Algebraic Manipulation Detection Codes and Related Structures

September 2020

SHAO MINFENG

Algebraic Manipulation Detection Codes and Related Structures

Graduate School of Systems and Information Engineering

University of Tsukuba

September 2020

SHAO MINFENG

Algebraic Manipulation Detection Codes and Related Structures

Minfeng Shao

University of Tsukuba, 2020

Supervisor: Ying Miao

Nowadays, large, diverse sets of information grow at ever-increasing rates. To protect data in private storage devices from tampering by adversaries becomes more and more challenging. Algebraic manipulation detection (AMD) codes were first introduced by Cramer *et al.* [27] to prevent a special type of tampering called algebraic manipulation. They used AMD codes to convert linear secret sharing schemes into robust secret sharing schemes and build robust fuzzy extractors. Generally speaking, an AMD code consists of a probabilistic encoding map and a determined decoding function, where the encoding map encodes a source into a message such that any tampering will be detected, except with a small error probability. For AMD codes, we consider the attack model such that an adversary may manipulate the valid message without having reading access by adding some offset noise of his choice. The attack model is divided into two sub-models by distinguishing two different settings: the adversary has full knowledge of the source (the strong attack model) and the adversary has no knowledge about the source (the weak attack model).

This dissertation is devoted to study AMD codes for both strong/weak attack models and related structures such as external difference families and highly nonlinear functions.

For AMD codes under weak attack model, we first define a new type of weighted external difference families which are proved equivalent with weak AMD codes. Based on this combinatorial characterization of weak AMD codes: (1) We improve the known lower bound, i.e., the R-bound by Paterson and Stinson [81] on the maximum probability of successful tampering for the adversary's all possible strategies; (2) We derive a necessary condition for the R-bound to be achieved; (3) We determine the exact combinatorial structure for a weak AMD code with the minimum possible probability of successful tampering, when the R-bound is not achievable. In this way, some weak AMD codes which have not been identified to be R-optimal previously now can be identified to be in fact optimal. Secondly, we exhibit several explicit constructions of optimal weighted external difference families to generate AMD codes under weak attack model. At last, we also build a relationship between highly nonlinear functions and systematic weak AMD codes. By choosing special highly nonlinear functions such as perfect nonlinear functions, a few infinite classes of new R-optimal systematic weak AMD codes and new systematic weak AMD codes with asymptotically optimal tag size are constructed.

For AMD codes under strong attack model, we also establish a relationship between highly nonlinear functions and systematic strong AMD codes. On one hand, by choosing some special highly nonlinear functions, a few infinite classes of new systematic strong AMD codes with minimum possible probability of successful tampering are constructed. On the other hand, from a subclass of AMD codes with more strict assumptions, highly nonlinear functions can also be generated, where their nonlinearities are determined by the parameters of the corresponding AMD codes. Especially, we prove that the well-known construction for AMD codes in [27, Theorem 2] can also be explained by highly nonlinear functions. In addition, a combinatorial construction of G-optimal AMD codes is introduced for strong attack model.

Finally, we make a conclusion on this dissertation and list some open problems related to the topics discussed in this dissertation.

ACKNOWLEDGMENTS

It is my greatest pleasure to acknowledge the patient guide, the generous assistance, helpful discussion and encouragement provided by my supervisor: Professor Ying Miao at University of Tsukuba, who taught me knowledge about combinatorics and also how to do research in combinatorics and cryptography; who led me to the path of my degree; who guided and encouraged me overcome some of the obstacles along the path. At the end of the path, his endless support and advice are helpful during the writing of this thesis.

I appreciate Prof. Hiroyasu Ando, Prof. Tuan Phung-Duc, and Prof. Yongbing Zhang who served as the members of my advisory group, for their time, support and friendly help during my Ph.D. program. I would like to thank Prof. Hiroyasu Ando, Prof. Masahiro Hachimori, Prof. Hiroki Koga, and Prof. Yongbing Zhang for serving as members of my degree evaluation committee. Their valuable comments improve the presentation of this dissertation. I am particularly grateful to Prof. Ryoh Fuji-Hara for many interesting discussions and suggestions for my study and daily life in Tsukuba. His encouragements help me to overcome the difficult I met in research.

I would like to appreciate Prof. Heguo Liu my master supervisor at Hubei University, who led me to research in group theory. This is the base of my currently research.

I would like to thank Prof. Sanpei Kageyama, Prof. Masakazu Jimbo, Prof. Hung-Lin Fu, Prof. Yanxun Chang, Prof. Masanori Sawa, Prof. Nobuko Miyamoto, Prof. Koji Momihara, Prof. Yuan-Hsun Lo, Prof. Minquan Cheng, Prof. Jing Jiang, Dr. Yuichiro Fujiwara, Dr. Kazuki Matsubara, Dr. Shoko Chisaki, and Dr. Xiaonan Lu for their help and discussions in workshops.

Many thanks to all my laboratory members, Dr. Yujie Gu, Prof. Xiaojing Liu, Dr. Donglai Ma, Satoshi Noguchi, Jinping Fan, Hua Shuai, Wei Li, Yueting Li, Xuli Zhang, Prof. Su Wang, and Dr. Xin Wang for their discussions in weekly seminars and help in daily life.

Finally, to my family, my daughter Chenxi Cai, and my husband Han Cai, I express my appreciation of their patience and encouragements.

Contents

CHAPTER 1

Introduction

Nowadays, large, diverse sets of information grow at ever-increasing rates. To protect data in private storage devices from tampering by adversaries becomes more and more challenging. Traditionally, since the devices are private, cryptographic security notions make assumptions that adversaries only have "black-box" access (without reading the message) to an attacked storage system. This is to say that an adversary may modify the information by adding some offset message without reading access to the original data. In [27], algebraic manipulation detection (AMD) codes were introduced to encode the source data in a way to make sure that whenever an adversary tampers with the original data the system may detect the attack except with a small given error probability. More specifically, an AMD code can solve the following problem.

- Encoding of the source: A private device encodes a source message x into an element $E(x)$ of a given finite group G;

- Decoding of the message: With $E(x)$ the original source x should be deterministically decoded;

- Adversary's ability: An adversary is capable to modify $E(x)$ to $E(x) + \Delta$ without reading $E(x)$;

- The goal: Whenever an adversary tampers with the original data the system may detect the attack except with a small error probability ρ.

AMD codes are an abstraction of several known methods for cheating detection in linear secret sharing schemes [12, 75–77, 80, 89]. Following the ideas of [42], Cramer et al. [27] also showed AMD codes can be used to design robust fuzzy extractors. Several other applications of AMD codes were found subsequently. For example, to unconditionally secure multiparty computation with dishonest majority [7], anonymous quantum communication [6], non-malleable codes [43], codes for computation simple channels [52], public key encryption against related key attacks [94], and secure memories [50, 67, 93]. Hereafter, we briefly describe the two original applications of AMD codes in [27].

1.1 Application scenarios for AMD codes: robust secret sharing schemes

One explicit application of AMD codes is to convert a linear secret sharing scheme into a robust secret sharing scheme. Usually, for a secret sharing scheme， we request the secret be shared among different users in such a way that if an authorized subset of users pool their shares, then they can decode the original secret, while if an unauthorized subset of users pool their shares, then they can determine nothing about the original secret. More specifically, we need a pair of functions $(Share, Recover)$, such that the function $Share$ maps the secret s into a set of shares $\mathcal{S} = Share(s)$ and

- any authorized subset $\widetilde{\mathcal{S}} \subseteq \mathcal{S}$ can reconstruct the original secret, i.e.,
$$Recover(\widetilde{\mathcal{S}}) = s;$$
- any unauthorized set of shares give absolutely no information about the original secret.

However, the system may suffer from tampering by some adversaries. We also require the system to notice the adversaries whenever there is any share being tampered. That is, we need the so-called robust secret sharing scheme [12], a pair of functions $(Share, Recover)$, such that the function $Share$ maps the secret s into a set of shares $\mathcal{S} = Share(s)$ and

- any authorized subset $\widetilde{\mathcal{S}} \subseteq \mathcal{S}$ can reconstruct the original secret, i.e.,
$$Recover(\widetilde{\mathcal{S}}) = s;$$
- any unauthorized set of shares gives absolutely no information about the o-riginal secret;
- whenever the system tries to recover the original secret with authorized set of shares that suffers from tampering, the system should be able to detect this except with a small error probability ρ, i.e., $\Pr(Recover(\widetilde{\mathcal{S}}+\Delta) \notin \{s, \perp\} \leq \rho)$, where $\widetilde{\mathcal{S}}+\Delta$ denotes shares being tampered with and $\perp$ denotes invalid secret.

In [27], AMD codes are included to convert a linear secret sharing scheme, i.e., $Share(s + \Delta) = Share(s) + Share(\Delta)$ into a robust secret sharing scheme. The main idea is that, before sharing the secret s, the system encodes the secret s using an AMD code and then shares the encoded message $E(s)$ using the linear secret sharing scheme. We use the well-known Shamir threshold scheme [86] to explain the main idea. In this case, we first encode the secret s_1 into $E(s_1)$ as in Fig. 1.1.

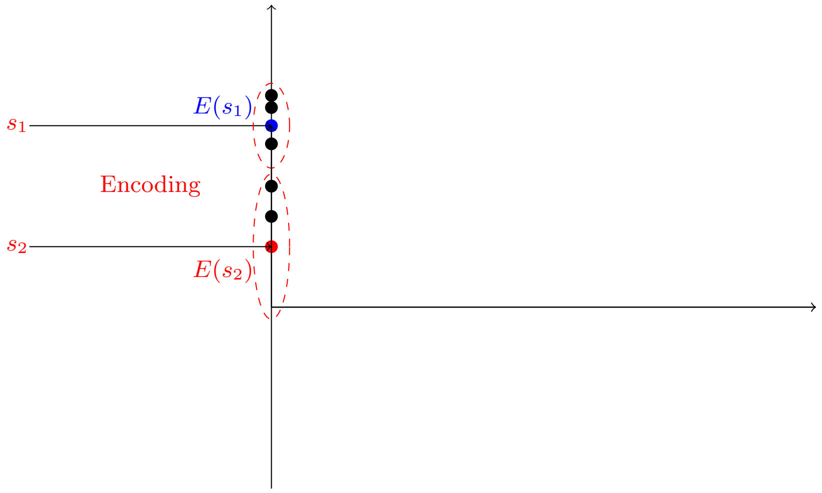

Figure 1.1: Encoding the secret via an AMD code

Secondly, the system uses a linear secret sharing scheme, for example the Shamir threshold scheme, to share the encoded message $E(s_1)$ into v shares as in Fig. 1.2. More specifically, the system randomly chooses $t-1$ coefficients $\{a_1, a_2 \dots, a_{t-1}\}$ from a finite field $\mathbb{F}_q$ and defines a polynomial $f(x) = E(s_1) + \sum_{1\leq i \leq t-1} a_i x^i$. The v different shares for $E(s_1)$ are $(x_i, f(x_i))$ for $1 \leq i \leq v$, where $x_i \in \mathbb{F}_q \setminus \{0\}$ for $1 \leq i \leq v$ are v fixed evaluation points. By polynomial interpolation, if the system gets t shares it is capable to recover $E(s_1)$, otherwise it can not get any information for $E(s_1)$ [86].

However, if there is an adversary who modifies his share $(x_t, f(x_t))$ into $(x_t, f'(x_t))$ then the system meets a problem for recovering the encoding message $E(s_1)$ and thus the original secret s_1 by accessing say $(x_i, f(x_i))$ for $1 \leq i \leq t$. As in Fig. 1.3, the information recovered from the t shares $(x_1, f(x_1)), (x_2, f(x_2)), \cdots, (x_{t-1}, f(x_{t-1}))$, $(x_t, f'(x_t))$ may be not $E(s_1)$.

Since the Shamir threshold scheme is a linear secret sharing scheme, we have

$$\begin{aligned} \mathit{Recover}(\mathit{Share}(E(s_1)) + \Delta) =& \mathit{Recover}(\mathit{Share}(E(s_1))) + \mathit{Recover}(\Delta) \\ =& E(s_1) + \mathit{Recover}(\Delta). \end{aligned}$$

Now, we have two cases. One is that $E(s_1) + \mathit{Recover}(\Delta) = E(s_2)$ for some valid secret s_2. Then the system can not detect the tampering by the adversary, since the system will regard $E(s_2)$ as the encoded message of the valid secret s_2 and then decode it into the secret s_2. Of course if this case happens, then the adversary wins, i.e., the original secret is tampered from s_1 to s_2 and the system has no knowledge about the tampering. The other case is that the system can not decode $E(s_1) + \mathit{Recover}(\Delta)$ into a valid secret, or it can decode $E(s_1) + \mathit{Recover}(\Delta)$ into

