used throughout this paper. We begin firstly by reviewing non-malleable codes [DPW10], secondly by stating some combinatorial results and finally, in Section 2.3, by stating our generalisation of classical algebraic manipulation detection codes [CPS02, DKRS06, CDF+08].

2.1 Non-Malleable Codes

We recall the notion of non-malleable codes for a class of tampering functions. Informally a non-malleable code is one which guarantees that after decoding either the original message is recovered or the message that is recovered is completely "unrelated" to the original.

Definition 1 (Non-Malleable Code [DPW10])**.** *Let $\mathcal{F}$ be a family of tampering functions. For each $F \in \mathcal{F}$ and $s \in \{0,1\}^k$, define the tampering experiment*

$$\mathsf{Tamper}_s^F =: \left\{ \begin{array}{c} c \leftarrow \mathsf{Enc}(s), \tilde{c} \leftarrow F(c), \tilde{s} = \mathsf{Dec}(c) \\ \textit{Output } \tilde{s}. \end{array} \right\}$$

defining a random variable over the randomness of the encoding function Enc*. Say that a coding scheme* $(\mathsf{Enc}, \mathsf{Dec})$ *is non-malleable w.r.t. $\mathcal{F}$ if for each $F \in \mathcal{F}$, there exists a distribution D_F over $\{0,1\}^k \cup \{\bot, \mathsf{same}^*\}$, such that, for all $s \in \{0,1\}^k$, we have:*

$$\mathsf{Tamper}_s^F \approx \left\{ \begin{array}{c} \tilde{s} \leftarrow D_F \\ \textit{Output } s \textit{ if } \tilde{s} = \mathsf{same}^*, \textit{ and } \tilde{s} \textit{ otherwise.} \end{array} \right\}$$

and D_F is efficiently samplable given oracle access to $F(\cdot)$.

Let $\mathcal{F}_{\mathsf{bit}}$ be the family of tampering functions that tamper every bit of a code-word of length n independently. Formally, $\mathcal{F}_{\mathsf{bit}}$ contains all functions $f : \{0,1\}^n \to \{0,1\}^n$ defined by n functions $f_i : \{0,1\} \to \{0,1\}$, namely $f(c_1, \ldots, c_n) = (f_1(c_1), \ldots, f_n(c_n))$. Each f_i is an affine function on $\mathbb{Z}_2$. We require the following proposition proved by [DPW10], concerning the existence of non-malleable codes against the family of bit-wise independent tampering functions with constant rate and negligible simulation error.

Lemma 6 (Theorem 4.2 [DPW10]). *For any $\delta > 0$ and $n \in \mathbb{N}$ there exist non-malleable codes w.r.t the family $\mathcal{F}_{\mathsf{bit}}$, with block length n, message size $k \geq (.18 - \delta)n$ and simulation error $2^{-\Omega(n)}$. Moreover there is an efficient procedure which, given k and n, outputs a description of such a code with probability $1 - 2^{-\Omega(n)}$.*

We will also use the notion of super non-malleability [FMVW14] in the split-state model [DPW10].

Definition 2 (Super Non-Malleability [FMVW14]). *Let $\mathsf{Enc} : \{0,1\}^k \to \{0,1\}^n$, $\mathsf{Dec} : \{0,1\}^n \to \{0,1\}^k$ be a coding scheme and $\mathcal{F}$ be a family of functions $f : \{0,1\}^n \to \{0,1\}^n$. We say that the scheme is $(\mathcal{F}, \epsilon)$-super non-malleable if for any $m_0, m_1 \in \{0,1\}^k$ and any $f \in \mathcal{F}$, we have $\mathsf{Tamper}^f_{m_0} \approx_\epsilon \mathsf{Tamper}^f_{m_1}$ where:*

$$\mathsf{Tamper}^f_m := \left\{ \begin{array}{c} c \leftarrow \mathsf{Enc}(x), c' = f(c) \\ \textit{Output } \mathsf{same}^* \textit{ if } c' = c, \textit{ output } \perp \textit{ if } \mathsf{Dec}(c') = \perp \\ \textit{and else output } c'. \end{array} \right.$$

Theorem 7 ([FMVW14]). *Let $\mathcal{H}_1 = \{h_1\}$ and $\mathcal{H}_2 = \{h_2\}$ be t-wise independent hashing families where $h_1 : \{0,1\}^{v_1} \to \{0,1\}^k$ and $h_2 : \{0,1\}^{k+v_1} \to \{0,1\}^{v_2}$. Then for any function family $\mathcal{F}$, consisting of functions $f : \{0,1\}^n \to \{0,1\}^n$ there exists an $(\mathcal{F}, \epsilon)$-super non-malleable code with probability $1 - p$ provided that*

$$\begin{gathered} t \geq O(\log |\mathcal{F}| + n + \log(1/p)) \\ v_1 > 3\log(1/\epsilon) + 3\log t + O(1) \\ v_2 > v_1 + 3. \end{gathered}$$

2.2 Combinatorial Tools

We describe some combinatorial tools used in our constructions of GAMDs.

Definition 3 (Balanced Block Design [CD06]). *Let v, c, λ be a positive integers. For point set V a balanced block design is a multiset $\mathcal{B}$ of blocks of points such that*

1. *$|V| = v$*
2. *$|P| = c$ for each $P \in \mathcal{B}$*
3. *Each pair of points is a subset of exactly λ blocks*

if $|\mathcal{B}| = v$ say that the (v, c, λ)-balanced block design is symmetric.

Definition 4 (Trace [CDN15]). *Let K and L be fields. Suppose that L is separable over K and $n := [L : K] > \infty$. Fix some algebraic closure $\bar{L}$ of L. Let $\sigma_1, \ldots, \sigma_n$ be the distinct K-embeddings of L into $\bar{L}$. The trace map $\mathrm{Tr}_{L/K}$ for each $x \in L$ is:*

$$\mathrm{Tr}_{L/K}(x) = \sum_{i=1}^{n} \sigma_i(x) \in K$$

Definition 5 (Difference Set [CD06]). *Let $(\mathcal{G}, +)$ be an additive abelian group of order v. A subset $D \subseteq \mathcal{G}$ is a (v, c, λ)-external difference set if $|D| = c$ and every non-zero element of $\mathcal{G}$ has exactly λ representations as a difference $d - d'$ for $d, d' \in D$. If every non-zero element of $\mathcal{G}$ has at most λ representations $d - d'$, say that D is a (v, c, λ)-bounded difference set.*

Definition 6 (Authentication Code [Sti90, Sti94]). *Let $\mathcal{S}$ be a set of source states, $\mathcal{K}$ a set of authentication keys and $\mathcal{A}$ be a mapping $\mathcal{A} : \mathcal{S} \times \mathcal{K} \to \mathcal{T}$ where $\mathcal{T}$ is a set of tags. Let Π be a probability distribution on $\mathcal{K}$. The probability of a successful substitution attack, with respect to family of substitution functions $\mathcal{F}$, is*

$$p_{\mathcal{F}}^{\mathsf{sub}} =: \max_{F \in \mathcal{F}, s \neq s' \in \mathcal{S}} \Pr_{K \leftarrow \Pi}[F(\mathcal{A}(s, K)) = \mathcal{A}(s', K)].$$

Lemma 8 (Schwartz-Zippel). *Let K be a field and let $P \in K[x_1, \ldots, x_n]$ where $(x_i)_{1 \leq i \leq n}$ are indeterminates. Let $S \subseteq K$ be a finite set and let $(u_i)_{1 \leq i \leq n}$ be selected independently and uniformly at random in S. Then*

$$\Pr[P(u_1, \ldots, u_n) = 0] \leq \frac{\deg(P)}{|S|}$$

Lemma 9 (Prime Number Theorem [Ros94]). *Let $\pi(x)$ denote the number of primes p which satisfy $2 \leq p \leq x$. Then*

$$\lim_{x \to \infty} \pi(x) \cdot \frac{\ln(x)}{x} = 1$$

.

2.3 Generalised Algebraic Manipulation Detection Codes

In this section we define a code which is a generalisation of the classical algebraic manipulation detection coding schemes. The main difference is simply that we allow manipulation functions be general algebraic functions over a field, rather than the restriction to point additions on its group considered by [CPS02, CDF+08]. In this paper K will always be a finite field or number field (finite extension of the rationals), however below we allow K to be arbitrary for completeness.

