Affine-evasive Sets Modulo a Prime

Divesh Aggarwal*

October 16, 2014

Abstract

In this work, we describe a simple and efficient construction of a large subset S of $\mathbb{F}_p$, where p is a prime, such that the set $A(S)$ for any non-identity affine map A over $\mathbb{F}_p$ has small intersection with S.

Such sets, called affine-evasive sets, were defined and constructed in [ADL14] as the central step in the construction of non-malleable codes against affine tampering over $\mathbb{F}_p$, for a prime p. This was then used to obtain efficient non-malleable codes against split-state tampering.

Our result resolves one of the two main open questions in [ADL14]. It improves the rate of non-malleable codes against affine tampering over $\mathbb{F}_p$ from $\log\log p$ to a constant, and consequently the rate for non-malleable codes against split-state tampering for n-bit messages is improved from $n^6 \log^7 n$ to n^6.

*Department of Computer Science, New York University. Email: divesha@cs.nyu.edu.

1 Introduction

Non-malleable Codes (NMCs). NMCs were introduced in [DPW10] as a beautiful relaxation of error-correction and error-detection codes. Informally, given a tampering family $\mathcal{F}$, an NMC $(\mathsf{Enc}, \mathsf{Dec})$ against $\mathcal{F}$ encodes a given message m into a codeword $c \leftarrow \mathsf{Enc}(m)$ in a way that, if the adversary modifies m to $c' = f(c)$ for some $f \in \mathcal{F}$, then the message $m' = \mathsf{Dec}(c')$ is either the original message m, or a completely "unrelated value". As has been shown by the recent progress [DPW10, LL12, DKO13, ADL14, FMVW13, FMNV14, CG14a, CG14b] NMCs 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'. NMCs are useful in situations where changing x to an unrelated x' is not useful for the attacker (for example, when x is the secret key for a signature scheme.)

Split-State Model. NMCs do not exist for the class of all functions $\mathcal{F}_{\mathsf{all}}$. In particular, it does not include functions of the form $f(c) := \mathsf{Enc}(h(\mathsf{Dec}(c)))$, since $\mathsf{Dec}(f(\mathsf{Enc}(m))) = h(m)$ is clearly related to m. One of the largest and practically relevant tampering families for which we can construct NMCs is the so-called split-state tampering family where the codeword is split into two parts $c_1 \| c_2$, and the adversary is only allowed to tamper with c_1, c_2 independently to get $f_1(c_1) \| f_2(c_2)$. A lot of the aforementioned results [LL12, DKO13, ADL14, CG14b, FMNV14] have studied NMCs against split-state tampering. [ADL14] gave the first (and the only one so far) information-theoretically secure construction in the split-state model from n-bit messages to $n^7 \log^7 n$-bit codewords (i.e., code rate $n^6 \log^7 n$). The security proof of this scheme relied on an amazing property of the inner-product function modulo a prime, that was proved using results from additive combinatorics.

Affine-evasive Sets and Our Result. One of the crucial steps in the construction of [ADL14] was the construction of NMC against affine tampering modulo p. This was achieved by constructing an affine-evasive set of size $p^{1/\log\log p}$ modulo a prime p. It was asked as an open question whether there exists an affine-evasive set of size $p^{\Theta(1)}$, which will imply constant rate NMC against affine-tampering and rate n^6 NMC against split-state tampering.[1] We resolve this question in the affirmative by giving an affine-evasive set of size $\Theta(\frac{p^{1/4}}{\log p})$.

2 Explicit Construction

For any set $S \subset \mathbb{Z}$, let $aS + b = \{as + b | s \in S\}$. By $S \mod p \subseteq \mathbb{F}_p$, we denote the set of values of S modulo p.

We first define an affine-evasive set $S \subseteq \mathbb{F}_p$.

Definition 1 A non-empty set $S \subseteq \mathbb{F}_p$ is said to be (γ, ν)-*affine-evasive* if $|S| \leq \gamma p$, and for any $(a, b) \in \mathbb{F}_p^2 \setminus \{(1, 0)\}$, we have

$$|S \cap (aS + b \pmod p)| \leq \nu |S| \ .$$

[1] Under a plausible conjecture, this will imply constant rate NMC against split-state tampering. See Theorem 5 for more details.

Now we give a construction of an affine-evasive set.

Let $Q := \{q_1, \ldots, q_t\}$ be the set of all primes less than $\frac{1}{2}p^{1/4}$. Define $S \subset \mathbb{F}_p$ as follows:

$$S := \left\{ \frac{1}{q_i} \pmod p \mid i \in [t] \right\} . \tag{1}$$

Thus, S has size $\Theta(\frac{p^{1/4}}{\log p})$ by the prime number theorem.