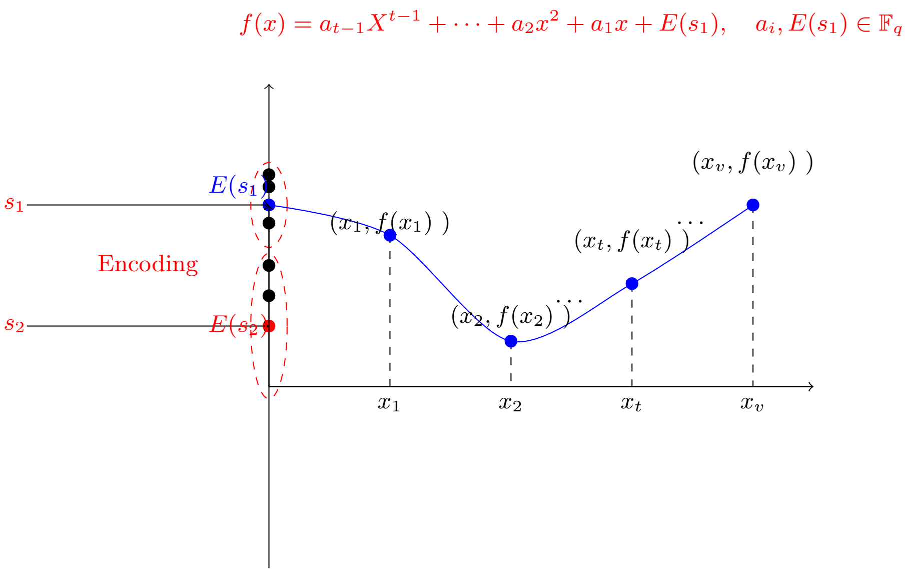

Figure 1.2: Encoding the secret via an AMD code

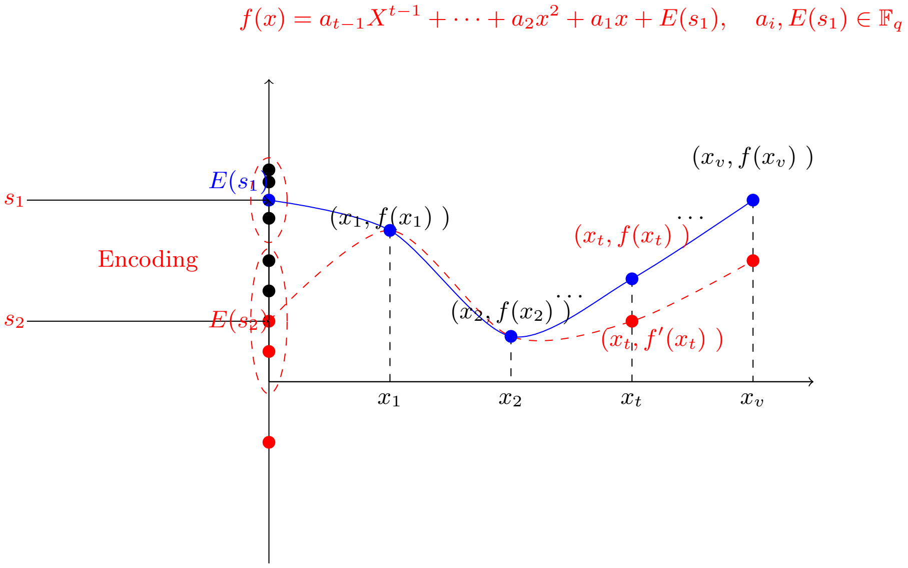

Figure 1.3: Encoding the secret via an AMD code

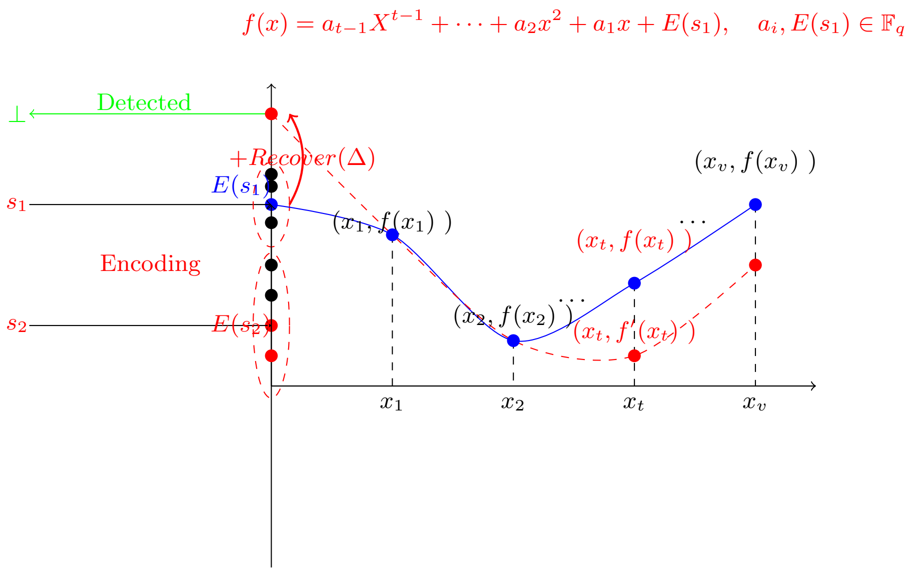

Figure 1.4: Encoding the secret via an AMD code

the original secret s_1, i.e.,

$$\text{Dec}(E(s_1) + Recover(\Delta)) \in \{s_1, \perp\}.$$

For this case, the system either can recover the correct secret or can detect the tampering as described in Fig. 1.4.

Note that we apply an AMD code to encode s_1. By the property of an AMD code, we have

$$\Pr(\text{Dec}(E(s_1) + Recover(\Delta)) \notin \{s_1, \perp\}) \leq \rho, \quad \text{for any } Recover(\Delta) \in \mathbb{F}_q \tag{1.1}$$

which means that the system can detect the tampering except for a given small error probability ρ. Thus, with the help of an AMD code, we can convert a linear secret sharing scheme into a robust secret sharing scheme referring to Fig. 1.5. From this viewpoint, it is important and interesting to further analyze AMD codes and construct new AMD codes.

1.2 Application scenarios for AMD codes: robust fuzzy extractors

Another application for AMD codes is to construct robust fuzzy extractors. Generally speaking, a fuzzy extractor extracts a uniformly random key R from a non-uniform some secret w (*e.g.* biometric data) in such a way that the key R can be recovered from any w' close to w in some appropriate metric space, say when $d(w, w') \leq t$ for a small given positive integer t. In this dissertation, we consider the case w and w' in

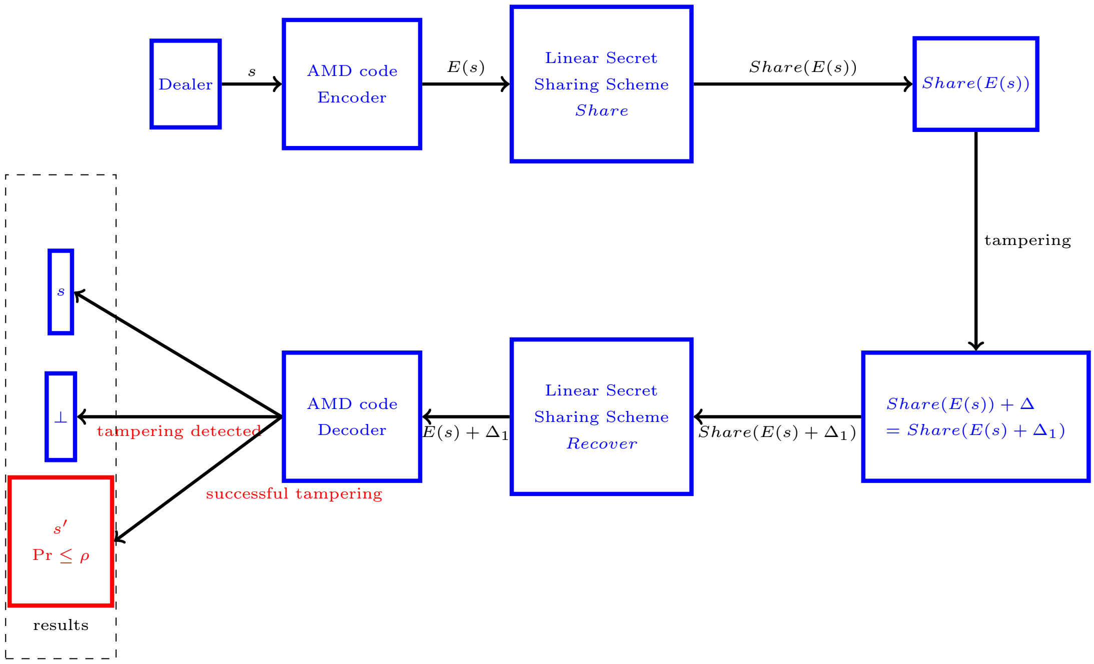

Figure 1.5: Robust secret sharing scheme based on an AMD code

some Hamming space and d denotes the Hamming distance. One of the well-known scenario has been considered for fuzzy extractors are the following key recovering problem:

- A user utilizes his biometric data w to generate a random key R together with some public string P, which is stored on a (possibly untrusted) server. The key R is used to encrypt some data for long-term storage.

- At a later point in time, the user obtains a refreshed biometric scan w' along with the value P from the server. Applying these information enables the user to recover R. In this way, the user may decrypt the original data.

However, this system may suffer some adversaries who can modify the information P stored on the public server. Robustness for fuzzy extractors requires that if the adversary modifies P into $\widetilde{P}$, then with high probability the user will detect this tampering of P and reject the refreshed biometric along with $\widetilde{P}$.

In [27], Cramer *et al.* introduced a method to generate a robust fuzzy extractor based on an AMD code and a known fuzzy extractor. In what follows, we simply explain the main idea of this construction. To this end, we begin with a special fuzzy extractor which contains three functions (SS, Rec, Ext):

- SS: From any biometric data w the function SS computes w into a public string P, i.e., $SS(w) = P$;

- *Ext*: From any biometric data w the function *Ext* extracts a uniformly random key R, i.e., $Ext(w) = R$;

- *Rec*: For any biometric data w' at a later point in time, the function *Rec* is capable to recover the key $R = Ext(w)$ by using w' and P, if the distance between w and w' is bounded by a fixed threshold t, i.e., if $d(w, w') \leq t$ then

  $$Rec(w', P) = w,$$

  and then can recover the key R with the help of *Ext*.

Cramer *et al.* also assumed that the fuzzy extractor has a kind of linearity such that, for the case $d(w, w') \leq t$, (1) $\widetilde{\Delta} = Rec(w', \widetilde{P}) - Rec(w, P)$ is determined by $\Delta = w' - w$, $\widetilde{P}$ and P, i.e., $\widetilde{\Delta}$ can be determined by a deterministic function h only related with Δ, P, and $\widetilde{P}$, say $\widetilde{\Delta} = h(\Delta, P, \widetilde{P})$, and (2) the extractor *Ext* satisfies the property that for any biometric data a, b,

$$Ext(a - b) = Ext(a) - Ext(b).$$

Let $E(x)$ be the probabilistic encoding map of an AMD code with form $E(x) = (x, y, f(x, y))$, where $f(x, y)$ is a map from $S \times G_1$ to G_2 and $y \in_R G_1$, where S denotes the source space, G_1 and G_2 are two groups. Now construct the new fuzzy extractor as

- Step 1: Divide the original key into two parts: from any biometric data w the function *Ext* extracts a key R, i.e., $Ext(w) = R = (R_a, R_{out})$;

- Step 2: Encode the public string by the AMD code: for any biometric data w the function *SS* computes a public string P and let $\sigma = f(P, R_a)$;

- Step 3: The new fuzzy extractor is given by (SS^*, Rec^*, Ext^*) with $Ext^*(w) = R^* = R_{out}$, $SS^*(w) = P^* = (P, \sigma)$, and $Rec^*(w, P^* = (P, \sigma)) = Rec(w, P)$ as explained in Fig. 1.6.

It is easy to check that for w' with $d(w, w') \leq t$, we have $Rec^*(w', P^*) = Rec(w', P) = w$, and then system can recover the key R_{out} by Ext^*. Consider the case that there exists an adversary who tampers with the public message from P^* to $\widetilde{P^*} = (P + \Delta_1, \sigma + \Delta_2)$. Now the system runs the Rec^* to get

$$\begin{aligned}
\widetilde{w} = Rec^*(w', \widetilde{P^*}) &= Rec^*(w, P^*) + Rec^*(w', \widetilde{P^*}) - Rec^*(w, P^*) \\
&= Rec(w, P) + Rec(w', P + \Delta_1) - Rec(w, P) \\
&= w + h(w' - w, P + \Delta_1, P) \\
&= w + \widetilde{\Delta},
\end{aligned} \tag{1.2}$$

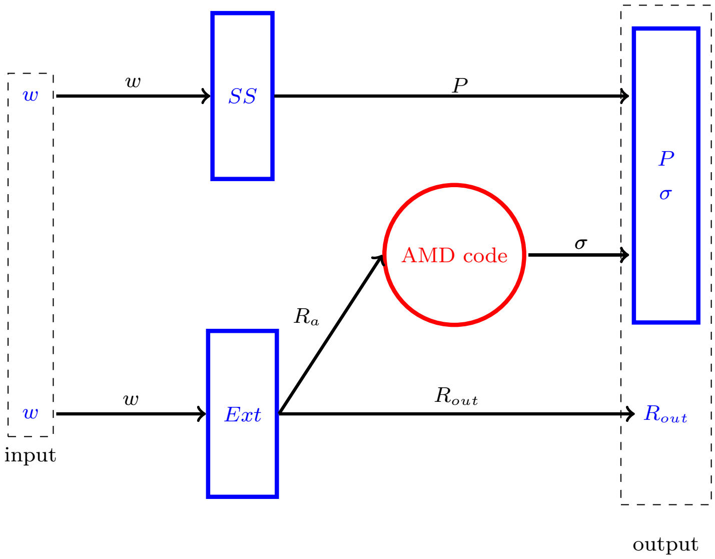

Figure 1.6: New fuzzy extractor based on an AMD code

where the fourth equality holds by the linearity of Rec and $\widetilde{\Delta} \triangleq h(w'-w, P+\Delta_1, P)$. Finally, the authenticity of $\widetilde{w}$ may be verified by checking whether the following holds

$$\sigma + \Delta_2 = f(P + \Delta_1, Ext_a(\widetilde{w})) = f(P + \Delta_1, Ext_a(w + \widetilde{\Delta})),$$

where $Ext(\widetilde{w}) = (\widetilde{R_a}, \widetilde{R_{out}}) \triangleq (Ext_a(\widetilde{w}), Ext_{out}(\widetilde{w}))$. By the linearity of the Ext we have

$$\begin{aligned}\sigma + \Delta_2 =& f(P + \Delta_1, Ext_a(w + \widetilde{\Delta}))\\ =& f(P + \Delta_1, Ext_a(w) + Ext_a(\widetilde{\Delta}))\\ =& f(P + \Delta_1, Ext_a(w) + \Delta_3)\\ =& f(P + \Delta_1, R_a + \Delta_3),\end{aligned}$$

where $\Delta_3 \triangleq Ext_a(\widetilde{\Delta})$. Note that this is also the security condition for the AMD code with probabilistic encoding map $E(x) = (x, y, f(x, y))$ to detect a tampering as in Fig. 1.7. The robustness of the new fuzzy extractor is determined by the ability of AMD code to detect a tampering. Thus, again the AMD code is the key point to construct a robust fuzzy extractor.

1.3 Known results for AMD codes

Based on the important applications introduced above, in the past decades, AMD codes have received much attention in the literature.

The progress is mainly twofold. For the first part, basic relationships among parameters of AMD codes are considered, i.e., to derive theoretic bounds for AMD

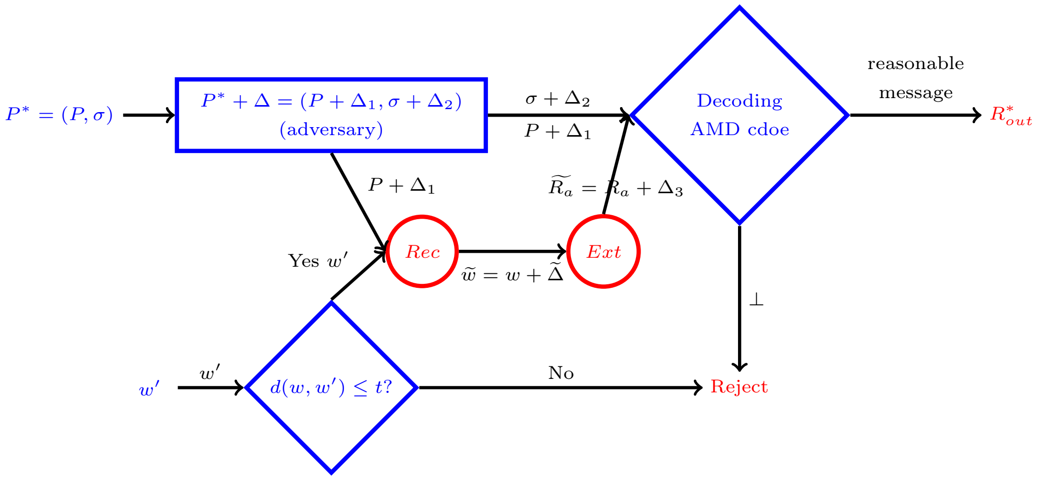

Figure 1.7: Decrypt process when there is an adversary.

codes. In [27], Cramer *et al.* built a lower bound for the minimum tag size of AMD codes, where the tag size denotes the difference between the length of the source and its corresponding encoding message. In [28, 29, 93], the well-known Singleton bound in coding theory was applied to show lower bounds for the minimum tag length of systematic AMD codes. In [23], Chen et al. established a lower bound on the minimum probability of successful tampering for adversaries for systematic AMD codes in terms of the nonlinearity of functions. Later in [81], Paterson and Stinson analyzed the case that the random choosing tampering is the best strategy for the adversaries, and then derived a lower bound on the minimum probability of successful tampering for adversaries that is the so-called R-bound for AMD codes. Another basic problem considered in [81] was when guessing the source message is the best strategy for the adversaries. For this case, Paterson and Stinson also provided a lower bound on the minimum probability of successful tampering for adversaries, that is, the G-bound.