Definition 7. *Let K be a field with associated metric $d : K^2 \to \mathbb{R}^+ \cup \{0\}$. Let $\mathcal{G} := K$ and let $\mathcal{F}$ be a family of algebraic tampering functions on $\mathcal{G}$. Let $\mathcal{S}$ be a set of symbols. Let $\mathcal{E} : \mathcal{S} \to \mathcal{G}$ be a probabilistic encoding and $\mathcal{D} : \mathcal{G} \to \mathcal{S} \cup \{\perp\}$ be a deterministic decoding procedure such that $\Pr_{\mathcal{E}}[\mathcal{D}(\mathcal{E}(s)) = s] = 1$ for all $s \in \mathcal{S}$.*

- *The tuple $(\mathcal{E}, \mathcal{D})$ is an ϵ-generalised algebraic manipulation detection (GAMD) code if $\forall s \in \mathcal{S}, \forall F \in \mathcal{F}$ $\Pr_{\mathcal{E}}[\mathcal{D}(F(\mathcal{E}(s))) \notin \{s, \perp\}] \leq \epsilon$.*
- *The tuple $(\mathcal{E}, \mathcal{D})$ is a weak ϵ-generalised algebraic manipulation detection code if $\forall F \in \mathcal{F}$ $\Pr_{\mathcal{E}, s \in_R \mathcal{S}}[\mathcal{D}(F(\mathcal{E}(s))) \notin \{s, \perp\}] \leq \epsilon$.*

Let $B_d(0, \delta)$ be the set of points at distance at most δ from $0_{\mathcal{G}}$. The (information) rate of a GAMD code is defined as $r = \lim_{\delta \to \infty} \frac{\log_2 |\mathcal{E}(\mathcal{S}) \cap B_d(0,\delta)|}{\log_2 |\mathcal{G} \cap B_d(0,\delta)|}$.

2.3.1 Families of Tampering Functions

In this paper we consider two classes of tampering functions on a GAMD $(\mathcal{E}, \mathcal{D})$ with co-domain $\mathcal{G} = \mathbb{F}_{p^n}$ for some prime p and positive integer n.

- **Point Additions:** let $\mathcal{F}_{\mathsf{add}} = \{F_\Delta\}_{\Delta \in \mathcal{G}}$ where $F_\Delta := x \mapsto x + \Delta$ over $\mathcal{G}$.
- **Polynomial Functions:** let $\mathcal{F}_{\mathcal{P}\leq d} = \{F_{(\vec{a})}\}_{\vec{a} \in \mathcal{G}^{d+1}}$ where $F_{(\vec{a})} := x \mapsto \sum_{i=0}^{d} a_i x^i$ over $\mathcal{G}$.

2.4 Notation

Write $f = o(g)$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$. Write $f = \Omega(n)$ if $\exists\ c > 0$ and $N_0 > 0$ such that for all $n > N_0$, $f(n) \geq c \cdot g(n)$. Let $e(\cdot)$ denote the real-valued exponential function. Let $\mathsf{SD}(,\cdot,)$ denote the statistical distance. For discrete probability distributions with outcome space $\mathcal{X}$, $\mathsf{SD}(P_0, P_1) = \frac{1}{2}\sum_{x\in\mathcal{X}} |P_0(x) - P_1(x)|$. For probability distributions P_0 and P_1 let $D(P_0\|P_1)$ denote the KL-divergence. Pinsker's inequality states that $D(P_0\|P_1) \geq 2\mathsf{SD}(P_0\|P_1)^2$. Say that random variables $X_1, \ldots, X_n$ are k-wise independent if $\Pr[X_{i_1} = a_1, \ldots, X_{i_k} = a_k] = \prod_{j=1}^{k} \Pr[X_{i_j} = a_j]$ for all $\{i_1, \ldots, i_k\} \subseteq [1, n]$. A function is algebraic iff it is the root of a polynomial equation. Let $\mathbb{Q}$ be the set of rationals. For field K, let $\mathcal{P}_{\leq d}$ be the space of univariate polynomials of degree at most d over K. For even integer n denote by I_n, the subset of permutations on n objects consisting of involutions with no fixed points. The independence number of a finite graph G is the size of the largest complete graph in the edge complement of G.

2.5 Tail Bounds on Sums of Dependent Variables

Lemma 10 (Multiplicative Chernoff Bound)**.** *Let $\{X_i\}_{1\leq i\leq n}$ be a sequence of independent random variables such that $0 \leq X_i \leq 1$, $E[X_i] = p$ for $1 \leq i \leq n$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X] = np$. Fix $0 < \delta < 1$. Then*

$$\Pr[X < \mu(1-\delta)] \leq e(-\frac{\delta^2\mu}{2})$$
$$\Pr[X > \mu(1+\delta)] \leq e(-\frac{\delta^2\mu}{3})$$

Lemma 11 (Theorem 1.12 [PR15])**.** *Let $\{X_i\}_{1\leq i\leq n}$ be a sequence of k-wise independent random variables such that $0 \leq X_i \leq 1$, $E[X_i] = p$ for $1 \leq i \leq n$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X] = np$. Fix $\delta > 0$. Then*

$$\Pr[X > \mu(1+\delta)] \leq \frac{1}{(p-p^2)^{n-k}} e\left(-nD(p(1+\delta)\|p)\right)$$

Lemma 12 (Theorem 1.14 [PR15])**.** *Let $G = (V, E)$ be a finite graph with vertices $v_1, \ldots, v_n$ and let α be its independence number. To each v_i, $i = 1, \ldots, n$ we associate a Bernoulli*

0/1 random variable B_i, such that $Pr[B_i = 1] = p$. Suppose that each random variable $B_i, i = 1, \ldots, n$ is independent of the set $\{B_j : (v_i, v_j) \notin E\}$. Let $0 < \delta < 1$ be a constant and $t = np(1 + \delta)$. Then

$$\Pr[\sum_{i=1}^{n} B_i \geq t] \leq p^{\alpha} \cdot e(-\frac{\delta^2 n}{2}) \cdot 2^n$$

3 Constructions

In this section we review some constructions for GAMD codes against the class of tampering functions corresponding to point additions and also polynomial functions. Our results show that efficient GAMDs (i.e, one ones with constant rate and low error probability) exist for the former class, while for the latter, the rate degrades quadratically in the degree of the function. For the class of point additions, we present two constructions of GAMDs based upon difference sets. Our first can be seen as a specific instantiation of the AMD codes in Section 4.1 [CPS02]. Our second which is based upon the probabilistic method allows the construction of GAMDs for a broader class of functions.

3.1 Point Additions

Cabello et al. [CPS02] constructed a difference set in $\mathbb{F}_{p^l} \times \mathbb{F}_{p^k}$ from any surjective map $\phi : \mathbb{F}_{p^l} \to \mathbb{F}_{p^k}$. An efficient instantiation of ϕ for arbitrary p can be found using the field trace (Definition 4). Using this construction we can build a weak-GAMD with rate $1 - o(1)$ and arbitrarily low error probability, described in Lemma 14.

Lemma 13. *[CPS02] Let p be an odd prime and l and k be positive integers such that $l \equiv 0 \pmod{k}$. Let $(\mathcal{G}, +)$ be the product of groups, $\mathbb{F}_{p^l} \times \mathbb{F}_{p^k}$ under addition. Define*

$$D_{k,l} = \{(\alpha, \phi_{\mathbb{F}_{p^l}/\mathbb{F}_{p^k}}(\alpha^2)) : \alpha \in \mathbb{F}_{p^l}\} \subseteq \mathcal{G}$$

Then $D_{l,k}$ is a (p^{l+k}, p^l, p^{l-k})-external difference set.