Theorem 1 *For any prime p, the set S defined in Equation (1) is $(\frac{1}{2}p^{-3/4}, O(p^{-1/4} \cdot \log p))$-affine-evasive.*

Proof. Clearly,

$$|S| = t \leq \frac{1}{2}p^{1/4} = \frac{1}{2}p^{-3/4} \cdot p .$$

Fix $a, b \in \mathbb{F}_p$, such that $(a, b) \neq (1, 0)$. Now, we show that $|S \cap (aS + b \pmod p)| \leq 3$. Assume, on the contrary, that there exist distinct $\alpha_i \in Q$ for $i \in \{0, 1, 2, 3\}$ such that $1/\alpha_i \pmod p \in S \cap (aS + b \pmod p))$. We have

$$\frac{a}{\beta_i} + b = \frac{1}{\alpha_i} \pmod p \text{ for } i = 0, 1, 2, 3 , \tag{2}$$

where $\beta_i, \alpha_i \in Q$ for $i \in \{0, 1, 2, 3\}$, and $\alpha_i \neq \alpha_j$ for any $i \neq j$.

For any i, if $\beta_i = \alpha_i$, then $b \cdot \beta_i = 1 - a \mod p$, which has at most one solution (since we assume $(a, b) \neq (1, 0)$). Thus, without loss of generality, we assume that $\beta_i \neq \alpha_i$, for $i \in \{1, 2, 3\}$, and $\beta_1 < \beta_2 < \beta_3$.

From Equation (2), we have that

$$\frac{\frac{a}{\beta_1} + b - \frac{a}{\beta_2} - b}{\frac{a}{\beta_1} + b - \frac{a}{\beta_3} - b} = \frac{\frac{1}{\alpha_1} - \frac{1}{\alpha_2}}{\frac{1}{\alpha_1} - \frac{1}{\alpha_3}} \pmod p ,$$

which on simplification implies

$$(\alpha_3 - \alpha_1)(\beta_2 - \beta_1)\beta_3\alpha_2 = (\alpha_2 - \alpha_1)(\beta_3 - \beta_1)\beta_2\alpha_3 \pmod p .$$

Note that both the left-hand and right-hand side of the above equation takes values between $\frac{-p}{16}$ and $\frac{p}{16}$, and hence the equality holds in $\mathbb{Z}$ (and not just in $\mathbb{Z}_p$).

$$(\alpha_3 - \alpha_1)(\beta_2 - \beta_1)\beta_3\alpha_2 = (\alpha_2 - \alpha_1)(\beta_3 - \beta_1)\beta_2\alpha_3 . \tag{3}$$

By equation 3, we have that β_3 divides $(\alpha_2 - \alpha_1)(\beta_3 - \beta_1)\beta_2\alpha_3$. Clearly, β_3 is relatively prime to α_3, β_2, and $\beta_3 - \beta_1$. Therefore, β_3 divides $(\alpha_2 - \alpha_1)$. This implies

$$\beta_3 \leq |\alpha_2 - \alpha_1| . \tag{4}$$

Also, from equation 3, we have that α_2 divides $(\alpha_2 - \alpha_1)(\beta_3 - \beta_1)\beta_2\alpha_3$, which by similar reasoning implies α_2 divides $\beta_3 - \beta_1$. Thus, using that $\beta_3 > \beta_1$,

$$0 < \alpha_2 \leq \beta_3 - \beta_1 < \beta_3 . \tag{5}$$

Similarly, we can obtain α_1 divides $\beta_3 - \beta_2$, which implies

$$0 < \alpha_1 \leq \beta_3 - \beta_2 < \beta_3 . \tag{6}$$

Equation (5) and (6) together imply that $|\alpha_2 - \alpha_1| < \beta_3$, which contradicts Equation (4). □

3 Affine-evasive function and Efficient NMCs

Affine-evasive function. We recall here the definition of affine-evasive functions from [ADL14]. Affine-evasive functions immediately give efficient construction of NMCs against affine-tampering.

Definition 2 *A surjective function $h : \mathbb{F}_p \mapsto \mathcal{M} \cup \{\perp\}$ is called* (γ, δ)-affine-evasive *if for any $a, b \in \mathbb{F}_p$ such that $a \neq 0$, and $(a, b) \neq (1, 0)$, and for any $m \in \mathcal{M}$,*

1. $\Pr_{U \leftarrow \mathbb{F}_p}(h(aU + b) \neq \perp) \leq \gamma$
2. $\Pr_{U \leftarrow \mathbb{F}_p}(h(aU + b) \neq \perp \mid h(U) = m) \leq \delta$
3. *A uniformly random X such that $h(X) = m$ is efficiently samplable.*

We now mention a result that shows that we can construct an affine-evasive function from an affine-evasive set S.