For the second part, constructions of AMD codes are the focus. In this scenario, Cramer *et al.* [27] first introduced a construction of AMD codes with nearly optimal tag size based on polynomial evaluations. They also showed a construction for AMD codes via authentication codes. In [60] and [29], linear codes such as Reed-Muller codes and BCH codes were included to generate AMD codes with small tag size, respectively.

For constructions of AMD codes, the other method is via combinatorial method, which generates AMD codes by carefully designing the underlying combinatorial structures, i.e., the structure of image sets of the probabilistic encoding map E. In [28], Cramer *et al.* first introduced a kind of differential structures to construct AMD codes. The differential structure from caps in projective space were also used in [28] to construct AMD codes. In [81], various types of generalized external difference families (say, strong external difference families) were included by Paterson and Stinson to characterize optimal AMD codes for different merits of optimality, respectively. In [55], under the weak attack model assumption, a combinatorial characterization was given for AMD codes via weighted external difference families.

In addition, in the literature, there are many papers focus on external difference families which correspond to AMD codes as showed in [81]. For this part the reader may for examples refer to [4, 54, 57, 66, 68, 81, 95, 96].

Arrangement of this dissertation

In this paper we mainly focus on AMD codes for both strong and weak attack models. It is, however, somehow wider in scope and includes some constructions of external difference families related to the constructions about AMD codes.

In Chapter 2, we recall some necessary mathematical notation and concepts that will be used in this dissertation.

Chapter 3 contains a discussion about external difference families and their relationships to AMD codes. A new type of weighted external difference families is defined in this chapter. Some constructions of external difference families are also presented in this chapter.

In Chapter 4, the focus is AMD codes under the weak attack model (weak AMD codes). In this chapter both external difference families and highly nonlinear functions are included as tools to construct weak AMD codes. As results, a few infinite classes of new optimal weak AMD codes and new systematic weak AMD codes with asymptotically optimal tag size are constructed.

Chapter 5 contains the results about AMD codes under the strong attack model (strong AMD codes). In this chapter, we establish a relationship between highly nonlinear functions and systematic strong AMD coeds. On one hand, by choosing some special highly nonlinear functions, a few infinite classes of new systematic strong AMD codes with minimum possible probability of successful tampering are constructed. On the other hand, from a subclass of AMD codes with more strict assumptions, highly nonlinear functions can also be generated, where their nonlinearities are determined by the parameters of the corresponding AMD codes. Especially, we prove that the well-known construction for AMD codes in [27, Theorem 2] can also be explained by highly nonlinear functions. In addition, a combinatorial construction of G-optimal AMD codes is introduced for strong attack model.

Conclusions and open problems for further research are included in Chapter 6.

Chapter 2

Preliminary

In this chapter, we recall some mathematical notation and concepts that will be used in this dissertation.

2.1 Groups and finite fields

In this section, we recall some necessary algebraic conceptions and related results.

Definition 2.1.1. *A group is a set G together with a binary operation $*$ such that*

- *Associative: for any $a, b, c \in G$*

$$a * (b * c) = (a * b) * c;$$

- *Identity: there is an identity $e \in G$ with $a * e = e * a = a$;*
- *Inverse: for any $a \in G$, there exists an inverse element $a^{-1} \in G$ with*

$$a * a^{-1} = a^{-1} * a = e.$$

If the operation $*$ is also commutative, i.e.,

$$a * b = b * a \text{ for } a, b \in G,$$

then the group $(G, *)$ is an Abelian group.

For simplicity, we use G to denote $(G, *)$ and ab to denote $a * b$ if the operation $*$ is clear and without ambiguity.

Definition 2.1.2. *A group G is said to be cyclic if there exists an element $a \in G$ such that for any $b \in G$ there exists an integer m*

$$b = a^m \triangleq \underbrace{aa \ldots a}_{m}.$$

In this case, we also denote G as $\langle a \rangle$. And such an element a is named as a generator of G. Furthermore, if the set G is finite, then we call it as finite group. For a finite group, the number of elements is called its order denoted as $|G|$.

Example 2.1.3. *Let $G = \{1, 2, 3, 4, 5, 6\}$ and $a * b = ab \pmod 7$, then it is easy to check that G is a cyclic group with order 6 and 3 is a generator with*

$$3 = 3^1,\ 2 = 3^2,\ 6 = 3^3,\ 4 = 3^4,\ 5 = 3^5,\ 1 = 3^6.$$

Definition 2.1.4. *For a group $(G, *)$, if $H \subseteq G$ and $(H, *)$ also forms a group, then $(H, *)$ is a subgroup of $(G, *)$. For an element $a \in G$, the subgroup $H = \langle a \rangle$ is said to be a subgroup generated by a, and the order of $(\langle a \rangle, *)$ is said to be the order of the element $a \in G$.*

Besides groups, we also need the basic conception of rings in algebra.

Definition 2.1.5. *A ring $(R, +, *)$ is a set R together with two binary operations $+$ and $*$ satisfying*

- *$(R, +)$ forms an Abelian group;*
- *The operation $*$ is associative, i.e.,*

$$a * (b * c) = (a * b) * c, \quad \textit{for } a, b, c \in R;$$

- *Distributive laws: for any $a, b, c \in R$,*

$$a * (b + c) = a * b + a * c$$

and

$$(b + c) * a = b * a + c * a.$$

Usually, in a ring we use 0 to denote the identity of the Abelian group $(R, +)$.

Definition 2.1.6. *A ring $(R, +, *)$ is said to have identity if there exists $e \in R$ such that $a * e = e * a = a$ for any $a \in R$. (The identity of a ring is usually denoted as 1.)*

A ring $(R, +, *)$ is said to commutative if for any $a, b \in R$ we have $a * b = b * a$.

A ring $(R, +, *)$ forms an integral domain if for $a, b \in R$, $a * b = 0$ implies $a = 0$ or $b = 0$.

A ring $(R, +, *)$ is said to be a division ring if $(R \setminus \{0\}, *)$ forms a group.

Definition 2.1.7. *A ring $(F, +, *)$ is said to be a field if it is a division ring and also commutative.*

Remark 2.1.8. *A finite integral domain forms a finite field.*

For a finite field the following results and definitions are frequently used in this thesis.

Lemma 2.1.9. *Let $\mathbb{F}_q$ be a finite field with q elements. Then*

- *A finite field $\mathbb{F}_q$ exists if and only if q is a prime power, i.e., there exists a prime p and a positive integer t with $q = p^t$.*
- *For a prime p, a positive integer t, and a finite field $\mathbb{F}_{p^t}$,*

$$pa \triangleq \sum_{1 \leq i \leq p} a = 0 \text{ for any } a \in \mathbb{F}_{p^t} \tag{2.1}$$

  and p is the smallest positive integer that (2.1) holds.

- *For positive integers m and t, if $m \mid t$, then there exists a finite field $\mathbb{F}_{q^m}$ with $\mathbb{F}_{q^m} \subseteq \mathbb{F}_{q^t}$.*
- *The group $(\mathbb{F}_q^* \triangleq \mathbb{F}_q \setminus \{0\}, *)$ is cyclic.*

Definition 2.1.10. *A generator of the cyclic group $\mathbb{F}_q^*$ is said to be a primitive element of the finite field $\mathbb{F}_q$.*

Definition 2.1.11. *For positive integers $m \mid t$, and a prime p, define the trace function from $\mathbb{F}_{p^t}$ to $\mathbb{F}_{p^m}$ as*

$$\mathrm{Tr}_{p^m}^{p^t}(x) = \sum_{0 \leq i \leq \frac{t}{m}-1} x^{p^{im}} \quad \text{for } x \in \mathbb{F}_{q^t}. \tag{2.2}$$

2.2 Secret sharing schemes

In cryptography, one important problem is to distribute a secret among a group of users, each of whom is allocated with a share of the secret. Generally speaking, a secret sharing scheme encodes a source information (or secret) into a set of shares $\mathcal{S}$ satisfying

- any authorized set of $\mathcal{S}$ can reconstruct the source information;
- any unauthorized set of shares gives absolutely no information about the source information.

Formally, a threshold secret sharing scheme (or threshold scheme) can be defined as follows.

Definition 2.2.1. *Suppose there is a secret s. A (t, k)-threshold secret sharing scheme is a set $\mathcal{S}$ of k elements called shares, and a threshold $t \leq k$ such that*

- *any t-subset of $\mathcal{S}$ can reconstruct the secret s;*
- *any subset of $\mathcal{S}$ with less than t shares gives absolutely no information about s.*

As an example we simply introduce the Shamir threshold scheme [86]. The basic idea of the Shamir threshold scheme is to evaluate a polynomial with degree less than or equal to $t-1$ into k values. As a result any t evaluation points are capable to recover the original polynomial.

Example 2.2.2. ((t, k)-*Shamir threshold scheme*) *Let q be a prime power and k and t be two positive integers, $t \leq k$. The secret $s \in \mathbb{F}_q$ is chosen by a special user called the dealer. When the dealer wants to share the secret s among k users, he gives each user some partial information called share.*

Initialization Phase:

1 The dealer chooses k distinct, non-zero elements of $\mathbb{F}_q$, denoted as α_i for $1 \leq i \leq k$. For $1 \leq i \leq k$, the dealer gives the value α_i to user U_i. The values α_i for $1 \leq i \leq k$ are public.

Share Distribution:

2 Suppose the dealer wants to share a secret $S \in \mathbb{F}_q$. The dealer secretly chooses, independent at random, $t-1$ elements of $\mathbb{F}_q$, which are denoted as $e_1, e_2, \cdots, e_{t-1}$.

3 For $1 \leq i \leq k$, compute $\beta_i = f(\alpha_i)$, where

$$f(x) = S + \sum_{1 \leq i \leq t-1} e_i x^i.$$

4 For $1 \leq i \leq k$, the dealer gives the share β_i to U_i.

Reconstruct Phase:

5 Any group of t users can compute the $f(x)$ by using the interpolation formula

$$S = \sum_{1 \leq j \leq t} \beta_{i_j} \frac{\prod_{1 \leq \tau \neq j \leq t} \alpha_{i_\tau}}{\prod_{1 \leq \tau \neq j \leq t} (\alpha_{i_\tau} - \alpha_{i_j})}.$$

Suppose we define

$$b_j = \frac{\prod_{1 \leq \tau \neq j \leq t} \alpha_{i_\tau}}{\prod_{1 \leq \tau \neq j \leq t} (\alpha_{i_\tau} - \alpha_{i_j})}, \quad \text{for } 1 \leq j \leq t.$$

Note that the $b_j s$ can be precomputed, if desired, and their values are not secret. Then we have

$$S = \sum_{1 \leq j \leq t} b_j \beta_{i_j}.$$

Hence the secret S is a linear combination of the shares. Such a secret sharing scheme is call linear.

2.3 Algebraic manipulation detection codes

In this section, we recall necessary definitions and notation related with algebraic manipulation detection codes.

Let S be the source space, i.e., the set of plaintext messages with size m, and G be the encoded message space. A probabilistic encoding function E maps $s \in S$ to some $g \in G$. Let $A_s \subseteq G$ denote the set of valid encodings of $s \in S$, where $A_s \cap A_{s'} = \emptyset$ is required for any $s \neq s'$ so that any message $g \in A_s$ can be correctly decoded as $\text{Dec}(g) = s$. Denote $\mathcal{A} \triangleq \{A_s : s \in S\}$.

Definition 2.3.1 ([81]). *For given $(S, G, \mathcal{A}, E)$, let*

- *The value $\Delta \in G^* \triangleq G \setminus \{0\}$ be chosen according to the adversary's strategy σ;*
- *The source message $s \in S$ be chosen uniformly at random by the encoder, i.e., we assume equiprobable sources;*
- *The message s be encoded into $g \in A_s$ using the encoding function E and there exists a deterministic decoding function* $\text{Dec} : G \to S \cup \{\perp\}$ *such that* $\text{Dec}(E(s)) = s$ *with probability* 1 *for any $s \in S$;*
- *The adversary wins (a successful tampering) if and only if $g + \Delta \in A_{s'}$ with $s' \neq s$.*

The probability of successful tampering is denoted by ρ_σ for strategy σ of the adversary. The code $(S, G, \mathcal{A}, E)$ is called an (n, m, K, ρ) algebraic manipulation detection code (or an (n, m, K, ρ)-AMD code for short) where $K = \{|A_s| : s \in S\}$ is a multiset and ρ denotes the maximum probability of successful tampering for all possible strategies, i.e.,

$$\rho = \max_{\sigma} \rho_\sigma.$$

Especially, if E encodes s to an element of A_s uniformly, i.e., $Pr(E(s) = g) = \frac{1}{|A_s|}$ for any $s \in S$ and $g \in A_s$, then we use $(S, G, \mathcal{A}, E_u)$ to distinguish this kind of AMD codes.

Throughout this thesis, we fix the following notation for AMD codes.

- An (n, m, K, ρ)-AMD code is said to have *equiprobable sources* if $\Pr(s) = \frac{1}{m}$ for any $s \in S$.
- An (n, m, K, ρ)-AMD code is said to be *equiprobable encoding* if $\Pr(E(s) = g) = \frac{1}{|A_s|}$ for any $s \in S$ and $g \in A_s$.
- An (n, m, K, ρ)-AMD code is said to be *uniform* if $|A_s|$ is constant for any $s \in S$.

- A uniform (n, m, K, ρ)-AMD code with $|A_s| = t$ for $s \in S$ is said to be *t-regular* if it has equiprobable sources and equiprobable encoding.

Remark 2.3.2. *Generally speaking, the distribution $K = (a_1, a_2, \cdots, a_m)$ is an essential parameter for an AMD code. However, for the convenience of discussion, the notation (n, m, a, ρ)-AMD code is used to denote an $(n, m, K = (a_1, a_2 \cdots, a_m), \rho)$-AMD code with $a = \sum_{1 \le i \le m} a_i$, when K is unknown or regarded as a variable.*

Definition 2.3.3. *An (n, m, K, ρ)-AMD code is said to be a strong (n, m, K, ρ) algebraic manipulation detection (AMD) code if for any $s \in S$, $\Delta \in G \setminus \{0\}$, the probability of* $\mathrm{Dec}(E(s) + \Delta) \notin \{s, \perp\}$ *is at most ρ, i.e.,*

$$\Pr(\mathrm{Dec}(E(s) + \Delta) \notin \{s, \perp\}) \le \rho < 1.$$

Definition 2.3.4. *An (n, m, K, ρ)-AMD code is called a weak (n, m, K, ρ)-AMD code if for any $\Delta \in G \setminus \{0\}$ and any random $s \in_R S$ rather than an arbitrary one, the probability*

$$\sum_{s \in S} \Pr(s) \sum_{g \in A_s} \Pr(E(s) = g) \Pr(\mathrm{Dec}(g + \Delta)) \notin \{s, \perp\}) \le \rho < 1.$$

The following special AMD codes are used in robust fuzzy extractors.

Definition 2.3.5. *An AMD code is called systematic if $S = G_1$ is an Abelian group, G is an Abelian group $G_1 \times G_2 \times B$, and the encoding has the form*

$$E : G_1 \to G_1 \times G_2 \times B \quad \text{with } E(s) = (s, x, f(s, x)) \tag{2.3}$$

for some function $f : G_1 \times G_2 \to B$ and $x \in_R G_2$.