Lemma 14. *For a prime p and positive integer n let $\mathcal{G} = \mathbb{F}_p^n$. Then there exists a explicit weak (p^{-1})-GAMD code with respect to the family of point additions, $\mathcal{F}_{\mathsf{add}}$, on $\mathcal{G}$, with efficient encoding and decoding procedure and rate $1 - o(1)$.*

Proof. Note $\mathcal{G} \sim (\mathbb{F}_{p^n}, +)$. By Lemma 13 we know that for any $n > 1$ there exists a (p^n, p^{n-1}, p^{n-2})-external difference set $D_{1,n-1} \subseteq \mathcal{G}$. Let $\mathcal{E}(\mathcal{S}) = D_{1,n-1}$ and consider the quantity $p_\Delta := \Pr_{s \in_R S}[F_\Delta(\mathcal{E}(s)) \notin \{s, \perp\}]$. Since $\mathcal{E}$ is deterministic, and s is chosen uniformly at random, $p_\Delta = \#\{s' \in S : \mathcal{E}(s') - \mathcal{E}(s) = \Delta\}/|\mathcal{S}|$. Thus for each $\Delta \in \mathcal{G}$, since $\mathcal{E}(\mathcal{S})$ is a (p^n, p^{n-1}, p^{n-2})-difference set, $p_\Delta \leq p^{n-2}/p^{n-1} = p^{-1}$. The rate of $\mathcal{E}$ is $\frac{\log |D_{1,n-1}|}{\log |\mathcal{G}|} = 1 - n^{-1} = 1 - o(1)$, as required. □

3.1.1 A New Construction

We note that so far the constructions of GAMD codes against the class of point additions have followed a similar recipe to the constructions of AMD codes presented in [CPS02, CDF+08]. In this section we present a new construction for this class based upon the probabilistic method.

Lemma 15. *Let $\mathcal{G}$ be an abelian group of order n where n is even. Let $0 \leq c < 1$ be arbitrary. Let $I'_n \subset I_n$ be of polynomial size. Then there exists a subset $S \subset \mathcal{G}$ and maps $\mathcal{E} : [|S|] \to \mathcal{G}$ and $\mathcal{D} : \mathcal{G} \to [|S|]$ which define a weak n^{c-1}-GAMD with respect to the set I'_n. The rate is ρ is $c - o(1)$. The sampling error is $e(-\frac{1}{4}n^{\rho}) + |I'_n| \cdot e(-2n^{2\rho-1})$.*

Proof. We show that, by analogy to the polynomial-evasive set case (see Section 3.2 following) for any positive constants $0 \leq \gamma < \nu < 1$, there exists a set $S \subset \mathcal{G}$ for which $|S| \in \gamma|\mathcal{G}|(1 \pm \epsilon)$ and $|S \cap F(S)| \leq \nu|S|$ hold for any $F \in I'_n$. Taking $\mathcal{S} = [|S|]$, $\nu = n^{c-1}$, $\gamma = n^{(c-1)-o(1)}$ and $\mathcal{E}$ and $\mathcal{D}$ as in the statement of the lemma, yields a code with error probability n^{c-1} and rate $\frac{\log \gamma n}{\log n} = c - o(1)$ (see also Theorem 29, Appendix C). We will demonstrate the existence of S via a probabilistic argument. Consider the set S defined by sampling each element of $\mathcal{G}$ independently with probability γ. Clearly the size of S, N_0, has a Binomial distribution with parameters (n, γ). We now analyse the size of the intersection $S \cap F(S)$, where $F \in I'_n$ is arbitrary. Observe that each such F induces a matching on $\mathcal{G}$ given by $(x, F(x)) : x < F(x)$. Moreover, since F contains no fixed points, each such pair occurs independently with probability γ^2. Thus $N_1 := |S \cap F(S)|/2$ follows a Binomial distribution, with parameters $(\frac{n}{2}, \gamma^2)$. Now by applying Lemma 10, if ϵ is such that $\gamma < \nu(1-\epsilon) < 2\gamma$ then

$$\Pr[N_0 \leq n\gamma(1-\epsilon)] \leq e(\frac{-n\gamma\epsilon^2}{2}) \tag{1}$$

$$\Pr[N_1 \geq \frac{\nu n\gamma(1-\epsilon)}{2}] \leq e(\frac{-n(\nu(1-\epsilon)-\gamma)^2}{6}) \tag{2}$$

Secondly, applying a union bound over all $F \in I'_n$, we have $\Pr_S[|S| \geq n\gamma(1-\epsilon) \cap |S \cap F(S)| \leq \nu n\gamma(1-\epsilon)$ for all $F \in I'_n] \geq 1 - e(-\frac{n\gamma\epsilon^2}{2}) - |I'_n|e(-\frac{n(\nu(1-\epsilon)-\gamma)^2}{6})$. As $|I'_n|$ is polynomial in n, for large enough n this probability is strictly greater than 0. Let $k > 1$ and $\nu = \epsilon^{k-1}, \gamma = \epsilon^k$. Then the function $g(\cdot) = (\epsilon^{k-1}(1-\epsilon) - \epsilon^k)^2$ is maximised on the interval $(0,1)$ by $\epsilon_0 = \frac{k-1}{2k}$. In particular for $\epsilon = \epsilon_0$, $\rho \geq \frac{\log_2(n \cdot 2^{-(k+1)})}{\log_2 n}$ and Equation 1 implies $\Pr[N_0 \leq n\gamma(1-\epsilon_0)] \leq e(\frac{-n\epsilon_0^{k+2}}{2}) \leq e(-n \cdot (\frac{1}{2})^{k+3}) \leq e(-\frac{n^{\rho}}{4})$. Equation 2 on the other hand implies $\Pr[N_1 \geq \frac{\nu n\gamma(1-\epsilon_0)}{2}] \leq e(-\frac{n\epsilon_0^{2k-1}(1-2\epsilon_0)^2}{6}) \leq e(-n \cdot (\frac{1}{2})^{2k+1}) \leq e(-2n^{2\rho-1})$. Thus the sampling error is $e(\frac{-n^{\rho}}{4}) + |I'_n| \cdot e(-2n^{2\rho-1})$. □

Corollary 16. *Let $G = (\mathbb{F}_n, +)$ where n is an arbitrary power of two. Then there exists a weak $(n^{-1/2})$-GAMD with respect to the family $\mathcal{F}_{\mathsf{add}}$, with rate $\frac{1}{2} - o(1)$. The sampling error is $e(-\frac{1}{4}n^{1/2}) + n^{-0.1}$.*

Proof. The family $\mathcal{F}_{\mathsf{add}}$ defines a subset of I_n of order n. Thus $S \subseteq \mathcal{G}$ exists with the properties of Lemma 15, taking $c = 1/2$ yields a $n^{-1/2}$-GAMD with rate $1/2 - o(1)$. Let $\rho = \frac{(1-\ln 2+\ln(1.1 \ln n))}{2}$. Then the sampling error is $e(-\frac{n^\rho}{4}) + n \cdot e(-2n^{2\rho-1}) \leq e(-\frac{n^{1/2}}{4}) + n \cdot e(-\ln(1.1 \cdot n))$. S defines an $(n, \sqrt{n}, 1)$-bounded difference set. □

We remark that the parameters achieved by Lemma 15 are essentially optimal - matching those of classical parameter sets modulo two [CD06]. In Appendix A we also prove the following result concerning the class $\mathcal{F}_{\mathsf{add}}$ over the cartesian power of a field K corresponding to the finite extensions of K under addition.

Lemma 17. *Let $(\mathcal{E}', \mathcal{D}')$ be a weak γ-GAMD over field $(K, +)$ for the class $\mathcal{F}_{\mathsf{add}}$ with rate ρ'. Then there exists $(\mathcal{E}, \mathcal{D})$, a weak γ-GAMD for $\mathcal{F}_{\mathsf{add}}$ over $(K^m, +)$, with rate $\rho := \rho'$ and $\gamma = 1 - (1 - \gamma')^m$.*

3.2 Polynomial Functions

In this section we show to construct explicit GAMDs secure against the class of all polynomials of finite degree modulo a prime, extending the constructions in [ADL14, Agg15]. We first present an informal overview of our construction, while the construction itself is described in section 3.2.1.

Our Construction In A Nutshell Aggarwal [Agg15] constructed codes secure against affine functions by constructing affine-evasive sets modulo a prime. The construction uses the reciprocals of all primes less than some inverse power in the underlying modulus. Fix an affine function F and let the reciprocal primes in its domain be denoted a_i and the primes in its range be denoted b_i. In that case an explicit bi-variate quartic relation is derived on the a_i and b_i [Agg15]. We follow this principle but instead use Lagrange interpolation to derive a (cyclically) symmetric relation on the a_i and b_i. Unfortunately the setting $d > 1$ necessitates some changes. Firstly there is no longer symmetry between the a_i and b_i which appears to be unique to the affine setting only. This implies divisibility relations appear possible only from the b_i (primes in the range of the polynomial). We are able to utilise these at slight expense (roughly $O(\log \log k)$ in bit-length) by an additive combinatorics-like construction of a set of primes with the property that no difference of elements of the set is divisible by another element. We believe this construction, which Lemma 18 is devoted to, may be of independent interest.