Lemma 1 ([ADL14, Claim 5]) *Let $S \subseteq \mathbb{F}_p$ be a (γ, ν)-affine-evasive set with $\nu \cdot K \leq 1$, and K divides $|S|$.*[2] *Furthermore, let S be ordered such that for any i, the i-th element is efficiently computable in $O(\log p)$. Then there exists a $(\gamma, \nu \cdot K)$-affine-evasive function $h : \mathbb{F}_p \mapsto \mathcal{M} \cup \{\perp\}$.*

Note that the above result requires that for any i, the i-th element of S is efficiently computable for some ordering of the set S. This is not possible for our construction since for our construction this would mean efficiently sampling the i-th largest prime. However, this requirement was made just to make sure that h^{-1} is efficiently samplable. We circumvent this problem by giving a slightly modified definition of the affine-evasive function h in the proof of Lemma 2. Before proving this, we state the following result that we will need.

Theorem 2 ([HB88]) *For any $n \in \mathbb{N}$, and any $n' \leq n$ such that $n'^{12/7} \geq n$,*

$$\pi(n) - \pi(n - n') = \Theta\left(\frac{n'}{\log n}\right) ,$$

where $\pi(n)$ denote the number of primes less than n.

Lemma 2 *Let $\mathcal{M}$ be a finite set such that $|\mathcal{M}| \geq 2$, and let $p \geq |\mathcal{M}|^{16}$ be a prime. There exists an efficiently computable $(p^{-3/4}, O(|\mathcal{M}| \log p \cdot p^{-1/4}))$-affine-evasive function $h : \mathbb{F}_p \mapsto \mathcal{M} \cup \{\perp\}$.*

Proof. Without loss of generality, let $\mathcal{M} = \{1, \ldots, K\}$, for some integer K. Let $S \subseteq \mathbb{F}_p$ be as defined in Section 2. Define $S_1, \ldots, S_K$ to be a partition of S as follows.

$$S_i := \left\{ s \in S \;\middle|\; \frac{1}{s} \in \left[\frac{i-1}{2K} p^{1/4} \,,\, \frac{i}{2K} p^{1/4}\right) \right\} . \tag{7}$$

Now let $n_i = \frac{p^{1/4} i}{2K}$ and $n' = \frac{p^{1/4}}{2K}$. By the construction of S, $|S_i| = \pi(n_i) - \pi(n_i - n')$. We will bound $|S_i|$ for all $i \in [K]$ using Theorem 2. To do this, we need to verify that for all i, $n'^{12/7} \geq n_i$. Since $n_i < n_j$ for all $i < j$, it is sufficient to show this for $i = K$, i.e., $n_i = \frac{p^{1/4}}{2}$.

$$\frac{n'^{12/7}}{n_K} = \frac{2p^{3/7}}{(2K)^{12/7} p^{1/4}} = \frac{p^{5/28}}{2^{5/7} \cdot K^{12/7}} \geq \frac{K^{5 \cdot 16/28}}{2^{5/7} \cdot K^{12/7}} = \frac{K^{8/7}}{2^{5/7}} > 1 ,$$

[2] The assumption K divides $|S|$ is just for simplicity.

where we used the fact that $p \geq K^{16}$, and $K \geq 2$. Also note that n_i is upper bounded by $\frac{p^{1/4}}{2}$, and hence $\log n_i = O(\log p)$. Thus, using Theorem 2, we get that each S_i has size at least $\Theta(\frac{p^{1/4}}{K \log p})$.

Let $h : \mathbb{F}_p \mapsto \mathcal{M} \cup \{\perp\}$ be defined as follows:

$$h(x) = \begin{cases} i & \text{if } x \in S_i \\ \perp & \text{otherwise }. \end{cases}$$

The statement $\Pr(h(aU + b) \neq \perp) \leq p^{-3/4}$ is obvious by the definition of S, and the observation that $aU + b$ is uniform in $\mathbb{F}_p$.

Also, for any $m \in \mathcal{M}$, and for any $(a, b) \neq (1, 0)$, and $a \neq 0$,

$$\begin{aligned} \Pr(h(aU + b) \neq \perp | h(U) = m) &= \frac{\Pr(aU + b \in S \wedge U \in S_m)}{\Pr(U \in S_m)} \\ &\leq \frac{\Pr(aU + b \in S \wedge U \in S)}{|S_m|/p} \\ &= \frac{p}{|S_m|} \Pr(U \in S \cap (a^{-1}S - ba^{-1}) \pmod p) \\ &= O(K \log p \cdot p^{-1/4}) . \end{aligned}$$

Also, sampling a uniformly random X such that $h(X) = m$ is equivalent to sampling a uniformly random prime q in the interval

$$I := \Big[\frac{m-1}{2K} p^{1/4} , \frac{m}{2K} p^{1/4}\Big)$$

and computing $1/q \mod p$. Sampling q can be done in time polynomial in $\log p$ by repeatedly sampling a random element in I until we get a prime. Computing $1/q \mod p$ can be done efficiently using Extended Euclidean Algorithm. □

Note that the proof of Lemma 2 is identical to the proof of Lemma 1, except the proof that a uniformly random X such that $h(X) = m$ is efficiently samplable for any given m.