For a systematic AMD code, the decoding function is naturally given by

$$\mathrm{Dec}(s, x, t) = \begin{cases} s, & \text{if } t = f(s, x), \\ \perp, & \text{otherwise.} \end{cases} \tag{2.4}$$

For a systematic AMD code, if $\Pr(E(s) = (s, x, f(s, x)) \neq 0$ for any $s \in G_1$ and $x \in_R G_2$, then it is $|A_2|$-uniform. Thus, an equiprobable encoding systematic AMD code with equiprobable sources is $|G_2|$-regular.

Definition 2.3.6 ([27])**.** *The tag size of an (n, m, a, ρ)-AMD code is defined as*

$$\varpi = \log |G| - \log |S| = \log n - \log m.$$

For the convenience of theoretic analysis, for any $u, k \in \mathbb{N}$, define *effective tag size* as

$$\varpi^*(k, u) = \min\{\log |G|\} - u, \tag{2.5}$$

where the minimum is over all $(|G|, |S|, a, \rho)$-AMD codes such that $|S| \geq 2^u$ and $\rho \leq 2^{-k}$. We point out that in this definition, the group G and the size distribution K can be different in this family of codes with $|S| \geq 2^u$ and $\rho \leq 2^{-k}$, and are left unspecified. In [27], Cramer *et al.* derived a lower bound for $\varpi^*(k, u)$ as follows.

Lemma 2.3.7 ([27]). *For any $u, k \in \mathbb{N}$, the effective tag size is lower bounded by*

$$\varpi^*(k, u) \geq 2k - 2^{-u+1} \geq 2k - 1$$

for strong AMD codes, and

$$\varpi^*(k, u) \geq k - 2^{-u+1} \geq k - 1$$

for weak AMD codes, respectively.

2.4 Highly nonlinear functions

In this section, we recall some necessary definitions about the nonlinearity of functions.

Let $(A, +)$ and $(B, +)$ be two Abelian groups with order n and m, respectively. Let f be a function from A to B. One way to measure the nonlinearity of a function f from A to B is to use the derivatives $D_a(f(x)) = f(x+a) - f(x)$ for $a \in A$, which is closely related to the differential cryptanalysis of stream ciphers [5, 74].

Definition 2.4.1 ([74]). *The nonlinearity N_f of a function f from A to B is defined as*

$$\begin{aligned} N_f &\triangleq \max_{a \in A \setminus \{0\}} \max_{b \in B} \Pr\left(D_a(f(x)) = b\right) \\ &= \max_{a \in A \setminus \{0\}} \max_{b \in B} \frac{|\{x \in A : \ D_a(f(x)) = b\}|}{|A|}, \end{aligned} \tag{2.6}$$

where $\Pr(D_a(f(x)) = b)$ denotes the probability of the occurrence of the event $D_a(f(x)) = b$.

Remark 2.4.2. *The Hamming distance between two functions f and g from A to B is defined to be $d(f, g) = |\{x \in A : f(x) \neq g(x)\}|$. A function f is linear if and only if $f(x + y) = f(x) + f(y)$ for all $x, y \in A$. A function g is affine if and only if $g = f + b$, where f is linear and b is a constant. An alternative method of measuring the nonlinearity of a function $f : A \to B$ is given by the minimum Hamming distance between f and all possible affine functions from A to B [78]. This measure of nonlinearity is closely related to linear cryptanalysis of stream ciphers [69]. For the relationship between these two definitions of nonlinearity, the reader is referred to [16, 26], for instances. In this paper, the former definition of nonlinearity is used.*

It is easy to check (see, for example, [16]) that $N_f = 1$ if f is a linear function from A to B, and $N_f \geq \frac{1}{|B|}$ for any function f from A to B. The smaller the value of N_f, the higher the corresponding nonlinearity of f.

Definition 2.4.3 ([74]). *A function f from A to B is said to have perfect nonlinearity if $N_f = \frac{1}{|B|}$. In this case, the function f is a perfect nonlinear function (PN function) if it has nonlinearity $\frac{1}{|B|}$.*

It is well-known that PN functions from $\mathbb{F}_q$ to $\mathbb{F}_q$ only exist over finite fields with odd characteristic, since for a function $f(x)$ from $\mathbb{F}_{2^m}$ to $\mathbb{F}_{2^m}$, the fact that $\theta \in \mathbb{F}_{2^m}$ is a root of $f(x+a) - f(x) = b$ for $b \in \mathbb{F}_q$ and $a \neq 0 \in \mathbb{F}_q$ always implies the fact that $\theta + a$ is also a root. That is, for this case, the possible minimum nonlinearity is $\frac{2}{|B|}$.

Definition 2.4.4. *A function $f(x)$ is said to be an almost perfect nonlinear (APN) permutation, if it is a permutation over $\mathbb{F}_{2^m}$ with nonlinearity $\frac{2}{|B|} = 2^{1-m}$.*

2.5 Error correcting codes

In this section we recap some necessary preliminaries about error correcting codes.

Definition 2.5.1. *Let n and q be positive integers and Q be a finite set with cardinality q. A q-ary code of length n is a set of vectors $\mathcal{C} \subseteq Q^n$, denoted (n, q) code. The vector $C \in \mathcal{C}$ is called a codeword. The code rate of $\mathcal{C}$ is given by*

$$r = \frac{\log_q |\mathcal{C}|}{n},$$

where the cardinality of $\mathcal{C}$, i.e., $|\mathcal{C}|$ is called the code size.

Definition 2.5.2. *The Hamming distance between two codewords $U = (u_1, u_2, \ldots, u_n)$ and $V = (v_1, v_2 \ldots, v_n)$ is*

$$d(U, V) = |\{1 \leq i \leq n \ : \ u_i \neq v_i\}|.$$

Definition 2.5.3. *The Hamming weight of a codeword $C = (c_1, c_2, \ldots, c_n) \in Q^n$ is*

$$\mathrm{Wt}(C) = d(C, \mathbf{0}) = |\{1 \leq i \leq n : \ c_i \neq 0\}|,$$

where $\mathbf{0} = (0, 0, \ldots, 0)$.

If $Q = \mathbb{F}_q$ and the code $\mathcal{C}$ forms a k-dimensional linear subspace of $\mathbb{F}_q^n$, then this code is a linear code denoted as $[n, k]_q$ linear code. More specifically,

Definition 2.5.4. *Let $H \in \mathbb{F}_q^{(n-k)\times n}$ be an $(n-k) \times n$ matrix with rank $n - k$. The set $\mathcal{C}$ of all the vectors with length n such that $Hc^\top = \mathbf{0}$ is called a linear (n, k) code also denoted as $[n, k]_q$ code for short. The matrix H is called the party-check matrix of $\mathcal{C}$. If H has form (A, I_{n-k}), then $\mathcal{C}$ is called a systematic code.*

Theorem 2.5.5. (*Singleton Bound*) *If $\mathcal{C}$ is an $[n,k,d]_q$ linear code, then*

$$n \geq k + d - 1.$$

Definition 2.5.6. *An $[n,k,d]_q$ linear code $\mathcal{C}$ is said to be maximum distance separable (MDS) code, if it achieves the Singleton Bound in Theorem 2.5.5 with equality, i.e., $d = n - k + 1$.*

Example 2.5.7. (*Generalized Reed-Solomon (GRS) code*) *Let $\theta = \{\theta_i \,:\, 1 \leq i \leq n\}$ be an n-subset of $\mathbb{F}_q$, where we assume $q \geq n$, then the well-known generalized Reed-Solomon (GRS) code with parameters $[n, k, n-k+1]_q$ can be defined as*

$$\begin{aligned} &GRS_k(\theta, \alpha) \\ &\triangleq \{(\alpha_1 f(\theta_1), \alpha_2 f(\theta_2), \ldots, \alpha_n f(\theta_n)) \;:\; f(x) \in \mathbb{F}_q[x] \text{ with } \deg(f(x)) < k\}, \end{aligned} \tag{2.7}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in (\mathbb{F}_q^*)^n$. Especially, if $(\alpha_1, \alpha_2, \ldots, \alpha_n) = \mathbf{1} = (1, 1, \ldots, 1)$ then the code $GRS_k(\theta, \mathbf{1})$ is named as Reed-Solomon (RS) code denoted as $RS_k(\alpha)$.

2.6 Difference systems of sets and comma-free codes

In this section, we recall some necessary background and definitions for difference systems of sets and comma-free codes.

Difference systems of sets (DSSs) are a special class of combinatorial structures which can be used to obtain comma-free codes for synchronization over erroneous channels. Because of this application and their own theoretical interest in combinatorics, DSSs have received much attention during these two decades (see [22,37,62,90], and the references therein).

Consider the process of transmitting data over a channel, where the data being sent is a stream of symbols from an alphabet F of size l, say, $F = \{0, 1, \ldots, l-1\}$. The data stream consists of consecutive messages, each being a sequence of v consecutive symbols

$$\cdots \underbrace{x_0 \cdots x_{v-1}}\,\underbrace{y_0 \cdots y_{v-1}} \cdots .$$

The synchronization problem that arises at the receiver is the task of correctly partitioning the data stream into messages of length v, as opposed to incorrectly conceiving a sequence of v symbols that is the concatenation of the end of one message with the beginning of another message as a single message, for example,

$$\cdots \underbrace{x_i \cdots x_{v-1} y_0 \cdots y_{i-1}} \cdots .$$

One way to solve the synchronization problem is to utilize comma-free codes. Let $\mathcal{C}$ be a subset of F^v. The elements in $\mathcal{C}$ are called codewords as usual. A code $\mathcal{C}$ is termed a *comma-free code* if the concatenation

$$T_i(x, y) = x_i \cdots x_{v-1} y_0 \cdots y_{i-1}$$

of any two (not necessarily distinct) codewords $x = (x_0, \dots, x_{v-1})$ and $y = (y_0, \dots, y_{v-1})$ is never a codeword in $\mathcal{C}$. Further, the *comma-free index* $\rho = \rho(\mathcal{C})$ [51] of a code $\mathcal{C} \subseteq F^v$ is defined as

$$\rho = \min\{d(z, T_i(x, y)) : x, y, z \in \mathcal{C} \text{ and } i = 1, 2, \dots, v-1\},$$

where $d(x, y)$ is the Hamming distance between $x, y \in F^v$. The comma-free index $\rho(\mathcal{C})$ allows one to distinguish a codeword from an overlap of two codewords (and hence provides for synchronization of codewords) even in case that up to $\lfloor(\rho(\mathcal{C}) - 1)/2\rfloor$ errors have occurred in the given codeword [51].

It turns out [61] that codes with prescribed comma-free index can be constructed by using difference systems of sets, a type of combinatorial structures. Let $\mathcal{S} = \{S_0, S_1, \dots, S_{l-1}\}$ be a collection of l disjoint subsets of $\mathbb{Z}_v$ and $\tau_i = |S_i|$ for each $0 \leq i \leq l-1$.

Definition 2.6.1. *$\mathcal{S}$ is said to be a $(v, (\tau_0, \tau_1, \dots, \tau_{l-1}), \rho)$ difference system of sets (DSS) if the multiset*

$$\{a - b \pmod{v} : a \in S_i, b \in S_j, i \neq j, 0 \leq i, j \leq l-1\} \tag{2.8}$$

contains every element $x \in \mathbb{Z}_v \setminus \{0\}$ at least ρ times. A DSS is perfect if every element $x \in \mathbb{Z}_v \setminus \{0\}$ is contained exactly ρ times in the multiset of (2.8), and is regular if all subsets S_i's are of the same size, i.e., $\tau_0 = \tau_1 = \cdots = \tau_{l-1}$.

In fact, the comma-free code $\mathcal{C}(\mathcal{S})$ corresponding to the DSS $\mathcal{S} = \{S_i : 0 \leq i \leq l-1\}$ is given as

$$\mathcal{C}(\mathcal{S}) = \left\{ (c_0, c_1, \dots, c_{v-1}) : c_i = j \text{ for } i \in S_j \text{ and } c_i = *, \text{ for } i \notin \bigcup_{0 \leq i \leq l-1} S_i \right\}.$$

Note that information symbols of the comma-free code $\mathcal{C}(\mathcal{S})$ are in the places where $c_i = *$. Thus, the redundancy of $\mathcal{C}(\mathcal{S})$ is $\sum\limits_{0 \leq i \leq l-1} \tau_i$.

Definition 2.6.2. *The code rate of the comma-free code from a $(v, (\tau_0, \tau_1, \dots, \tau_{l-1}), \rho)$ DSS (also called the code rate of the DSS) is given as*

$$1 - \frac{\sum\limits_{0 \leq i \leq l-1} \tau_i}{v}. \tag{2.9}$$

From a practical view of point, comma-free codes with high code rate are preferred [45, 61]. It is very desirable that the redundancy $\sum\limits_{0 \leq i \leq l-1} \tau_i$ is as small as possible. We denote by $r_l(v, \rho)$ the minimum redundancy of all DSSs with parameters $(v, (\tau_0, \tau_1, \dots, \tau_{l-1}), \rho)$.

Levenshtein [61] proved the following bound on $r_l(v,\rho)$:

$$r_l(v,\rho) \geq \sqrt{\frac{l\rho(v-1)}{l-1}},$$

with equality if and only if the DSS is perfect and regular. From the above Levenshtein bound, Wang [90] derived a sharper bound.

Lemma 2.6.3 ([90]). *For a DSS with parameters* $(v,(\tau_0,\tau_1,\dots,\tau_{l-1}),\rho)$*,*

$$r_l(v,\rho) \geq \left\lceil \sqrt{\rho(v-1)+\left\lceil \frac{\rho(v-1)}{l-1} \right\rceil} \right\rceil . \tag{2.10}$$

A $(v,(\tau_0,\tau_1,\dots,\tau_{l-1}),\rho)$ DSS is said to be *optimal* if its redundancy is equal to $r_l(v,\rho)$, i.e., $\sum\limits_{0\leq i\leq l-1} \tau_i = r_l(v,\rho)$.

Chapter 3

External Difference Families

In this chapter, we introduce some definitions and constructions of external difference families, which will be the main tool to construct algebraic manipulation detection codes for both weak and strong models. We begin with definitions of some special kinds of external difference families.

3.1 External difference families: definitions

In this section, we recall some notation and definitions about external difference families. First of all, we describe some necessary notation.

- For a multi-set B and a positive integer k, let $k \boxtimes B$ denote the multi-set, where each element of B repeated k times.
- For a subset $B \subseteq G$, $D(B)$ denotes the multi-set $\{a - b \in G : a, b \in B,\ a \neq b\}$.
- For subsets $B_1, B_2 \subseteq G$, $D(B_1, B_2)$ denotes the multi-set $\{a - b \in G : a \in B_1, b \in B_2\}$.

Definition 3.1.1 ([26]). *Let G be an addictive Abelian group of order n. Let $\mathcal{B} = \{B_i : 1 \leq i \leq m\}$ be a family of subsets of G. Then $\mathcal{B}$ is called a difference family (DF) if each nonzero element of G appears exactly λ times in the multi-set $\bigcup_{1 \leq i \leq m} D(B_i)$, i.e.,*

$$\bigcup_{1 \leq i \leq m} D(B_i) = \lambda \boxtimes (G \setminus \{0\}).$$

Let $K = (|B_1|, |B_2|, \ldots, |B_m|)$. One briefly says that $\mathcal{B}$ is an (n, K, λ)-DF.