3.2.1 Construction of Polynomial Evasive GAMDs

Lemma 18. *For any positive integer N there exists a positive integer B, so that N primes lie in the interval $[0, B]$ and such that no prime divides the difference of two others for $B = O(N \ln^{1+o(1)} N)$.*

Proof. By Lemma 9 we can find $\Theta(\frac{B}{\ln B})$ primes q_i in the interval $(B/2, B]$. Suppose $q_i \mid q_j - q_k$ for some $q_i \neq q_j \neq q_k$. Then $B/2 < q_i \leq |q_j - q_k| \leq B/2$ which is a contradiction. □

For positive integer N, denote the above set D_N.

Theorem 19. *Let p a prime of k bits. There exists an explicit weak ϵ-GAMD secure against the class $\mathcal{F}_{\mathcal{P}\leq \mathsf{d}}$ modulo p of rate $2/\Theta(d^2)$ and error probability $\epsilon = \frac{O(k)}{d} \cdot 2^{-\frac{k}{\Theta(d^2)}}$ for any positive integer d.*

Proof. As mentioned above, define $N(p) = d^2 p^{2/(d^2+3d-2)}/4\ln^{1.1} p$ so that $q \in D_{N(p)}$ satisfies $q < (1-d^{-1.9}) \cdot p^{2/(d^2+3d-2)}$. Let

$$\mathfrak{P}_d := \{q^{-1} |\ q \text{ prime}, q \in D_{N(p)}\}$$

Fix $\vec{a} = (a_0, \ldots, a_{d-1}) \in \mathbb{F}_p$ and define $F_{\vec{a}}(x) = \sum_{i=0}^{d-1} a_i x^i$. We will prove that $|S \cap F_{\vec{a}}(S)| \leq d$. Suppose to the contrary that there exist distinct $(x_i)_{i=1}^{d+1}$ and $(y_i)_{i=1}^{d+1}$ in $\mathbb{F}_p$ such that $F_{\vec{a}}(x_i) = y_i$. Let $\mathcal{L}_j$ be the j^{th} Lagrange basis polynomial in the interpolation of $(x_i, y_i)_{i=1}^{d+1}$. In that case one has

$$L(x) = \sum_{j=1}^{d+1} \mathcal{L}_j(x) = \sum_{j=1}^{d+1} y_j \frac{\prod_{k\neq j}(x-x_k)}{\prod_{k\neq j}(x_j - x_k)}$$

Observe that $F_{\vec{a}}(x) = \sum_{i=0}^{d-1} a_i x^i$ is of degree $d-1$, while $L(x)$ is nominally of degree d. It follows that the leading coefficient of $L(\cdot)$ is zero and hence that

$$\sum_{j=1}^{d+1} \frac{y_j}{\prod_{k\neq j}(x_j - x_k)} \equiv 0 \bmod p \tag{3}$$

Write $x_j = a_j^{-1}$ and $y_j = b_j^{-1}$. WLOG $a_1 \neq b_1$, since for any non-trivial $F_{\vec{a}}$ the polynomial $F_{\vec{a}}(x) - x$ has at most $d-1$ roots. Therefore

$$\sum_{j=1}^{d+1} \frac{a_j^d \prod_{k\neq j} a_k}{b_j \cdot \prod_{k\neq j}(a_j - a_k)} \equiv 0 \bmod p$$

Multiplying out and clearing common terms

$$\sum_{j=1}^{d+1} ((-1)^j a_j^{d-1} \cdot \prod_{k\neq j} b_k \cdot \prod_{l>k, k\neq j} (a_l - a_k)) \equiv 0 \bmod p \tag{4}$$

Since $a_j, b_j < (1-d^{-1.9}) \cdot p^{2/(d^2+3d-2)}$ and $|a_l - a_k| < \max\{a_k, a_l\}$ for every $k < l$, Equation 4 holds over the integers. In particular, since b_1 appears in every summand except the first

$$b_1 \mid a_1^{d-1} \cdot \prod_{k=2}^{d+1} b_k \cdot \prod_{l>k, k\geq 2} (a_l - a_k) \tag{5}$$

We now derive a contradiction as follows. By assumption b_1 is distinct from and hence coprime to a_1 and $(b_i)_{i\geq 2}$. Then $b_1 \mid (a_l - a_k)$ for some $l > k$ which by our construction of $\mathfrak{P}_d$ is impossible. □

In Appendix A we also prove

Theorem 20. *Let p be a prime. There exists some constant c so that for any $0 < \epsilon < 1$ there exists a ϵ-non-malleable code $(\mathsf{Enc}, \mathsf{Dec})$ for the class $\mathcal{F}_{\mathcal{P}\leq d}$ where $\mathsf{Enc} : \mathbb{Z}_T \to \mathbb{F}_p$ and $\mathsf{Dec} : \mathbb{F}_p \to \mathbb{Z}_T$ whenever $p > (\frac{T}{\epsilon})^{c \cdot d^2}$.*

We remark that Theorem 19 extends to all finite centred Laurent expansions, i.e., *two-sided polynomial expressions about zero*, as well as to finite fields with similar parameters.

4 A Weak GAMD to GAMD Transformation

In this section we present a sufficient result for transforming any weak GAMD to a GAMD following a similar idea to that presented in Section 4 [CDF+08]. Our main result here is Lemma 22 which states that if the classes of tampering functions can be represented by a set of polynomials in one or more variable of bounded degree $d \ll |\mathcal{K}|$ then any weak GAMD for this family can transformed to a GAMD. In particular this implies asymptotically efficient GAMDs for the class of polynomial functions with negligible error probability.

Proposition 21. *Suppose that $(\mathcal{E}', \mathcal{D}')$ is a weak ϵ'-GAMD with respect to $\mathcal{F}$ where $\mathcal{E}' : \mathcal{S}' \to \mathcal{G}'$. Let $\mathcal{A} : \mathcal{S} \times \mathcal{S}' \to \mathcal{T}$ be an authentication code. Let $\mathcal{G} = \mathcal{S} \times \mathcal{G}' \times \mathcal{T}$. Define $\mathcal{E} : \mathcal{S} \to \mathcal{G}$ by $\mathcal{E}(s) = (s, \mathcal{E}'(k), \mathcal{A}(s,k))$, where $k \in_R \mathcal{S}'$. Define $\mathcal{D} : \mathcal{G} \to \mathcal{S} \cup \{\perp\}$ by $\mathcal{D}(s, c', \tau) = s$ iff $\mathcal{D}'(c') \neq \perp$ and $\tau = \mathcal{A}(s, \mathcal{D}'(c'))$. Then $(\mathcal{E}, \mathcal{D})$ is an ϵ-GAMD with respect to $\mathcal{F}$ where $\epsilon = \epsilon' + p^{\mathsf{sub}}_{\mathcal{F}}$.*

Proof. Suppose that $c = (\tilde{s}, \tilde{c'}, \tilde{\tau})$ is a received code-word for source symbol s under key k. Suppose that $s \neq \tilde{s}$. Then $\Pr[\mathcal{D}'(\tilde{c'}) \neq \{k, \perp\}] \leq \epsilon'$ since $(\mathcal{E}', \mathcal{D}')$ is a weak ϵ'-GAMD and k is chosen uniformly at random in $\mathcal{K}$. Moreover, $\Pr[\mathcal{A}(\tilde{s}, k) = \tilde{\tau}] \leq p^{\mathsf{sub}}_{\mathcal{F}}$ since $s \neq \tilde{s}$. Thus the event $\mathcal{D}(c) = \tilde{s}$ occurs with probability at most $\epsilon' + p^{\mathsf{sub}}_{\mathcal{F}}$. The result follows. □

Lemma 22. *Let ℓ be an arbitrary positive integer and K be a field. Let $\mathcal{K} \subseteq K^2$ be a finite set and $\mathcal{A} : \mathcal{S} \times \mathcal{K} \to \mathcal{T}$ be the message authentication code defined by $\mathcal{A}((s_1, \ldots, s_\ell), (x, y)) = \sum_{i=1}^{\ell} s_i x^i + y$. Then $p^{\mathsf{sub}}_{\mathcal{F}_{\mathcal{P}_{\leq d}}} \leq \frac{\ell d}{|\mathcal{K}|}$.*