Efficient NMCs. We recall here the definition of non-malleable codes for completeness.

Definition 3 *Let $\mathcal{F}$ be some family of tampering functions. For each $f \in \mathcal{F}$, and $m \in \mathcal{M}$, define the tampering-experiment*

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

which is a random variable over the randomness of the encoding function Enc. We say that a coding scheme $(\mathsf{Enc}, \mathsf{Dec})$ is ε-non-malleable w.r.t. $\mathcal{F}$ if for each $f \in \mathcal{F}$, there exists a distribution (corresponding to the simulator) D_f over $\mathcal{M} \cup \{\perp, \mathsf{same}^*\}$, such that, for all $m \in \mathcal{M}$, we have that the statistical distance between Tamper_m^f and

$$\mathsf{Sim}_m^f := \left\{ \begin{array}{c} \tilde{m} \leftarrow D_f \\ \textit{Output: } m \textit{ if } \tilde{m} = \mathsf{same}^*, \textit{ and } \tilde{m}, \textit{ otherwise.} \end{array} \right\}$$

is at most ε. Additionally, D_f should be efficiently samplable given oracle access to $f(\cdot)$.

Using Lemma 2 and the construction of [ADL14], we get the following results.

Theorem 3 *There exists an efficient coding scheme* $(\mathsf{Enc}, \mathsf{Dec})$ *encoding* k*-bit messages to* $\Theta(k + \log(\frac{1}{\varepsilon}))$ *bit codewords that is* ε*-non malleable w.r.t. the family of affine tampering functions* $\mathcal{F}_{\mathsf{aff}}$*.*

Theorem 4 *There exists an efficient coding scheme* $(\mathsf{Enc}, \mathsf{Dec})$ *encoding* k*-bit messages to* $\Theta((k + \log(\frac{1}{\varepsilon}))^7)$ *bit codewords that is* ε*-non malleable w.r.t. the family of split-state tampering functions* $\mathcal{F}_{\mathsf{split}}$*.*

Also, assuming the following conjecture from [ADL14], our result gives the first NMC with constant rate in the split-state model.

Conjecture 1 ([ADL14, Conjecture 2]) *There exists absolute constants* $c, c' > 0$ *such that the following holds. For any finite field* $\mathbb{F}_p$ *of prime order, and any* $n > c'$*, let* $L, R \in \mathbb{F}_p^n$ *be uniform, and fix* $f, g : \mathbb{F}_p^n \to \mathbb{F}_p^n$*. Let* $\mathcal{D}$ *be the family of convex combinations of* $\{(U, aU + b) : a, b \in \mathbb{F}_p\}$ *where* $U \in \mathbb{F}_p$ *is uniform. Then there exists* $D \in \mathcal{D}$ *such that*

$$\Delta(\langle L, R\rangle, \langle f(L), g(R)\rangle \ ; \ D) \leq p^{-cn} \ .$$

Theorem 5 *Assuming Conjecture 1, there exists an efficient coding scheme* $(\mathsf{Enc}, \mathsf{Dec})$ *encoding* k*-bit messages to* $\Theta(k + \log(\frac{1}{\varepsilon}))$ *that is* ε*-non malleable w.r.t. the family of split-state tampering functions* $\mathcal{F}_{\mathsf{split}}$*.*

References

[ADL14] D. Aggarwal, Y. Dodis, and S. Lovett. Non-malleable codes from additive combinatorics. In *STOC*, 2014. To appear.

[CG14a] M. Cheraghchi and V. Guruswami. Capacity of non-malleable codes. In *Innovations in Theoretical Computer Science*. ACM, 2014. To appear.

[CG14b] M. Cheraghchi and V. Guruswami. Non-malleable coding against bit-wise and split-state tampering. In *Theory of Cryptography Conference - TCC*. Springer, 2014. To appear.

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

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

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

[FMVW13] S. Faust, P. Mukherjee, D. Venturi, and D. Wichs. Efficient non-malleable codes and key-derivation for poly-size tampering circuits. *IACR Cryptology ePrint Archive*, 2013.

[HB88] D. R. Heath-Brown. The number of primes in a short interval. *J. Reine Angew. Math.*, 389:22–63, 1988.

[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.