Example 3.1.2. *Let $n = 7$ and $k = 3$, then the set $\{\{2, 4, 1\}, \{6, 5, 3\}\}$ is a difference family with parameter $(7, (3, 3), 2)$ over $(\mathbb{Z}_7, +)$.*

Definition 3.1.3. *When $m = 1$, the set B_1 is also called an $(n, k = |B_1|, \lambda)$ difference set, or briefly an (n, k, λ)-DS.*

Example 3.1.4. *Let $n = 7$ and $k = 3$, then the set $\{2, 4, 1\}$ is a difference set with parameter $(7, 3, 1)$ over $(\mathbb{Z}_7, +)$.*

Definition 3.1.5. *Let G be an addictive group of order n. A k elements subset B of G is said to be an (n, k, λ, μ) partial difference set (PDS), if $D(B)$ contains each non-identity elements of B exactly λ times and non-identity elements of $G \setminus B$ exactly μ times.*

If $\mathcal{B}$ forms a partition of G, then $\mathcal{B}$ is called a *partitioned difference family* (PDF) [37] and denoted as an (n, K, λ)-PDF. In the literature, PDFs are proved to be useful in frequency-hopping sequence design and coding theory.

The parameters of some known PDFs over cyclic groups are listed in Tables 3.1-3.3.

Fact 3.1.6 ([37])**.** *Let $S = \{S_1, S_2, \dots, S_l\}$ be a partition of $\mathbb{Z}_v$, where $|S_i| = \tau_i$, $1 \leq i \leq l$. Then S is a $(v, (\tau_1, \tau_2, \dots, \tau_l), \lambda)$-PDF if and only if S is a $(v, (\tau_1, \tau_2, \dots, \tau_l), v-\lambda)$ perfect DSS.*

Thus, every PDF in Tables 3.1-3.3 corresponds to a perfect DSS given by Definition 2.6.1.

Perfect DSSs are also known as external difference families, which were first defined to construct authentication codes and secret sharing schemes. More specifically, an external difference family can be defined as:

Definition 3.1.7 ([77])**.** *Let $\mathcal{B} = \{B_i : 1 \leq i \leq m\}$ be a family of disjoint subsets of G. Then $\mathcal{B}$ forms an external difference family (EDF) if each nonzero element of G appears exactly λ times in the union of multi-sets $D(B_i, B_j)$ for $1 \leq i \neq j \leq m$, i.e.,*

$$\bigcup_{1\leq i\neq j\leq m} D(B_i, B_j) = \lambda \boxtimes (G \backslash \{0\}).$$

We briefly denote $\mathcal{B}$ as an (n, m, K, λ)-EDF, where $K = (|B_1|, |B_2|, \dots, |B_m|)$. An EDF is regular if $|B_1| = |B_2| = \cdots = |B_m| = k$, denoted as an (n, m, k, λ)-EDF.

Example 3.1.8. *Let $n = 7$, $m = 3$, and $G = (\mathbb{Z}_7, +)$, then it is easy to check that*

$$\mathcal{B}_1 = \{\{0\}, \{3, 6, 5\}, \{2, 4, 1\}\}$$

and

$$\mathcal{B}_2 = \{\{3, 6, 5\}, \{2, 4, 1\}\}$$

is a $(7, 3, (1, 3, 3), 5)$-EDF and a $(7, 2, 3, 2)$-EDF, respectively.

A DSS considers the external differences of a set system which contain nonzero elements at least λ times. A perfect DSS is exactly an external difference family. Thus, each PDF in Tables 3.1-3.3 corresponds to an EDF.

In [62], to construct algebraic manipulation detection codes, Paterson and Stinson considered the opposite case of DSSs, that is, the external differences of a set system contain nonzero elements at most λ times.

Table 3.1: Some known PDFs over cyclic groups with parameters $(v, \mathcal{W}, \lambda)$

Parameters	Constraints	Ref.
$\left(tp,\ (t^1, (t+1)^{\frac{t(p-1)}{t+1}}),\ 2\right)$	$p \equiv 1 \pmod{2(t+1)}$, and $t = 2, 4$	[9]
$((2e-1)v,\ \mathcal{W},\ e-1)$ $\mathcal{W} = ((e-1)^1, e^{\frac{(2e-1)v-e+1}{e}})$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $e\|(p_t - 1)$ for $1 \le t \le r$ $2e-1$ is a prime	[9]
$\left(sv,\ (2^1, 3^{\frac{sv-2}{3}}),\ 2\right)$,	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $3\|(p_t - 1)$ for $1 \le t \le r$, $s \in \{2, 8, 11, 17, 23, 29, 35, 41\}$	[9]
$\left(sv,\ (4^1, 5^{\frac{sv-4}{5}}),\ 4\right)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $10\|(p_t - 1)$ for $1 \le t \le r$ $s \in \{4, 19, 29, 39\}$	[9]
$\left(sv,\ (5^1, 6^{\frac{sv-5}{6}}),\ 5\right)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $6\|(p_t - 1)$ for $1 \le t \le r$ $s \in \{11, 23, 29, 41\}$	[9]
$\left(5v,\ (5^1, 6^{\frac{5v-5}{6}}),\ 5\right)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $12\|(p_t - 1)$ for $1 \le t \le r$	[9]
$\left(41v,\ (6^1, 7^{\frac{41v-6}{7}}),\ 6\right)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $14\|(p_t - 1)$ for $1 \le t \le r$	[9]
$\left(6v,\ (6^1, 7^{\frac{6v-6}{7}}),\ 6\right)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $28\|(p_t - 1)$ for $1 \le t \le r$	[9]
$\left(sv,\ (7^1, 8^{\frac{sv-7}{8}}),\ 7\right)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $8\|(p_t - 1)$ for $1 \le t \le r$ $s \in \{7, 31, 47, 71, 79, 103\}$	[9]
$\left(v,\ (1^1, e^{\frac{v-1}{e}}),\ e-1\right)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, and $e\|(p_t - 1)$ for $1 \le t \le r$	[13][47] [65][82]

Herein p, p_i's and q_i's are primes; t, e, s, r and m are positive integers; q is a prime power;

We use $\mathcal{W}^*(l)$ to denote the case where $|\mathcal{W}| = l$ but the distribution of $\mathcal{W}$ is not clear;

For PDFs over noncyclic groups, the reader is referred to, for example, Table 1 in [65].

Table 3.2: Some known PDFs over cyclic groups with parameters $(v, \mathcal{W}, \lambda)$

Parameters	Constraints	Ref.
$\Big(v,\ (1^1, (e-1)^{\frac{v-1}{e-1}}),\ e-2\Big)$	$v = ep_1^{m_1}p_2^{m_2}\dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $e(e-1) \| (p_t - 1)$ for $1 \leq t \leq r$	[14]
$(p^2,\ ((2p-1)^1, (p-1)^{\frac{p^2-2p+1}{p-1}}),\ p)$	p is an odd prime	[?]
$\Big(\frac{q^t-1}{m},\ \mathcal{W}^*(q),\ \frac{q^{t-1}-1}{m}\Big)$	$m \| (q-1)$, and $\gcd(m,t) = 1$	[?][37]
$(q^2+1,\ \mathcal{W}^*(q),\ q+1)$	$q = 2^t$, $t \geq 1$	[?]
$(q-1, ((\frac{q}{d}-1)^1, \frac{q}{d}^{d-1}), \frac{q-d}{d})$	$d \| q$	[38]
$\Big(2^m-1,\ (1^1, m^{\frac{2^m-2}{m}}),\ m-1\Big)$	m is a prime	[39]
$\Big(2^m-1,\ (1^1, (2m)^{\frac{2^m-2}{2m}}),\ 2m-1\Big)$	m is an odd prime	[39]
$\Big((e+1)v,\ (1^1, e^{\frac{(e+1)v-1}{e}}),\ e-1\Big)$	$v = p_1^{m_1}p_2^{m_2}\dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $e \| (p_t - 1)$ for $1 \leq t \leq r$	[65]
$\big((v,\ \mathcal{W},\ \frac{p+3}{8})$ $\mathcal{W} = (\frac{p+11}{8}^{\frac{8p(v-1)}{p+11}}, 2^{\frac{p-1}{4}}, \frac{p-1}{4}^1, \frac{p+3}{4}^1)$	$v = pp_1^{m_1}p_2^{m_2}\dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $\frac{p+11}{8} \| (p_t - 1)$ $\gcd(p, p_t) = 1$, for $1 \leq t \leq r$, $p = 25 + 4s^2$ or $p = 49 + s^2$	[65]
$\big((v,\ \mathcal{W},\ 2e^2+1)$ $\mathcal{W} = ((2e^2+2)^{\frac{pv-p}{2e^2+2}}, 2^{3s^2}, (s^2)^1, (s^2+1)^1)$	$v = pp_1^{m_1}p_2^{m_2}\dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $(2e^2+2) \| (p_t - 1)$ $\gcd(p, p_t) = 1$, for $1 \leq t \leq r$, $p = 1 + 8s^2 = 9 + 64e^2$	[65]
$\Big((v,\ (\frac{3p+1}{8}^{\frac{8p(v-1)}{3p+1}}, \frac{p-1}{4}^1, \frac{p+3}{4}^1, \frac{p-1}{2}^1),\ \frac{3p-7}{8}\Big)$	$v = pp_1^{m_1}p_2^{m_2}\dots p_r^{m_r}$ $2 < p_1 < p_2 < \cdots < p_r$, $\frac{3p+1}{8} \| (p_t - 1)$ $\gcd(p, p_t) = 1$, for $1 \leq t \leq r$, $p = 4 + s^2$	[65]
$\Big((v,\ (\frac{3p+5}{8}^{\frac{8p(v-1)}{3p+5}}, \frac{p-1}{4}^2, \frac{p+1}{2}^1),\ \frac{3p-3}{8}\Big)$	$v = pp_1^{m_1}p_2^{m_2}\dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $\frac{3p+5}{8} \| (p_t - 1)$ $\gcd(p, p_t) = 1$, for $1 \leq t \leq r$, $p = 9 + 4s^2$	[65]
$\Big(p,\ (2^{\frac{p-1}{4}}, \frac{p-1}{4}^1, \frac{p+3}{4}^1),\ \frac{p+3}{8}\Big)$	$p = 4t^2 + m^2$, and $m = 5, 7$	[92]

Herein p, p_i's and q_i's are primes; t, e, s, r and m are positive integers; q is a prime power;

We use $\mathcal{W}^*(l)$ to denote the case where $|\mathcal{W}| = l$ but the distribution of $\mathcal{W}$ is not clear;

For PDFs over noncyclic groups, the reader is referred to, for example, Table 1 in [65].

Table 3.3: Some known PDFs over cyclic groups with parameters $(v, \mathcal{W}, \lambda)$

Parameters	Constraints	Ref.
$\left(p,\ ({\frac{p-1}{4}}^1, {\frac{p+3}{4}}^1, {\frac{p-1}{2}}^1),\ \frac{3p-7}{8}\right)$	$p = 4 + m^2$, and $m \equiv 1 \pmod 4$	[92]
$\left(p,\ (2^{3t^2}, (t^2)^1, (t^2+1)^1),\ 2^m + 1\right)$	$p = 8t^2 + 1 = 64m^2 + 9$	[92]
$\left(p,\ ({\frac{p-1}{4}}^2, {\frac{p+1}{2}}^1),\ \frac{3p-3}{8}\right)$	$p = 4t^2 + 9$	[92]
$\left(p,\ (1^1, t^{\frac{p-1}{t}}),\ t-1\right)$	$p \equiv 1 \pmod t$, and $t = 3, 4$	[99]
$\left(p,\ \mathcal{W},\ \frac{t^2+m}{2}\right)$ $\mathcal{W} = (2^{t^2+m-1}, (t^2+m-1)^1, (t^2+m)^1)$	$p = 4t^2 + m^2$, and $m = 1, 3$	[99]
$\left(p,\ (f^m, (mf+1)^1),\ \frac{(m+1)f}{2}\right)$	$p = ef + 1$, $p \equiv 3 \pmod 4, e = 2m$	[99]
$\left(p,\ (6^{2t}, (12n+1)^1),\ 6t+3\right)$	$p = 24t + 1 \pmod 4$ $(-3)^{6t} \not\equiv 1 \pmod p$	[99]
$(q^r - 1,\ \mathcal{W}^*(q^s),\ q^{r-s} - 1)$	$1 \le s \le r$	[101]
$\left(t\frac{q^r-1}{m},\ \mathcal{W}^*(q^s),\ t\frac{q^{r-s}-1}{m}\right)$	$m \mid (q-1)$, $\gcd(m, r) = 1$, $1 \le t \le m$, $1 \le s \le r$	[101]

Herein p, p_i's and q_i's are primes; t, e, s, r and m are positive integers; q is a prime power;

We use $\mathcal{W}^*(l)$ to denote the case where $|\mathcal{W}| = l$ but the distribution of $\mathcal{W}$ is not clear;

For PDFs over noncyclic groups, the reader is referred to, for example, Table 1 in [65].

Definition 3.1.9 ([81]). *Let $\mathcal{B} = \{B_i : 1 \leq i \leq m\}$ be a family of disjoint subsets of G. Then $\mathcal{B}$ is a bounded external difference family (BEDF) if each nonzero element of G appears at most λ times in the union of multi-sets $D(B_i, B_j)$ for $1 \leq i \neq j \leq m$, i.e.,*

$$\bigcup_{1 \leq i \neq j \leq m} D(B_i, B_j) \subseteq \lambda \boxtimes (G \backslash \{0\}).$$

We briefly denote $\mathcal{B}$ as an (n, m, K, λ)-BEDF, where $K = (|B_1|, |B_2|, \ldots, |B_m|)$.

Example 3.1.10. *Let $n = 13$, $m = 2$, and $G = (\mathbb{Z}_{13}, +)$, then*

$$\mathcal{B}_1 = \{\{1\}, \{2, 6, 8, 10, 12\}\}$$

is a $(13, 2, (1, 5), 1)$-BEDF and

$$\mathcal{B}_2 = \{\{2, 4\}, \{11, 12\}\}$$

is a $(13, 2, (2, 2), 1)$-BEDF.

It is easy to check that EDFs are a special case of BEDFs according to Definitions 3.1.7 and 3.1.9.

Fact 3.1.11. *If $\mathcal{B}$ is an (n, m, K, λ)-EDF over group G, then $\mathcal{B}$ is also an (n, m, K, λ)-BEDF over group G.*

To construct AMD codes, in [81], the following generalizations of EDFs were also introduced by distinguishing the differences $D(B_i, B_j)$ and $D(B_j, B_i)$.

Definition 3.1.12 ([81]). *Let $\mathcal{B} = \{B_i : 1 \leq i \leq m\}$ be a family of disjoint subsets of G. $\mathcal{B}$ is called an $(n, m; k_1, k_2, \ldots, k_m; \lambda_1, \lambda_2, \ldots, \lambda_m)$-generalized strong external difference family (GSEDF) if for any given $1 \leq i \leq m$, each nonzero element of G appears exactly λ_i times in the union of multi-sets $D(B_i, B_j)$ for $1 \leq j \neq i \leq m$, i.e.,*

$$\bigcup_{\{j: 1 \leq j \leq m, j \neq i\}} D(B_i, B_j) = \lambda_i \boxtimes (G \backslash \{0\}), \tag{3.1}$$

where $k_i = |B_i|$ for $1 \leq i \leq m$. Furthermore, if $k = k_1 = k_2 = \cdots = k_m$ and $\lambda = \lambda_1 = \lambda_2 = \cdots = \lambda_m$, then it is named as a strong external difference family, also denoted as an (n, m, k, λ)-SEDF for short.

Herein, we include the first SEDF with $m \geq 5$ found in [57, 95] as an example.

Example 3.1.13 ([57, 95]). *Let $n = 243$, $m = 11$, and $G = (\mathbb{F}_{3^5}, +)$, then there exists a $(243, 11, 22, 20)$-SEDF.*

By Definitions 3.1.7 and 3.1.12, we know that SEDFs are a special case of EDFs.