Proof. Let F be a fixed polynomial in $\mathcal{F}_{\mathcal{P}_{\leq d}}$. Let $s \neq s' \in \mathcal{S}$. Consider the polynomial $P(x, y) = F(\sum_{i=1}^{\ell} s_i x^i + y) - (\sum_{i=1}^{\ell} s'_i x^i + y)$ in $K[x, y]$. We argue this is a non-zero polynomial as follows. First observe that if $P \equiv 0$, then $\deg(F) = 1$, since otherwise $P(x, y)$ contains a non-trivial power of y. So let $F(u) = a_0 u + a_1$. Then $a_0 = 1$ by a similar argument. Thus $P = \sum_{i=1}^{\ell} (s_i - s'_i) x^i + a_1$, which is a contradiction since $s \neq s'$ implies there exists i for which $s_i \neq s'_i$. On the other hand the degree of P is at most $\deg(F) \cdot \ell \leq \ell d$. Thus by Lemma 8, as $k = (x, y)$ is chosen uniformly in $\mathcal{K}$, the event $P = 0$ occurs with probability at most $\frac{\ell d}{|\mathcal{K}|}$. Finally, $P = 0$ occurs iff $F(\mathcal{A}(s, k)) = \mathcal{A}(s', k)$, concluding the proof. □

Corollary 23. *For any $n \in \mathbb{N}$ and large enough prime p there exists an ϵ-GAMD of block length n with respect to the family $\mathcal{F}_{\mathcal{P}\leq d}$ over $\mathbb{F}_p$ where $\epsilon = 2^{-n/\Theta(d^2)}$. The rate is $1 - o(1)$.*

Proof. Pick prime p so that $p > 2^n$. By Theorem 19 we can construct $\mathcal{E}'$ over $\mathbb{F}_{p^2}$ so that $\epsilon' \leq \frac{O(\log p)}{d} p^{-1/\Theta(d^2)}$. Let $\mathcal{A} : \mathbb{F}_p^{n-3} \times \mathbb{F}_p^2 \to \mathbb{F}_p$ be as in Lemma 22. Then as $\deg(F) \leq d$ for all $F \in \mathcal{F}_{\mathcal{P}\leq d}$, we have $p^{\mathsf{sub}}_{\mathcal{F}_{\mathcal{P}\leq d}} \leq \frac{(n-3)d}{p^2}$ by Lemma 22. The rate of $\mathcal{E}$ is $\frac{n-3}{n} = 1 - o(1)$. The error probability is bounded by $\epsilon = p^{\mathsf{sub}}_{\mathcal{F}_{\mathcal{P}\leq d}} + \epsilon' \leq \frac{n}{d} \cdot 2^{-n/\Theta(d^2)} + 2^{-\Omega(n)} = 2^{-n/\Theta(d^2)}$. □

5 Separations

We describe some separations regarding non-malleable codes and GAMDs generalising separations noted in previous works [DPW10, DKO13, FMVW14]. Although non-malleable codes have already proved a valuable digression from the classical notion of error correction and detection, here we provide evidence that GAMD codes provide a strengthening of classical algebraic manipulation detection distinct to that provided by non-malleable cryptography. Specifically we are able to prove (Theorem 25) that any non-malleable code can be broken by some tampering family for which a GAMD with high rate and low error probability exists. This family actually corresponds to a re-coding functionality in which a code-word is decoded, one is added to the message which is then again encoded, so is a natural candidate for this task. On the negative side, however, we show that for at least one family of tampering functions, non-malleable codes exist but GAMDs do not. We also prove that in the two-state model super non-malleable codes exist for arbitrary ϵ agains the class $\mathcal{F}_{\mathcal{P}\leq d}$ with inverse polynomial rate.

Theorem 24. *Some family of tampering functions $\mathcal{F}$ exists for which for any $n \in \mathbb{N}$, non-malleable codes of block length n exist with constant rate and simulation error $2^{-\Omega(n)}$, but ϵ-GAMD codes with constant rate do not exist, for any choice of non-negligible (in block-length) ϵ.*

Proof. Let $\mathcal{F} = \mathcal{F}_{\mathsf{bit}}$ be the family of bit-wise independent tampering functions (see Section 2.1). By Lemma 4 we know that for any n there exists a non-malleable code $(\mathsf{Enc}, \mathsf{Dec})$ with block size n, constant rate and simulation error $2^{-\Omega(n)}$. Now suppose that an ϵ-GAMD $(\mathcal{E}, \mathcal{D})$ exists for $\mathcal{F}_{\mathsf{bit}}$ where $\mathcal{E} : \mathcal{S} \to \mathcal{G}$ where $\mathcal{G} = \mathbb{F}_2^n$. There exists distinct code-words $c = \mathsf{Enc}(s), c' = \mathsf{Enc}(s') : c \neq c'$. Then the function $F_{c'}(x) = c'$ is contained in $\mathcal{F}_{\mathsf{bit}}$ and is not detectable except with probability at most $\frac{1}{|\mathcal{S}|}$. By assumption $(\mathcal{E}, \mathcal{D})$ is constant rate, so $\epsilon \leq \frac{1}{|\mathcal{S}|} = 2^{-\Omega(n)}$. The theorem follows. □

Theorem 25. *For any non-malleable code $\mathcal{C}$ of block length n there exists a family of tampering functions $\mathcal{F}_{\mathcal{C}}$ such that $\mathcal{C}$ is malleable with respect to $\mathcal{F}_{\mathcal{C}}$ but there exists a $(2^{-\Omega(n)})$-GAMD code $\mathcal{C}'$ with respect to $\mathcal{F}_{\mathcal{C}}$ with rate $r \cdot O(1) - o(1)$, where r is the rate of $\mathcal{C}$.*

Proof. Let (E, D) be a non-malleable code where $\mathsf{E} : K^k \to K^n$. We construct GAMD code $(\mathcal{E}, \mathcal{D})$ and family of tampering functions $\mathcal{F}_{\mathcal{C}}$ as follows. Let $\mathbf{1} = 0^{k-1} \| 1 \in K^k$ and let $F_{\mathsf{E},\mathsf{D}}$ be the function $F_{\mathsf{E},\mathsf{D}}(c) = \mathsf{E}(\mathsf{D}(c) + \mathbf{1})$ if $\mathsf{D}(c) \neq \perp$ otherwise c. Define $\mathcal{F}_{\mathcal{C}} = \{F_{\mathsf{E},\mathsf{D}}\}$ and let $(\mathcal{E}', \mathcal{D}')$ with $\mathcal{E}' : \mathcal{S} \to K^k$ be an ϵ'-GAMD with respect to the family $\mathcal{F}_{\mathsf{add}}$ of point addition functions on K^k. Define $\mathcal{E} : \mathcal{S} \to K^k$, $\mathcal{D} : K^k \to \mathcal{S}$ by $\mathcal{E}(s) = \mathsf{E}(\mathcal{E}'(s))$,

$\mathcal{D}(c) = \mathcal{D}'(\mathsf{D}(c))$ if $\mathsf{D}(c) \neq \perp$ otherwise $\perp$. We claim that $(\mathcal{E}, \mathcal{D})$ is an ϵ'-GAMD for $\mathcal{F}_{\mathcal{C}}$. We have

$$\Pr_{\mathsf{E},\mathcal{E}'}[\mathcal{D}(\mathcal{E}(s)) = s] = \Pr_{\mathsf{E},\mathcal{E}'}[\mathcal{D}'(\mathcal{E}'(s))) = s] = 1$$

by the correctness of non-malleable (E, D) and ϵ'-GAMD $(\mathcal{E}', \mathcal{D}')$ respectively. On the other hand for any $s \neq s'$ in $\mathcal{S}$, $\Pr_{\mathsf{E},\mathcal{E}'}[\mathcal{D}(F(\mathcal{E}(s))) = s'] = \Pr_{\mathsf{E},\mathcal{E}'}[\mathcal{D}'(\mathsf{D}(F(\mathcal{E}(s)))) = s'] \leq \Pr_{\mathsf{E},\mathcal{E}'}[\mathcal{D}'(\mathsf{D}(\mathsf{E}((\mathcal{E}(s) + \mathbf{1})))) = s'] \leq \Pr_{\mathsf{E},\mathcal{E}'}[\mathcal{D}'((\mathcal{E}(s) + \mathbf{1})) = s'] \leq \epsilon'$ as $(\mathcal{E}', \mathcal{D}')$ detects tampering by point additions on K^k with probability at least $1 - \epsilon'$. The rate of $(\mathcal{E}, \mathcal{D})$ is $r \cdot (O(1)) - o(1)$ in view of Lemma 14 and Theorem 27. □