Fact 3.1.14. *If $\mathcal{B}$ is an $(n, m; k_1, k_2, \ldots, k_m;\ \lambda_1, \lambda_2, \ldots, \lambda_m)$-GSEDF then $\mathcal{B}$ is an $(n, m, (k_1, k_2, \ldots, k_m), \sum_{1\le i\le m}\lambda_i)$-EDF. Particularly, if $\mathcal{B}$ is an (n, m, k, λ)-SEDF then $\mathcal{B}$ is an $(n, m, k, m\lambda)$-EDF.*

Known results of SEDFs are summarized in the following lemma.

Theorem 3.1.15 ([4, 54, 57, 81])**.** *An (n, m, k, λ)-SEDF over a group G exists if:*

I $(n, m, k, \lambda) = (k^2 + 1, 2, k, 1)$ *with* $G = (\mathbb{Z}_{k^2+1}, +)$ *[81];*

II $(n, m, k, \lambda) = (n, 2, \frac{n-1}{2}, \frac{n-1}{4})$ *and* $n \equiv 1 \pmod 4$*, when partial different set over G with parameters* $(n, \frac{n-1}{2}, \frac{n-5}{4}, \frac{n-1}{4})$ *exists [54];*

III $(n, m, k, \lambda) = (p, 2, \frac{p-1}{4}, \frac{p-1}{16})$*, when t is an integer, $p = 16t^2 + 1$ is a prime, and $G = (\mathbb{Z}_p, +)$ [4];*

IV $(n, m, k, \lambda) = (p, 2, \frac{p-1}{6}, \frac{p-1}{36})$*, when t is an integer, $p = 108t^2 + 1$ is a prime, and $G = (\mathbb{Z}_p, +)$ [4];*

V $(n, m, k, \lambda) = (243, 11, 22, 20)$ *with* $G = (\mathbb{Z}_3^5, +)$ *[81].*

Remark 3.1.16. *For SEDFs, there are also some papers focusing on the nonexistence. For this part, refer to [4, 54, 57] for examples.*

In [81], GSEDFs are further generalized as follows.

Definition 3.1.17 ([81])**.** *Let $\mathcal{B} = \{B_i : 1 \le i \le m\}$ be a family of disjoint subsets of G. Then $\mathcal{B}$ forms an $(n, m; k_1, k_2, \ldots, k_m;\ \lambda_1, \lambda_2, \ldots, \lambda_m)$-bounded generalized strong external difference family (BGSEDF) if for any given $1 \le i \le m$, each nonzero element of G appears at most λ_i times in the union of multi-sets $D(B_i, B_j)$ for $1 \le j \ne i \le m$, i.e.,*

$$\bigcup_{\{j:1\le j\le m,\, j\ne i\}} D(B_i, B_j) \subseteq \lambda_i \boxtimes (G\backslash\{0\}), \tag{3.2}$$

where $k_i = |B_i|$ for $1 \le i \le m$.

Definition 3.1.18 ([81])**.** *Let $\mathcal{B} = \{B_i : 1 \le i \le m\}$ be a family of disjoint subset-s of G. Then $\mathcal{B}$ is an $(n, m; c_1, c_2, \ldots, c_l; w_1,\ w_2, \ldots, w_l; \lambda_1, \lambda_2, \ldots, \lambda_l)$-partitioned external difference family (PEDF) if for any given $1 \le t \le l$,*

$$\bigcup_{\{i\,:\,|B_i|=w_t\}} \bigcup_{\{j:1\le j\le m,\, j\ne i\}} D(B_i, B_j) = \lambda_t \boxtimes (G\backslash\{0\}), \tag{3.3}$$

where $c_t = |\{i\ :\ |B_i| = w_t,\ 1 \le i \le m\}|$ for $1 \le t \le l$.

Example 3.1.19. *Let $n = 5$, $m = 2$, and $G = (\mathbb{Z}_5, +)$, then $\mathcal{B} = \{\{0\}, \{1\}, \{2, 4\}\}$ is a $(5, 2; 2, 1; 1, 2; 1, 1)$-PEDF.*

According to Definitions 3.1.12 and 3.1.18, GSEDFs are a special case of PEDFs.

Fact 3.1.20. *If $\mathcal{B}$ is an $(n, m; k_1, k_2, \ldots, k_m; \lambda_1, \lambda_2, \ldots, \lambda_m)$-GSEDF, then it is also an $(n, m; 1, 1, \ldots, 1; k_1, k_2, \ldots, k_m; \lambda_1, \lambda_2, \ldots, \lambda_m)$-PEDF.*

To finish this section, we summarize the relationship among the above mentioned difference families in Fig. 3.1.

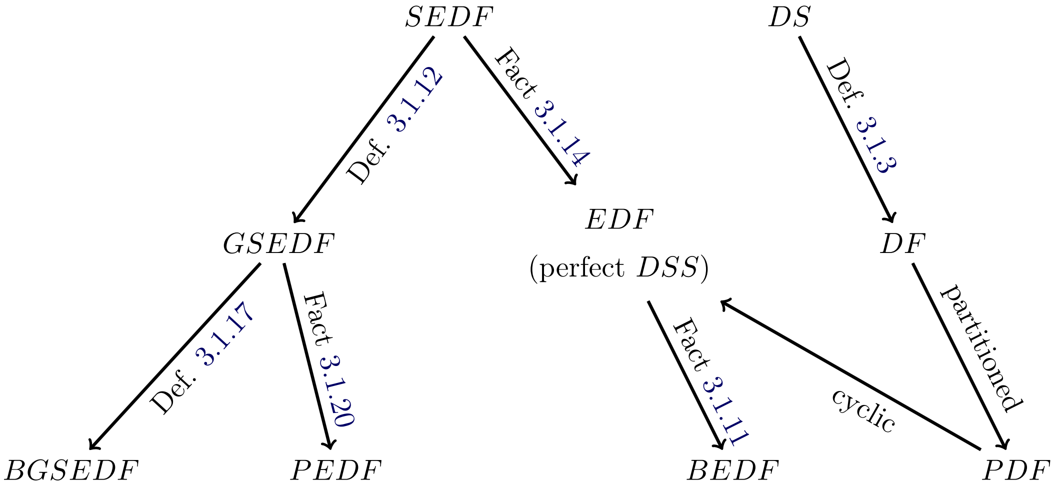

Figure 3.1: Summary of the relationship among DS, DF, PDF, EDF, $PEDF$, $BEDF$, $SEDF$, $GSDEF$ and $BGSEDF$.

3.2 A direct construction of PDFs

In this section, we shall introduce a direct construction of PDFs based on the Chinese Remainder Theorem [31] and a kind of cyclotomy.

For a positive integer v with $v > 1$, let $\mathbb{Z}_v^*$ denote the set of invertible elements in $\mathbb{Z}_v$, i.e., $\mathbb{Z}_v^* = \{0 \leq i \leq v-1 \ : \ \gcd(i, n) = 1\}$, and let $\varphi(v)$ be the Euler's totient function, i.e.,

$$\varphi(v) = |\mathbb{Z}_v^*| = |\{0 \leq i \leq v-1 \ : \ \gcd(i, n) = 1\}|.$$

It is well known that $(\mathbb{Z}_v^*, \cdot)$ is a group with order $\varphi(v)$, where the multiplication is given by $\mathbb{Z}_v$, i.e, $ab = ab \pmod{v}$, for $a, b \in \mathbb{Z}_v^*$. For $a \in \mathbb{Z}_v^*$, define the order of a in the group $(\mathbb{Z}_v^*, \cdot)$, i.e., the smallest integer $t > 0$ such that

$$a^t \equiv 1 \pmod{v}$$

to be the *multiplicative order* of a modulo v. For a positive integer v with $v > 1$, an element $a \in \mathbb{Z}_v^*$ is called a primitive root modulo v if the multiplicative order of a

modulo v is $\varphi(v)$. It is well known that if there is a primitive root modulo v, then v can only equal 2, 4, p^t, or $2p^t$, where p is an odd prime and t is a positive integer [3]. Furthermore, for any odd prime p, there exists at least one primitive root g modulo p which is also a primitive root modulo p^t for all $t \geq 1$ [3].

According to the unique factorization theorem, a positive integer n has the following unique decomposition

$$n = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}, \tag{3.4}$$

where $p_1 < p_2 < \cdots < p_k$ are primes and $m_1, m_2, \ldots, m_k$ are positive integers. The cardinality of the multiplicative group $\mathbb{Z}_n^*$ is equal to

$$\varphi(n) = \prod_{i=1}^{k} p_i^{m_i - 1}(p_i - 1).$$

From now on, we always suppose that n is an odd integer with $p_i \equiv 1 \pmod 8$ for each $1 \leq i \leq k$, where p_i is given in (3.4). Then we know that $\gcd(n, 7) = 1$ and 8 is a common factor of $p_1 - 1, p_2 - 1, \ldots, p_k - 1$. Write $p_i = 8f_i + 1$ for k positive integers f_i with $1 \leq i \leq k$. For $i = 1, 2, \ldots, k$, let g_i be a primitive root modulo p_i^t for all $t \geq 1$, i.e., the multiplicative order of g_i modulo p_i^t is $\varphi\left(p_i^t\right) = p_i^{t-1}(p_i - 1)$ for $t \geq 1$. Since $p_1, p_2, \ldots, p_k$ are distinct primes, by the Chinese Remainder Theorem [31], there exists a unique $\alpha \in \mathbb{Z}_n$ such that

$$\alpha \equiv g_i^{f_i p_i^{m_i - 1}} \pmod{p_i^{m_i}} \text{ for all } 1 \leq i \leq k.$$

It is easy to check that the multiplicative order of α modulo n is 8, i.e., $G_n = \{\alpha^i : i \in \mathbb{Z}_8\}$ is a subgroup of $\mathbb{Z}_n^*$ with order 8. Therefore, there are $N = \varphi(n)/8$ elements $a_0 = 1, a_1, \ldots, a_{N-1} \in \mathbb{Z}_n^*$ such that

$$D_i^{(n)} = a_i G_n = \{a_i \alpha^j : 0 \leq j \leq 7\}, \quad 0 \leq i \leq N - 1,$$

are exactly all the cosets of G_n in $\mathbb{Z}_n^*$, i.e.,

$$\mathbb{Z}_n^* = \bigcup_{0 \leq i \leq N-1} D_i^{(n)} \tag{3.5}$$

and

$$D_i^{(n)} \bigcap D_j^{(n)} = \emptyset, \quad \text{for any } 0 \leq i \neq j \leq N - 1.$$

We remark here that the discussion above is still valid for any factor $n_1 > 1$ of n.

Define a collection of subsets of $\mathbb{Z}_7 \times \mathbb{Z}_n^*$ as

$$\mathcal{B}_n = \{B_{i,j} : i \in \{0, 1, 2, 3\}, 0 \leq j \leq N - 1\}, \tag{3.6}$$

where for each $i \in \{0,1,2,3\}$ and $0 \le j \le N-1$,

$$\begin{aligned} B_{i,j} = \big\{&(1, a_j\alpha^i), (1, a_j\alpha^{4+i}), (2, a_j\alpha^{1+i}), (2, a_j\alpha^{5+i}), \\ &(0, a_j\alpha^{2+i}), (0, a_j\alpha^{6+i}), (4, a_j\alpha^{3+i}), (4, a_j\alpha^{7+i})\big\}. \end{aligned} \tag{3.7}$$

The following results will be needed in the sequel.

Lemma 3.2.1. *The sets in $\mathcal{B}_n$ form a partition of $\{0,1,2,4\} \times \mathbb{Z}_n^*$.*

Proof. The conclusion follows directly from (3.5), (3.6), and (3.7). □

Lemma 3.2.2. *With the notation as above,*

$$\bigcup_{B \in \mathcal{B}_n} D(B) = 4 \boxtimes (\mathbb{Z}_7 \times \mathbb{Z}_n^*).$$

Proof. Note that $\alpha^4 \equiv -1 \pmod{n}$ since the multiplicative order of α modulo n is 8. Thus, for given $i \in \{0,1,2,3\}$ and $0 \le j \le N-1$, $B_{i,j}$ can be rewritten as

$$\begin{aligned} B_{i,j} = \big\{&(1, a_j\alpha^i), (1, -a_j\alpha^i), (2, a_j\alpha^{i+1}), (2, -a_j\alpha^{i+1}), \\ &(0, a_j\alpha^{i+2}), (0, -a_j\alpha^{i+2}), (4, a_j\alpha^{i+3}), (4, -a_j\alpha^{i+3})\big\}. \end{aligned}$$

This means that for $T = \{0,1,2,3\}$,

$$\begin{aligned} &\bigcup_{i \in T} D(B_{i,j}) \\ &= \left(\{0\} \times \bigcup_{i \in T} \pm 2a_j\alpha^i\{1,\ \alpha,\ \alpha^2,\ \alpha^3\}\right) \\ &\quad \bigcup \left(\{\pm 1\} \times \bigcup_{i \in T} \pm a_j\alpha^i\{\alpha-1,\ \alpha+1, \alpha^2-1,\ \alpha^2+1\}\right) \\ &\quad \bigcup \left(\{\pm 2\} \times \bigcup_{i \in T} \pm a_j\alpha^i\{\alpha-\alpha^2,\ \alpha+\alpha^2,\ \alpha^3-\alpha,\ \alpha^3+\alpha\}\right) \\ &\quad \bigcup \left(\{\pm 3\} \times \bigcup_{i \in T} \pm a_j\alpha^i\{\alpha^2-\alpha^3,\ \alpha^2+\alpha^3, \alpha^3-1,\ \alpha^3+1\}\right) \\ &= \left(4 \boxtimes \left(\{0\} \times 2D_j^{(n)}\right)\right) \bigcup \left(\bigcup_{\substack{i \in \{1,2\} \\ \beta \in A}} \left(\{\pm i\} \times \beta D_j^{(n)}\right)\right) \\ &\quad \bigcup \left(\bigcup_{\beta \in \{\alpha-1, \alpha+1\}} \left(2 \boxtimes \{\pm 3\} \times \beta D_j^{(n)}\right)\right), \end{aligned} \tag{3.8}$$

where $A = \{\alpha-1, \alpha+1, \alpha^2-1, \alpha^2+1\}$, and the last equality holds by the fact that $\alpha^r \bigcup_{i \in T} \{\pm a_j\alpha^i\} = \alpha^r D_j^{(n)} = D_j^{(n)}$ for any positive integer r. Therefore, by (3.5) and

(3.8),

$$\begin{aligned}\bigcup_{B\in\mathcal{B}_n} D(B) = &\bigcup_{0\leq j\leq N-1}\left(\left(4\boxtimes\left(\{0\}\times 2D_j^{(n)}\right)\right)\bigcup\left(\bigcup_{\substack{i\in\{1,2\}\\ \beta\in A}}\left(\{\pm i\}\times\beta D_j^{(n)}\right)\right)\right.\\ &\left.\bigcup\left(\bigcup_{\beta\in\{\alpha-1,\alpha+1\}}\left(2\boxtimes\{\pm 3\}\times\beta D_j^{(n)}\right)\right)\right)\\ =&\left(4\boxtimes\left(\{0\}\times\mathbb{Z}_n^*\right)\right)\bigcup\left(\bigcup_{\substack{i\in\{1,2\}\\ \beta\in A}}\left(\{\pm i\}\times\beta\mathbb{Z}_n^*\right)\right)\\ &\bigcup 2\boxtimes\left(\bigcup_{\beta\in\{\alpha-1,\alpha+1\}}\left(\{\pm 3\}\times\beta\mathbb{Z}_n^*\right)\right)\\ =&4\boxtimes\left(\mathbb{Z}_7\times\mathbb{Z}_n^*\right),\end{aligned}$$

where the last equality follows from the fact that $2\in\mathbb{Z}_n^*$, and

$$A=\{\alpha-1,\alpha+1,\alpha^2-1,\alpha^2+1\}\subset\mathbb{Z}_n^*$$

since the order of $\alpha\in\mathbb{Z}_n^*$ is 8. This completes the proof of this lemma. □

Again, we remark that the discussion above is also valid for any factor $n_1>1$ of n.

Note that

$$\mathbb{Z}_n\setminus\{0\}=\bigcup_{1<n_1,\, n_1|n}\frac{n}{n_1}\mathbb{Z}_{n_1}^*. \tag{3.9}$$

Let

$$\begin{aligned}\Psi_{(7,n)}=&\left\{\left(1,\frac{n}{n_1}\right)B : 1<n_1,\ n_1\mid n,\text{ and } B\in\mathcal{B}_{n_1}\right\}\\ &\bigcup\{0,1,2,3,4\}\times\{0\}\bigcup\{(3,1),(5,1)\}\bigcup\{(3,2),(6,2)\},\end{aligned}\tag{3.10}$$

where $(1,\frac{n}{n_1})B=\{(i,\frac{n}{n_1}a):(i,a)\in B\}$ is a subset of $\mathbb{Z}_7\times\mathbb{Z}_n$.

Lemma 3.2.3. *With notation as above, we have*

$$\bigcup_{B\in\Psi_{(7,n)}} B=\{0,1,2,4\}\times\mathbb{Z}_n\bigcup\{(3,0),(3,1),(5,1),(3,2),(6,2)\} \tag{3.11}$$

and

$$\bigcup_{B\in\Psi_{(7,n)}} D(B)=4\boxtimes\left(\mathbb{Z}_7\times\mathbb{Z}_n\setminus\{(0,0)\}\right).$$

Proof. By Lemma 3.2.1, we can derive (3.11) directly from (3.10). For $\bigcup\limits_{B\in\Psi_{(7,n)}} D(B)$, Lemma 3.2.2 implies that

$$\begin{aligned}
&\bigcup_{B\in\Psi_{(7,n)}} D(B)\\
=&\left(\bigcup_{1<n_1,n_1|n} 4\boxtimes\left(\left(1,\frac{n}{n_1}\right)\mathbb{Z}_7\times\mathbb{Z}_{n_1}^*\right)\right)\bigcup D(\{0,1,2,3,4\}\times\{0\})\\
&\bigcup D(\{(3,1),(5,1)\})\bigcup D(\{(3,2),(6,2)\})\\
=&4\boxtimes(\mathbb{Z}_7\times(\mathbb{Z}_n\backslash\{0\}))\bigcup D(\{0,1,2,3,4\}\times\{0\})\\
&\bigcup D(\{(3,1),(5,1)\})\bigcup D(\{(3,2),(6,2)\})\\
=&4\boxtimes(\mathbb{Z}_7\times\mathbb{Z}_n\backslash\{(0,0)\}),
\end{aligned}$$

where the last two equalities follow from (3.9) and

$$\begin{aligned}
&D(\{0,1,2,3,4\}\times\{0\})\bigcup D(\{(3,1),(5,1)\})\bigcup D(\{(3,2),(6,2)\})\\
=&4\boxtimes((\mathbb{Z}_7\backslash\{0\})\times\{0\}),
\end{aligned}$$

respectively. □

Based on (3.11), a partition of $\mathbb{Z}_7\times\mathbb{Z}_n$ is given by

$$\Psi_{(7,n)}^*=\Psi_{(7,n)}\bigcup\{\{(a,b)\}:(a,b)\in\mathbb{Z}_7\times\mathbb{Z}_n,\ (a,b)\notin B\text{ for any }B\in\Psi_{(7,n)}\}.\quad(3.12)$$

According to the discussion above, it is easy to verify that $|\Psi_{(7,n)}|=(n-1)/2+3$ and $|\Psi_{(7,n)}^*|=|\Psi_{(7,n)}|+3n-5=(7n-5)/2$.

Theorem 3.2.4. *$\Psi_{(7,n)}^*$ is a $(7n,(1^{3n-5},2^2,\ 5^1,8^{(n-1)/2}),4)$-PDF over $\mathbb{Z}_{7n}$, where $n=p_1^{m_1}p_2^{m_2}\cdots p_k^{m_k}$, $2<p_1<p_2<\cdots<p_k$ are primes with $p_i\equiv 1 \pmod 8$ for each $1\le i\le k$, and $m_1,m_2,\ldots,m_k$ are positive integers.*

Proof. According to (3.12) and Lemma 3.2.3,

$$\bigcup_{B\in\Psi_{(7,n)}^*} D(B)=\bigcup_{B\in\Psi_{(7,n)}} D(B)=4\boxtimes(\mathbb{Z}_7\times\mathbb{Z}_n\setminus\{(0,0)\}).$$

Therefore, the desired conclusion follows from the fact for $B\in\Psi_{(7,n)}^*$,

$$|B|=\begin{cases}1, & 3n-5\text{ times},\\ 2, & 2\text{ times},\\ 5, & 1\text{ time},\\ 8, & \frac{n-1}{2}\text{ times}.\end{cases}$$

□

Remark 3.2.5. *In [13, 14, 39], PDFs were respectively generated by cosets of subgroups for $\mathbb{Z}_n^*$ and $\mathbb{F}^*_{p_1^{m_1}} \times \mathbb{F}^*_{p_2^{m_2}} \times \cdots \times \mathbb{F}^*_{p_k^{m_k}}$ as well. However, the constructions in [13, 14] can produce PDFs only in the case $(\lambda+1)|6$ if $n = 7p_1^{m_1}p_2^{m_2}\cdots p_k^{m_k}$, and the construction in [39] can generate PDFs over cyclic groups only in the case $m_i = 1$ for $1 \le i \le k$ and $(\lambda+1)|6$ if $n = 7p_1^{m_1}p_2^{m_2}\cdots p_k^{m_k}$. Thus, our construction can yield PDFs with new parameters in contrast to those constructions.*

Example 3.2.6. *Let $n = 17$. By our construction, we can obtain a $(119, (1^{46}, 2^2, 5^1, 8^8))$-PDF over $\mathbb{Z}_{119}$ as*

$\{\{1, 9, 98, 32, 50, 93, 21, 53\},$
$\{86, 77, 81, 15, 16, 42, 4, 36\},$
$\{35, 60, 64, 100, 84, 25, 106, 2\},$
$\{18, 43, 30, 49, 67, 8, 72, 70\},$
$\{52, 103\},\ \{87, 104\},$
$\{3\},\ \{5\},\ \{6\},\ \{10\},\ \{12\},\ \{13\},$
$\{33\},\ \{34\},\ \{38\},\ \{40\},\ \{41\},\ \{45\},$
$\{62\},\ \{66\},\ \{68\},\ \{69\},\ \{73\},\ \{75\},$
$\{94\},\ \{96\},\ \{97\},\ \{101\},\ \{108\},\ \{110\},$
$\{71, 44, 56, 11, 99, 58, 63, 74\},$
$\{37, 112, 39, 113, 65, 7, 46, 57\},$
$\{105, 95, 22, 79, 14, 109, 29, 23\},$
$\{88, 78, 107, 28, 116, 92, 114, 91\},$
$\{0, 85, 51, 17, 102\},$
$\{19\},\ \{20\},\ \{24\},\ \{26\},\ \{27\},\ \{31\},$
$\{47\},\ \{48\},\ \{54\},\ \{55\},\ \{59\},\ \{61\},$
$\{76\},\ \{80\},\ \{82\},\ \{83\},\ \{89\},\ \{90\},$
$\{111\},\ \{115\},\ \{117\},\ \{118\}\}.$

3.3 DSSs from PDFs

In this section, we shall introduce a new construction of DSSs from PDFs based on the following construction of optimal and perfect DSSs from PDFs developed by Ding [37] (refer to Fact 3.1.6).

The construction in Fact 3.1.6 is generic in the sense that it works for any PDF over a cyclic group. Based on this generic construction, several classes of optimal DSSs were obtained via PDFs in the literature (see [13, 14, 37, 39, 100, 101], for example). Similarly, PDFs proposed in Section 3.2 also generate optimal DSSs under this framework.

However, all those previous DSSs constructed from PDFs are of partitioned-type (i.e., $\mathbb{Z}_v = \cup_{i=0}^{l-1} S_i$ and then $\sum_{0 \le i \le l-1} \tau_i = v$), thus they can only lead to comma-free codes of zero code rate. Then a natural question one would ask is whether the construction in Fact 3.1.6 can be modified to obtain DSSs leading to comma-free codes with positive code rate. In the following, we shall answer this problem in the affirmative.

Theorem 3.3.1. *Let $\mathcal{S}' = \{S_i \ : \ 0 \le i \le l-1\}$ be a (v, l, λ)-PDF over $\mathbb{Z}_v$,*

$$\mathcal{S} = \{S_i : i \in I\}$$

and

$$T = \bigcup_{i \notin I} S_i, \tag{3.13}$$

where $I \triangleq \{i : 0 \leq i \leq l-1, |S_i| > 1\} = \{i_j : 1 \leq j \leq |I|\}$*. If each element of* $\mathbb{Z}_v \setminus \{0\}$ *appears at least* r *times in the multiset* $D(T)$*, i.e.,* $r \boxtimes (\mathbb{Z}_v \setminus \{0\}) \subseteq D(T)$*, then* $\mathcal{S}$ *is a* $(v, (|S_{i_1}|, |S_{i_2}|, \cdots, |S_{i_{|I|}}|), v + r - \lambda - 2(l - |I|))$ *DSS with code rate* $\frac{|T|}{v} = \frac{l-|I|}{v}$.

Proof. It is clear that $\mathcal{S}$ is nonempty since $v > l$. Recall that $\mathcal{S}' = \{S_i : 0 \leq i \leq l-1\}$. Then $\mathcal{S} \subseteq \mathcal{S}'$ and $|S_i| = 1$ for any $S_i \in \mathcal{S}' \setminus \mathcal{S}$. Since S' is a (v, l, λ)-PDF over $\mathbb{Z}_v$, we have

$$\bigcup_{S_i \in \mathcal{S}} D(S_i) = \bigcup_{S_i \in \mathcal{S}'} D(S_i) = \lambda \boxtimes (\mathbb{Z}_v \setminus \{0\}). \tag{3.14}$$

Since $\mathcal{S}'$ is a partition of $\mathbb{Z}_v$, we have

$$\begin{aligned} D(\mathbb{Z}_v) &= v \boxtimes (\mathbb{Z}_v \setminus \{0\}) \\ &= \left(\bigcup_{\substack{S_i, S_j \in \mathcal{S}' \\ S_i \neq S_j}} D(S_i, S_j) \right) \bigcup \left(\bigcup_{S_i \in \mathcal{S}'} D(S_i) \right) \\ &= \left(\bigcup_{\substack{S_i, S_j \in \mathcal{S}' \\ S_i \neq S_j}} D(S_i, S_j) \right) \bigcup \left(\bigcup_{S_i \in \mathcal{S}} D(S_i) \right). \end{aligned}$$

This, together with (3.14), leads to

$$\bigcup_{\substack{S_i, S_j \in \mathcal{S}' \\ S_i \neq S_j}} D(S_i, S_j) = (v - \lambda) \boxtimes (\mathbb{Z}_v \setminus \{0\}). \tag{3.15}$$

On the other hand, we have

$$\begin{aligned} \bigcup_{\substack{S_i, S_j \in \mathcal{S}' \\ S_i \neq S_j}} D(S_i, S_j) = &\left(\bigcup_{\substack{S_i, S_j \in \mathcal{S}' \setminus \mathcal{S} \\ S_i \neq S_j}} D(S_i, S_j) \right) \bigcup \left(\bigcup_{\substack{S_i \in \mathcal{S}' \setminus \mathcal{S} \\ S_j \in \mathcal{S}}} D(S_i, S_j) \cup D(S_j, S_i) \right) \\ &\bigcup \left(\bigcup_{\substack{S_i, S_j \in \mathcal{S} \\ S_i \neq S_j}} D(S_i, S_j) \right). \end{aligned} \tag{3.16}$$

Recall that $T = \bigcup_{S_i \in \mathcal{S}' \setminus \mathcal{S}} S_i$ and $|S_i| = 1$ for $S_i \in \mathcal{S}' \setminus \mathcal{S}$, i.e.,

$$D(T) = \bigcup_{\substack{S_i, S_j \in \mathcal{S}' \setminus \mathcal{S} \\ S_i \neq S_j}} D(S_i, S_j).$$

Thus, by (3.16),

$$\begin{aligned}\left(\bigcup_{\substack{S_i,S_j\in\mathcal{S}'\\ S_i\neq S_j}} D(S_i,S_j)\right)\bigcup D(T) &= \left(\bigcup_{\substack{S_i\in\mathcal{S}'\setminus\mathcal{S}\\ S_j\in\mathcal{S}',S_j\neq S_i}} D(S_i,S_j)\cup D(S_j,S_i)\right)\\ &\quad \bigcup\left(\bigcup_{\substack{S_i,S_j\in\mathcal{S}\\ S_i\neq S_j}} D(S_i,S_j)\right)\\ &= 2|\mathcal{S}'\setminus\mathcal{S}|\boxtimes(\mathbb{Z}_v\setminus\{0\})\bigcup\left(\bigcup_{\substack{S_i,S_j\in\mathcal{S}\\ S_i\neq S_j}} D(S_i,S_j)\right),\end{aligned}$$

where the last identity holds since

$$\bigcup_{\substack{S_j\in\mathcal{S}'\\ S_j\neq S_i}} D(S_i,S_j)=\bigcup_{\substack{S_j\in\mathcal{S}'\\ S_j\neq S_i}} D(S_j,S_i)=\mathbb{Z}_v\setminus\{0\}$$

for each $S_i\in\mathcal{S}'\setminus\mathcal{S}$.

Combining (3.15), (3.16) with the fact that $r\boxtimes(\mathbb{Z}_v\setminus\{0\})\subseteq D(T)$, we arrive at

$$(v+r-\lambda-2|\mathcal{S}'\setminus\mathcal{S}|)\boxtimes(\mathbb{Z}_v\setminus\{0\})\subseteq\bigcup_{\substack{S_i,S_j\in\mathcal{S}\\ S_i\neq S_j}} D(S_i,S_j),$$

which implies that $\mathcal{S}$ is a $(v,(|S_{i_1}|,|S_{i_2}|,\cdots,|S_{i_{|I|}}|),v+r-\lambda-2(l-|I|))$ DSS, where $I=\{i_1,i_2,\cdots,i_{|I|}\}$. By Definition 2.6.2, the code rate of the DSS $\mathcal{S}$ is $\frac{|T|}{v}=\frac{l-|I|}{v}$, which completes the proof. □

Remark 3.3.2. *For the modified construction of DSSs in Theorem 3.3.1, we have the following comments.*

1) It becomes the original construction of Ding in [37] if $\tau_i\neq 1$ for each $0\leq i\leq l-1$.

2) It generates DSSs leading to comma-free codes with positive code rate provided that $|I|<l$.

3) The resultant DSSs can be optimal in many cases (e.g., Theorem 3.3.3).