Theorem 26. *In the two-state model there exists, for any $0 < \epsilon < 1$, an explicit ϵ-super non malleable code for the class $\mathcal{F}_{\mathcal{P}_{\leq d}}$ with negligible sampling error. The rate is $\frac{1}{\Theta(d^2)}$.*

Proof. Consider ϵ_1-non-malleable $(\mathsf{E}_1, \mathsf{D}_1)$ mapping k_1 bits to n_1 bits and ϵ_2 super non-malleable $(\mathsf{E}_2, \mathsf{D}_2)$ mapping k_2 bits to n_2 bits using t-wise independent hash functions. Let $(\mathsf{Enc}, \mathsf{Dec})$ be given by $\mathsf{Enc}(s) = (\mathsf{E}_1(h_1 \| h_2), \mathsf{E}_2(s; h_1, h_2))$ where $h_1 \in_R \mathcal{H}_1, h_2 \in_R \mathcal{H}_2$ and $\mathsf{Dec}(c_1, c_2) = \mathsf{D}_2(c_2, \mathsf{D}_1(c_1))$ if $\mathsf{D}_1(c_1) \neq \perp$ otherwise $\perp$. Let $\epsilon = \epsilon_1 + \epsilon_2$. Then ϵ-super non-malleability will follow if we can prove that for any non-trivial f_1, the probability, over a random choice of h_1, h_2 that f_1 mauls c_1 to an equivalent but distinct codeword c_1' is at most ϵ_1. Let $m = h_1 \| h_2$. We have

$$\begin{aligned}
&\Pr[\mathsf{Dec}(c_1') = \mathsf{Dec}(c_1) \wedge c_1' \neq c_1 \mid c_1 \leftarrow \mathsf{Enc}(m)] \\
&= \Pr[f_1(c_1) \in S_m \wedge f_1(c_1) \neq c_1 \mid c_1 \in S_m] \\
&= \sum_m p_m \cdot \frac{|\{c_1 \mid c_1 \in S_m, f_1(c_1) \in S_m, c_1 \neq f_1(c_1)\}|}{|S_m|} \\
&\leq \sum_m p_m \cdot \frac{|S_m \cap f_1(S_m)|}{|S_m|} \leq \sum_m p_m \cdot \epsilon_1 \leq \epsilon_1
\end{aligned}$$

We now analyse the rate of Enc. We have $|\mathcal{F}_{\mathcal{P}_{\leq d}}| = p^d = 2^{nd}$. So by letting $t > n(d+1) + O(1), v_1 > 3\log_2 n + O(1), v_2 = v_1 + O(1)$, we have $k_2 = n_2 - 6\log_2 n_2 - O(1)$. The hash functions h_1, h_2 require $t(v_1 + v_2)$ bits hence $k_1 \geq 6n_2(d+1)\log_2 n_2$. Therefore $n_1 > 6\kappa d^2 n_1 (d+1)\log_2 n_2$ for some constant κ. We have $\rho = \frac{k_1+k_2}{n_1+n_2} = \frac{\Theta(n_2 d)}{\Theta(n_2 d^3)} = \frac{1}{\Theta(d^2)}$. □

6 Conclusion

We have defined a generalisation of algebraic manipulation detection codes to facilitate detection of tampering by algebraic functions over a field. We have demonstrated explicit constructions of these codes for the families of point additions and polynomial functions and randomised constructions for some broader classes over finite fields. In future work it would be interesting to extend these constructions as well as to investigate applications of these codes.

Acknowledgements The author would like to thank anonymous reviewers and Chaitanya Rao for helpful comments and suggestions.

References

[ADL14] Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing*, STOC '14, pages 774–783, New York, NY, USA, 2014. ACM.

[Agg15] Divesh Aggarwal. Affine-evasive sets modulo a prime. *Inf. Process. Lett.*, 115(2):382–385, February 2015.

[CD06] Charles J. Colbourn and Jeffrey H. Dinitz. *Handbook of Combinatorial Designs, Second Edition (Discrete Mathematics and Its Applications)*. Chapman & Hall/CRC, 2006.

[CDF+08] Ronald Cramer, Yevgeniy Dodis, Serge Fehr, Carles Padró, and Daniel Wichs. Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. In Nigel Smart, editor, *Advances in Cryptology – EUROCRYPT 2008: 27th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Istanbul, Turkey, April 13-17, 2008. Proceedings*, pages 471–488, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[CDN15] Ronald Cramer, Ivan Bjerre Damgård, and Jesper Buus Nielsen. *Secure Multiparty Computation and Secret Sharing*. Cambridge University Press, New York, NY, USA, 1st edition, 2015.

[CPS02] Sergio Cabello, Carles Padró, and Germán Sáez. Secret sharing schemes with detection of cheaters for a general access structure. *Des. Codes Cryptography*, 25(2):175–188, February 2002.

[CZ14] E. Chattopadhyay and D. Zuckerman. Non-malleable codes against constant split-state tampering. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 306–315, Oct 2014.

[DKO13] Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Non-malleable codes from two-source extractors. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology – CRYPTO 2013*, pages 239–257, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[DKRS06] Yevgeniy Dodis, Jonathan Katz, Leonid Reyzin, and Adam Smith. Robust fuzzy extractors and authenticated key agreement from close secrets. In *Proceedings of the 26th Annual International Conference on Advances in Cryptology*, CRYPTO'06, pages 232–250, Berlin, Heidelberg, 2006. Springer-Verlag.

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

[FMNV14] Sebastian Faust, Pratyay Mukherjee, Jesper Buus Nielsen, and Daniele Venturi. Continuous non-malleable codes. In Yehuda Lindell, editor, *Theory of Cryptography*, pages 465–488, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[FMVW14] Sebastian Faust, Pratyay Mukherjee, Daniele Venturi, and Daniel Wichs. Efficient non-malleable codes and key-derivation for poly-size tampering circuits. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology – EUROCRYPT 2014*, pages 111–128, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[KLT16] Aggelos Kiayias, Feng-Hao Liu, and Yiannis Tselekounis. Practical non-malleable codes from l-more extractable hash functions. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 1317–1328, New York, NY, USA, 2016. ACM.

[LL12] Feng-Hao Liu and Anna Lysyanskaya. Tamper and leakage resilience in the split-state model. In Reihaneh Safavi-Naini and Ran Canetti, editors, *Advances in Cryptology – CRYPTO 2012*, pages 517–532, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[PR15] Christos Pelekis and Jan Ramon. Hoeffding's inequality for sums of weakly dependent random variables, 2015. `https://arxiv.org/abs/1507.06871`.

[Ros94] H. E. Rose. *A Course in Number Theory, Second Edition*. Oxford University Press, 1994.

[Rus98] Imre Rusza. An infinite sidon sequence. *J. Number Theory*, 68(1):63–71, 1998.

[Sin38] James Singer. A theorem in finite projective geometry and some applications to number theory. *Transactions of the American Mathematical Society*, 43(3):377–385, 1938.

[Sti90] D. R. Stinson. The combinatorics of authentication and secrecy codes. *J. Cryptol.*, 2(1):23–49, January 1990.

[Sti94] D. R. Stinson. Universal hashing and authentication codes. *Designs, Codes and Cryptography*, 4(3):369–380, Jul 1994.

A Proof of Auxiliary Results

Proof of Lemma 17

Proof. Since $\gamma \leq 1-(1-\gamma')^m$ it suffices to prove that $\lim_{\delta\to\infty} \frac{|S\cap F(S)\cap B_d(0,\delta)|}{|S|} \leq (1-\gamma')^m$ for each choice of $F \in \mathcal{F}_{\mathsf{add}}$ over K^m. Therefore we need to show that for each $\epsilon > 0$ there exists $\delta_\epsilon > 0$ so that for all $F \in \mathcal{F}_{\mathsf{add}}$,

$$|S \cap F(S) \cap B_d(0,\delta)| \in |S \cap B_d(0,\delta)| \cdot ((1-\gamma')^m \pm \epsilon). \tag{6}$$

Decompose F as $\prod_{i=1}^m F_i$ where $F_i \in \mathcal{F}_{\mathsf{add}}$ acts on the i^{th} copy of K in K^m. Let $\epsilon' = (1-\gamma')\ln(1+\epsilon)m^{-1}$. Let $\delta'_{\epsilon'}$ be such that $\forall \delta' > \delta'_{\epsilon'}$, $|S'\cap F_i(S')\cap B_{d'}(0,\delta')| \in |S\cap B_{d'}(0,\delta')|\cdot((1-\gamma')\pm\epsilon')$. Then $\prod_{i=1}^m |S'\cap F_i(S')\cap B_{d'}(0,\delta')| \in \prod_{i=1}^m (|S'\cap B_{d'}(0,\delta)|\cdot((1-\gamma')\pm\epsilon'))$. Let $S = S'^m$ and $d = d'^m$ be the supremum metric on K^m. Then $(\prod_{i=1}^m |S'\cap B_{d'}(0,\delta')|\cdot(1-\gamma'+\epsilon')^m) \leq |\prod_{i=1}^m S'\cap\prod_{i=1}^m F_i(S')\cap\prod_{i=1}^m B_{d'}(0,\delta')| \leq (\prod_{i=1}^m |S'\cap B_{d'}(0,\delta')|)\cdot(1-\gamma'+\epsilon')^m)$. Now $((1-\gamma')-\epsilon')^m = (1-\gamma')^m(1-\epsilon'(1-\gamma')^{-1})^m \geq (1-\gamma')^m e^{-m\epsilon'/(1-\gamma')} \geq (1-\gamma')^m(1+\epsilon)^{-1}$. Similarly one can prove $(1-\gamma'+\epsilon')^m \leq (1-\gamma')^m(1+\epsilon)$. Thus taking $\delta_\epsilon = \delta'_{\epsilon'}$ shows that Equation 6 holds for each choice of ϵ and F in $\mathcal{F}_{\mathsf{add}}$ over K^m. To complete the proof, observe that the rate of $\mathcal{E}$ is $\lim_{\delta\to\infty}\frac{\log_2|S\cap B_d(0,\delta)|}{\log_2|B_d(0,\delta)|} = \lim_{\delta\to\infty}\frac{\log_2(\prod_{i=1}^m|S'\cap B_{d'}(0,\delta)|)}{\log_2(\prod_{i=1}^m|S'\cap B_{d'}(0,\delta)|)} \leq \lim_{\delta\to\infty}\frac{\log_2|S'\cap B_{d'}(0,\delta)|}{\log_2|B_{d'}(0,\delta)} = \rho'$. □

Proof of Theorem 20

Proof. By Theorem 19 we know that there exists a set $S \subset \mathbb{F}_p$ with the property that $|S| \leq (\log p \cdot p^{\frac{2}{d^2+5d+2}-1})\cdot p$ and $|S\cap F(S)| \leq \frac{\log p\cdot p^{\frac{-2}{d^2+5d+2}}}{2d}\cdot|S|$ for all $F \in \mathcal{F}_{\mathcal{P}\leq d}$. Consider partitioning S into sets $(S_m)_m$ of equal size $\frac{|S|}{T}$. Define $\mathsf{Enc} : \mathbb{Z}_T \to \mathbb{F}_p$ by $\mathsf{Enc}(m) = c : c \in_R \mathbb{Z}_m$ and $\mathsf{Dec}(c) = m : c \in S_m$. Fix $F \in \mathcal{F}_{\mathcal{P}\leq d}$ and define simulation experiment Sim_m^F as in Figure 1. Note that distribution D_F satisfies $\Pr[D_F = \mathsf{same}^*] = \Pr_{c\in_R\mathbb{F}_p}[F(c) = c]$ and $\Pr[D_F = m] = \Pr_{c\in_R\mathbb{F}_p}[F(c) \neq c \cap \mathsf{Dec}(F(c)) = m] : m \in \mathbb{Z}_T \cup \{\perp\}$. We claim that $\mathsf{SD}(\mathsf{Sim}_m^F, \mathsf{Tamper}_m^F) \leq \epsilon$ where Tamper_m^F is the tampering experiment of Definition 1. First suppose that $F(x) \equiv x$. In that case $\Pr[\mathsf{Tamper}_m^F = m] = \Pr[\mathsf{Sim}_m^F = m] = 1$ so that $\mathsf{SD}(\mathsf{Sim}_m^F, \mathsf{Tamper}_m^F) = 0$. Suppose $F(x) \equiv a$ where a is a constant in $\mathbb{F}_p$. Then $\Pr[\mathsf{Tamper}_m^F = \mathsf{Dec}(a)] = \Pr[\mathsf{Sim}_m^F = \mathsf{Dec}(a)] = 1$ so again $\mathsf{SD}(\mathsf{Sim}_m^F, \mathsf{Tamper}_m^F) = 0$. If $F \notin \{\mathrm{id}., \mathbb{F}_p\}$, then $\Pr_{c\in_R\mathbb{F}_p}[F(c) = c]$ occurs with probability at most $\frac{d}{p}$ by Lemma 8. Thus $\mathsf{SD}(\mathsf{Sim}_m^F, \mathsf{Dec}(F(c)) : c \in_R \mathbb{F}_p) \leq \frac{d}{p}$. Now

$$\begin{aligned}
&\mathsf{SD}(\mathsf{Tamper}_m^F, \mathsf{Dec}(F(c)) : c \in_R \mathbb{F}_p)\\
&= \sum_{m'} |\Pr[\mathsf{Dec}(F(c)) = m' : c \leftarrow \mathsf{Enc}(m)] - \Pr[\mathsf{Dec}(F((c)) = m' : c \in_R \mathbb{F}_p]|\\
&\leq \sum_{m'} |\Pr[\mathsf{Dec}(F(c)) = m' : c \leftarrow \mathsf{Enc}(m)]| + \sum_{m} |\Pr[\mathsf{Dec}(F(c)) = m' : c \in_R \mathbb{F}_p]|\\
&\leq \Pr[F(c) \in \bigcup_{m'\in\mathbb{Z}_T} S_{m'} : c \in_R S_m] + \Pr[F(c) \in \bigcup_{m'\in\mathbb{Z}_T} S_{m'} : c \in_R \mathbb{F}_p]\\
&\leq \frac{|S\cap F(S_m)|}{|S_m|} + \frac{|S\cap\mathbb{F}_p|}{|\mathbb{F}_p|} \leq \epsilon
\end{aligned} \tag{7}$$

To satisfy Equation 7 we need $\log p \cdot p^{\frac{1}{\Theta(d^2)}}\cdot(\frac{T}{\Theta(d)} + \frac{1}{p}) + \frac{d}{p} < \epsilon$ so that for some constant c it holds $p > (\frac{T}{\epsilon})^{c\cdot d^2}$ yielding the result. □

Sim_m^F :

1. Pick $c \in_R \mathbb{F}_p \mid c \in S_m$.
2. Output same^* if $F(c) = c$ else $\mathsf{Dec}(F(c))$.

Figure 1: Tampering simulation experiment.

B Degree-k Algebraic Extension

Rusza proved that there exists an infinite sequence of integers whose asymptotic density in any interval $[0, \epsilon)$ is at least $\epsilon^{0.41-o(1)}$ [Rus98]. To construct a GAMD for the class of point additions over the rationals we will only need the weaker result that for any prime power M there exists a Sidon set (integer 1-difference set) of size $M+1$ inside $\mathbb{Z}_q = \{1, \ldots, q\}$ where $q = M^2 + M + 1$ [Sin38]. We denote this set $\mathcal{D}_M$ and consider $q(\cdot)$ as function in M.

Theorem 27. *There exists an explicit weak ϵ-GAMD over the rationals against the class of point additions with rate $0.75 - o(1)$ and negligible error probability ϵ.*

Proof. Let $N > 0$ be an arbitrary integer. Let $r(N)$ be the largest prime such that $r^2+r+1 \leq N$. Let $S \subset \mathbb{Q}$ be given by

$$S := \{\frac{a}{p} |\ p \text{ prime}, a \in \mathcal{D}_{r(\lfloor \frac{p}{2} \rfloor)}\} \tag{8}$$

We prove that for any element $F \in \mathcal{F}_{\mathsf{add}}$, $|S \cap F(S)| \leq 1$. Suppose for contradiction that there exist $v_1, v_2, v_3, v_4 \in S$ such that $v_1 - v_2 = v_3 - v_4$. Let $v_1 = \frac{a}{p}, v_2 = \frac{b}{q}, v_3 = \frac{c}{r}, v_4 = \frac{d}{s}$ where $a < \frac{p}{2}, b < \frac{q}{2}, c < \frac{r}{2}, d < \frac{s}{2}$.

Case 1: $p \neq q \neq r \neq s$**.** We have $(aq-bp)rs = (cs-dr)pq$. Then $pq|(aq-bp)$ and $aq-bp \neq 0$ as $a < p$. One the other hand $|aq - bp| < \max\{aq, bp\} < \frac{pq}{2}$ which is a contradiction.

Case 2: At least two, not all p, q, r, s **distinct.** WLOG $p \neq r$ and $q \neq s$. Then either $p = s$ or $q = r$. If $p = s$, $\frac{a}{p} - \frac{b}{q} = \frac{c}{r} - \frac{d}{p}$ so that $(a + d)rq = p(br + cq)$. Then $r \mid cpq$. As $p \neq r$ and $c < r$, $q = r$. Thus $p \mid a + d$ which contradicts $a, d < \frac{p}{2}$. The case $q = r$ is similar.

Case 3: $p = q = r = s$**.** In this case $a - b = c - d$ with $a \neq c$ and $b \neq d$, which contradicts $\mathcal{D}_{r(p)}$ being a 1-difference set.

We now analyse the rate of $\mathcal{E}$. We have $\rho = \lim_{N\to\infty} \frac{\log_2 \#\{x \in S : x = \frac{a}{N} : a \leq N\}}{\log_2 \#\{x \in \mathbb{Q} : x = \frac{a}{N} : a \leq N\}}$. By Lemma 9 for sufficiently large N there are at least $\frac{N}{\ln N} - 1.5\frac{(N/2)}{\ln(N/2)}$ primes in the interval $[N/2, N]$. We may also choose prime M so that $q(M) \approx \lfloor \frac{(N/2)}{2} \rfloor$. Thus S contains at least $\sqrt{(\lfloor \frac{N}{4} \rfloor)} \cdot (\frac{N}{\ln N} - \frac{3N}{4\ln(N/2)}) = O(\frac{N^{3/2}}{\ln N})$ elements whose denominator is at most N. Thus $\rho = \lim_{N\to\infty} \frac{1.5 \ln N - \ln\ln N}{\ln(N^2/2)} = 0.75$. □

Combining Theorem 27, Lemma 17 and Lemma 22 we have

Corollary 28. *Let K be a number field of index $k := [K : \mathbb{Q}]$. Then there exists a ϵ-GAMD for the class $\mathcal{F}_{\mathsf{add}}$ over K with rate $1 - o(1)$ and negligible ϵ for any choice of k at most polynomial in the message length.*

C Shifted Polynomial Collections with Bounded Root Set Size

Definition 8. *Let K be a finite field and $I = \{F_i\}_{1 \leq i \leq m}$ be a collection of polynomials in $K[x_1, \ldots, x_m]$. For $\mathbf{a} \in K^m$ let $I_{\mathbf{a}} = \{F_i - a_i\}_{1 \leq i \leq m}$ where $\mathbf{a} = (a_1, \ldots, a_m)$. Define $s(I) = \max_{\mathbf{a}} |\{x \in K^m \mid F(\mathbf{x}) = 0 \ \forall F \in I_{\mathbf{a}}\}|$.*

Theorem 29. *Let $\mathcal{I}$ be a collection of ideals for which each $I \in \mathcal{I}$ is as in Definition 8. Consider the map $\varphi_I : K^m \to K^m$ given by $\varphi_I(\mathbf{x}) = (F_1(\mathbf{x}), \ldots, F_m(\mathbf{x}))$. Let c be a positive integer and let $\mathcal{F}_{c,\mathcal{I}} = \{\varphi_I \mid I \in \mathcal{I}, s(I) \leq c\}$. Suppose that $|\mathcal{F}_{c,\mathcal{I}}| < n^\tau$, $\tau > 0$. Then there exists a weak $2e(-(c+1))$-GAMD for the class $\mathcal{F}_{c,\mathcal{I}}$ with rate $1 - o(1)$. The sampling error is $e(-\Omega(|K|))$.*

Proof. Analogous to Lemma 15, consider the set S defined by sampling each element of K^m with probability γ. For each $\mathbf{x} \in K^m$ let $\beta_{\mathbf{x}}$ be an indicator variable which is equal to 1 iff the events $\mathbf{x} \in S$ and $\varphi_I(\mathbf{x}) \in S$ both occur. We claim that the dependency graph associated with $\{\beta_{\mathbf{x}}\}_{\mathbf{x}}$ satisfies $\alpha(G) > \frac{n}{2(c+1)}$. To see this, note that $\beta_{\mathbf{x}}$ is independent of $\beta_{\mathbf{y}}$ unless one of i) $\mathbf{y} = \varphi_I(\mathbf{x})$ ii) $\varphi_I(\mathbf{x}) = \varphi_I(\mathbf{y})$ iii) $\mathbf{x} = \varphi_I(\mathbf{y})$ occurs. As φ_I is well-defined, clearly there is at most one solution to the first case. For the second case, suppose for contradiction that there are $c+1$ values $\mathbf{y}, \mathbf{x}_1, \ldots, \mathbf{x}_c$ such that $\varphi_I(\mathbf{y}) = \varphi_I(\mathbf{x}_1) = \ldots = \varphi_I(\mathbf{x}_c)$. This means that $I_{\varphi(\mathbf{y})}$ has $c+1$ solutions $\mathbf{y}, \mathbf{x}_1, \ldots, \mathbf{x}_c$, which by the definition of $s(\cdot)$ is impossible. A similar argument shows that there are at most c solutions to the third case. Thus $V(G)$ can be partitioned into distinct sets $V_1, \ldots, V_{n'}$ such that $|V_i| = 2(c+1)$ and V_i possesses a representative $\mathbf{y}_i$ such that $\beta_{\mathbf{y}_i}$ is independent of $\beta_{\mathbf{y}_j} : i \neq j$. Let N_0 and N_1 be the sizes of the sets S and $S \cap \varphi_I(S)$ for a fixed $\varphi_I \in \mathcal{F}_{c,\mathcal{I}}$. Let $0 < \epsilon < 1$ be a parameter satisfying $\gamma < \nu(1-\epsilon) < 2\gamma$. Applying Lemma 10 and Lemma 12 we have

$$\Pr[N_0 \in (-\infty, \mu_0(1-\epsilon)] \cup [\mu_0(1+\epsilon), \infty)] \leq e(-\frac{\epsilon^2 \mu_0}{3}) + e(-\frac{\epsilon^2 \mu_0}{2})$$

$$\Pr[N_1 \in [\mu_1(1+\epsilon), \infty)] \leq (\gamma^2)^\alpha \cdot e(-\frac{n(\nu(1-\epsilon)-\gamma)^2}{2\gamma^2}) \cdot 2^n$$

Let $\gamma \leq e(-(c+1))$ so that $\gamma^{2\alpha} < e(-n)$ and $\nu(1-\epsilon) = \frac{3\gamma}{2}$. Then a ν-GAMD exists for $\nu = \frac{3}{2}e(-(c+1) - o(c))$ for some $\epsilon = e(o(c))$. The sampling error is $e(-\frac{\epsilon^2 n\gamma}{6}) + e(-\frac{\epsilon^2 n\gamma}{4}) + n^\tau \cdot (\gamma^{2\alpha}) \cdot e(-\frac{n}{8}) \cdot e(n \ln 2) = e(-\Omega(|K|))$. The rate is $\frac{\log_2 \gamma n}{\log_2 n}$, which for approximately constant γ is $1 - o(1)$. □

As a consequence we have the following randomised analogue to Theorem 19.

Corollary 30. *There exists a weak* $2e(-(d+1))$*-GAMD for the class* $\mathcal{F}_{\mathcal{P}_{\leq d}}$ *with asymptotically optimal rate and negligible error probability.*