Theorem 3.3.3. *With the notation in Theorem 3.3.1, the DSS $\mathcal{S}$ is optimal with respect to the bound in Lemma 2.6.3, if $|I|>1$ and*

$$\begin{aligned}&(|I|-1)(\lambda-1)v+((l-|I|)^2+1-\lambda)(|I|-1)\\ <&(v-\lambda-2(l-|I|))(v-1).\end{aligned}\tag{3.17}$$

Proof. Set $x = l - |I|$ and note that $|I| > 1$. It is easy to check that (3.17) holds if and only if

$$\begin{aligned}(v-x-1)^2 &< (v-\lambda-2x)(v-1)+\frac{(v-\lambda-2x)(v-1)}{|I|-1}\\ &\leq (v-\lambda-2x)(v-1)+\left\lceil\frac{(v-\lambda-2x)(v-1)}{|I|-1}\right\rceil.\end{aligned}$$

That is,

$$v-x-1 < \left\lceil\sqrt{(v-\lambda-2x)(v-1)+\left\lceil\frac{(v-\lambda-2x)(v-1)}{|I|-1}\right\rceil}\right\rceil.$$

This, together with (2.10), leads to

$$v-x = \left\lceil\sqrt{(v-\lambda-2x)(v-1)+\left\lceil\frac{(v-\lambda-2x)(v-1)}{|I|-1}\right\rceil}\right\rceil,$$

since $r \geq 0$ and $r_l(v,\rho) \leq \sum\limits_{S_i \in \mathcal{S}} |S_i| = v - x$. Therefore, $\mathcal{S}$ is optimal with respect to the bound in Lemma 2.6.3. □

As will be shown in the next subsection, the condition given by inequality (3.17) enables us to obtain infinite families of optimal DSSs, although it looks quite complex.

3.3.1 Some optimal DSSs with relatively high code rate

Now we focus on DSSs leading to comma-free codes with relatively high code rate based on the modified construction. To this end, we need PDFs with a large number of i such that $|S_i| = 1$ for $0 \leq i \leq l-1$. Unfortunately, most of the known PDFs do not satisfy this property. In what follows, we are going to analyze DSSs generated by known PDFs constructed in Theorem 3.2.4, Section 3.2.

At first, we employ the PDFs generated by Theorem 3.2.4 to yield DSSs. Before that, we need some properties about those PDFs.

Lemma 3.3.4. *For the PDFs generated by Theorem 3.2.4, we have*

$$T = \bigcup_{B \in \Psi^*_{(7,n)} \setminus \Psi_{(7,n)}} B = \{3,5,6\} \times \mathbb{Z}_n \setminus \{(3,0),(3,1),(5,1),(3,2),(6,2)\}$$

and

$$(n-4) \boxtimes (\mathbb{Z}_7 \times \mathbb{Z}_n \setminus \{(0,0)\}) \subset D(T),$$

where $\Psi_{(7,n)}$ and $\Psi^*_{(7,n)}$ are defined by (3.10) and (3.12), respectively.

Proof. Firstly, by (3.11), (3.12) and (3.13),

$$\begin{aligned}T = \bigcup_{B\in\Psi^*_{7,n}\setminus\Psi_{7,n}} B = (\mathbb{Z}_7\times\mathbb{Z}_n)\setminus\left(\bigcup_{B\in\Psi_{7,n}} B\right)\\ =&\{3,5,6\}\times\mathbb{Z}_n\setminus\{(3,0),(3,1),(5,1),(3,2),(6,2)\}.\end{aligned}$$

On one hand, the fact $D(\{3,5,6\}) = \mathbb{Z}_7\setminus\{0\}$ means that

$$\begin{aligned}&D(\{3,5,6\}\times\mathbb{Z}_n)\\ =&\left(\bigcup_{i\in\{3,5,6\}} D(\{i\}\times\mathbb{Z}_n)\right)\bigcup\left(\bigcup_{\substack{i,j\in\{3,5,6\}\\ i\neq j}} D(\{i\}\times\mathbb{Z}_n,\{j\}\times\mathbb{Z}_n)\right)\\ =&3n\boxtimes(\{0\}\times(\mathbb{Z}_n\backslash\{0\}))\bigcup n\boxtimes((\mathbb{Z}_7\setminus\{0\})\times\mathbb{Z}_n).\end{aligned}\tag{3.18}$$

Note that for any given $i\in\mathbb{Z}_n$, the multisets

$$\begin{aligned}\Lambda_3 =&\{\pm((3,i)-(a,b)) : (a,b)\neq(3,i),(a,b)\in\{3,5,6\}\times\mathbb{Z}_n\}\\ =&2\boxtimes(\{0\}\times(\mathbb{Z}_n\setminus\{0\}))\bigcup(\{2,5,3,4\}\times\mathbb{Z}_n),\end{aligned}\tag{3.19}$$

$$\begin{aligned}\Lambda_5 =&\{\pm((5,i)-(a,b)) : (a,b)\neq(5,i),(a,b)\in\{3,5,6\}\times\mathbb{Z}_n\}\\ =&2\boxtimes(\{0\}\times(\mathbb{Z}_n\setminus\{0\}))\bigcup(\{2,5,1,6\}\times\mathbb{Z}_n),\end{aligned}\tag{3.20}$$

$$\begin{aligned}\Lambda_6 =&\{\pm((6,i)-(a,b)) : (a,b)\neq(6,i),(a,b)\in\{3,5,6\}\times\mathbb{Z}_n\}\\ =&2\boxtimes(\{0\}\times(\mathbb{Z}_n\setminus\{0\}))\bigcup(\{1,6,3,4\}\times\mathbb{Z}_n).\end{aligned}\tag{3.21}$$

On the other hand,

$$D(\{3,5,6\}\times\mathbb{Z}_n)\subset D(T)\bigcup 3\boxtimes\Lambda_3\bigcup\Lambda_5\bigcup\Lambda_6.$$

Thus, according to (3.18)-(3.21), we can conclude that

$$\begin{aligned}&3n\boxtimes(\{0\}\times(\mathbb{Z}_n\setminus\{0\}))\bigcup n\boxtimes((\mathbb{Z}_7\setminus\{0\})\times\mathbb{Z}_n)\\ \subset&D(T)\bigcup 10\boxtimes(\{0\}\times(\mathbb{Z}_n\setminus\{0\}))\bigcup 2\boxtimes((\mathbb{Z}_7\setminus\{0\})\times\mathbb{Z}_n)\\ &\bigcup 2\boxtimes(\{2,5,3,4\}\times\mathbb{Z}_n)\\ \subset&D(T)\bigcup 10\boxtimes(\{0\}\times(\mathbb{Z}_n\setminus\{0\}))\bigcup 4\boxtimes((\mathbb{Z}_7\setminus\{0\})\times\mathbb{Z}_n),\end{aligned}$$

which means

$$(n-4)\boxtimes(\mathbb{Z}_7\times\mathbb{Z}_n\setminus\{(0,0)\})\subset D(T),$$

where we use the fact $n\geq 17$ to make sure $3n-10 > n-4$. □

Based on Theorem 3.2.4, Theorem 3.3.1, and Lemma 3.3.4, we obtain the following result.

Corollary 3.3.5. *If $n = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$, $8|(p_i - 1)$, and p_i's are prime factors of n for $1 \le i \le k$, then there exists a $(7n, (2^2, 5^1, 8^{(n-1)/2}), 2n+2)$ DSS with code rate $(3n-5)/(7n)$.*

Remark 3.3.6. *In [45], the authors pointed out that the optimal or asymptotically optimal DSSs that can yield comma-free codes with code rate higher than a half are quite rare (refer to Table II [45] for the detailed parameters). Herein, the comma-free codes constructed from the DSSs in Corollary 3.3.5 have code rate $\frac{3n-5}{7n}$. Although these DSSs in Corollary 3.3.5 are not optimal with respect to the bound in Lemma 2.6.3, they perform quite well in the sense that*

$$\lim_{n\to+\infty}\left(\frac{\sum\limits_{B\in\Psi_{(7,n)}}|B|}{r_l(v,\rho)}\right)^2 \le \lim_{n\to+\infty}\left(\frac{\sum\limits_{B\in\Psi_{(7,n)}}|B|}{\left\lceil\sqrt{\rho(v-1)+\left\lceil\frac{\rho(v-1)}{l-1}\right\rceil}\right\rceil}\right)^2 = \frac{8}{7},$$

where $v = 7n$, $\rho = 2n+2$, $l = (n-1)/2+3$, and $\sum\limits_{B\in\Psi_{(7,n)}}|B| = 4n+5$.

Example 3.3.7. *Based on the PDF in Example 3.2.6 and Theorem 3.3.1, a $(119, (2^2, 5^1, 8^8), 36)$ DSS with code rate $\frac{46}{119}$ can be given as*

$\mathcal{S} = \{S_0 = \{1, 9, 98, 32, 50, 93, 21, 53\}$, $S_1 = \{71, 44, 56, 11, 99, 58, 63, 74\}$,
$S_2 = \{86, 77, 81, 15, 16, 42, 4, 36\}$, $S_3 = \{37, 112, 39, 113, 65, 7, 46, 57\}$,
$S_4 = \{35, 60, 64, 100, 84, 25, 106, 2\}$, $S_5 = \{105, 95, 22, 79, 14, 109, 29, 23\}$,
$S_6 = \{18, 43, 30, 49, 67, 8, 72, 70\}$, $S_7 = \{88, 78, 107, 28, 116, 92, 114, 91\}$,
$S_8 = \{52, 103\}$, $S_9 = \{87, 104\}$, $S_{10} = \{0, 85, 51, 17, 102\}\}$.

According to Theorems 3.3.1 and 3.3.3, one can obtain the following optimal DSSs from known PDFs.

Definition 3.3.8. *Let $D = (d_{i,j})_{0\le i\le k-1, 0\le j\le m-1}$ be a $k \times m$ matrix with entries from $\mathbb{Z}_m$. If every element of $\mathbb{Z}_m$ occurs exactly once among the differences $d_{i_1,j} - d_{i_2,j}$, $0 \le j \le m-1$ for any $0 \le i_1 \ne i_2 \le k-1$, then D is called an $(m, k, 1)$ cyclic difference matrix, or $(m, k, 1)$ CDM for short.*

Theorem 3.3.9 ([9])**.** *Let $m = p_1^{m_1} p_2^{m_2} \cdots p_t^{m_t}$, where $p_1, p_2, \cdots, p_t$ are t distinct primes. Let λ be a nonnegative integer with $(\lambda+1)|(p_i - 1)$ for each $1 \le i \le t$.*

1) *There exists an $(nm, (1^{(n-w-1)m+1}, w^m, (\lambda+1)^{(m-1)/(\lambda+1)}), \lambda)$-PDF, if there exists an (n, w, λ) cyclic different set with $2 \le w < \min_{1\le i\le t} p_i$;*

2) *There exists an $(nm, (1^{n-w}, w^1, (\lambda+1)^{n(m-1)/(\lambda+1)}), \lambda)$-PDF, if there exist an $(n, \lambda+2, 1)$-CDM and an (n, w, λ) cyclic different set with $w \ge 2$.*

Corollary 3.3.10. *Let $m = p_1^{m_1} p_2^{m_2} \cdots p_t^{m_t}$, where $p_1, p_2, \ldots, p_t$ are t distinct primes. Let λ be a nonnegative integer with $(\lambda+1)|(p_i - 1)$ for each $1 \le i \le t$. Then there exists an $(nm, W = (w^1, (\lambda+1)^{n(m-1)/(\lambda+1)}), nm - 2(n-w) - \lambda)$ DSS with code rate $\frac{n-w}{nm}$, if there exist an $(n, \lambda+2, 1)$ CDM and an (n, w, λ) cyclic different set with $w \ge 2$. Furthermore, the DSS is optimal with respect to the bound in Lemma 2.6.3 when*

$$2n^2m^2 + am + b > 0, \tag{3.22}$$

where $a = n(-n^2 - n\lambda - 3n + 2wn - w^2 + 2w + 2w\lambda - \lambda^2 - \lambda - 2)$ and $b = n^3 - 2wn^2 + w^2n + \lambda n + 3n + \lambda^2 + \lambda - 2w\lambda - 2w$.

Remark 3.3.11. *Let $\delta = a^2 - 8n^2b$. If $\delta < 0$, then the inequality (3.22) holds for any m, which means that the DSS in Corollary 3.3.10 is optimal in this case. If $\delta > 0$, the inequality (3.22) holds for any $m > \frac{-a+\sqrt{a^2-8n^2b}}{4n^2}$, which implies that we can take a large enough m to guarantee that the DSS in Corollary 3.3.10 is optimal.*

Example 3.3.12. *Let $n = 4, m = 7$, and $\mathcal{S}_1 = \{\{0,3,5,6\}, \{1\}, \{2\}, \{4\}\}$. It is easy to check that $\mathcal{S}_1$ is a $(7, (1^3, 4), 2)$-PDF and $\{2,4,1\}$ is a $(7,3,1)$-DS over $\mathbb{Z}_7$. Note that a $(7,4,1)$ CDM can be given as*

$$D = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 2 & 4 & 6 & 1 & 3 & 5 \\ 0 & 3 & 6 & 2 & 5 & 1 & 4 \end{pmatrix}.$$

By Theorem 3.3.9, there exists a $(28, (1^3, 3^7, 4^1), 20)$-PDF:

$$\{\{4\}, \{8\}, \{16\}, \{25,22,19\}, \{1,2,3\}, \{5,10,15\}, \{9,18,27\}, \{13,26,11\}, \{17,6,23\}, \{21,14,7\}, \{0,12,20,24\}\}.$$

From S_1, S_2 and D, by Theorem 3.3.1, we obtain the following $(28, (3^7, 4^1), 20)$ DSS:

$$\mathcal{S} = \{\{25,22,19\}, \{1,2,3\}, \{5,10,15\}, \{9,18,27\}, \{13,26,11\}, \{17,6,23\}, \{21,14,7\}, \{0,12,20,24\}\}.$$

It is easy to check that

$$r_l(v,\rho) = \left\lceil \sqrt{\rho(v-1) + \left\lceil \frac{\rho(v-1)}{l-1} \right\rceil} \right\rceil = \left\lceil \sqrt{618} \right\rceil = 25.$$

Therefore, $\mathcal{S}$ is optimal with respect to the bound in (2.10), which is consistent with the result in Corollary 3.3.10.

3.4 Weighted external difference families

In this section, we further generalize external difference families to weighted external difference families.

Definition 3.4.1. *Let $\mathcal{B} = \{B_i : 1 \le i \le m\}$ be a family of disjoint subsets of G. Let $K = (k_1, k_2, \ldots, k_m)$ with $k_i = |B_i|$ for $1 \le i \le m$ and $\widetilde{k} = lcm(k_1, k_2, \ldots, k_m)$. Define $\widetilde{\mathcal{B}} = \{\widetilde{B}_i : B_i \in \mathcal{B}\}$ as the standard weighted multi-sets of $\mathcal{B}$, where*

$$\widetilde{B}_i \triangleq \frac{\widetilde{k}}{|B_i|} \boxtimes B_i = \frac{\widetilde{k}}{k_i} \boxtimes B_i.$$

Then $\mathcal{B}$ is called an (n, m, K, a, λ)-bounded standard weighted external difference family (BSWEDF) if λ is the smallest positive integer such that

$$\bigcup_{1 \le i \ne j \le m} D(B_i, \widetilde{B}_j) \subseteq \lambda \boxtimes (G \backslash \{0\}),$$

where $a = \sum_{1 \le i \le m} k_i$. Furthermore, if $\mathcal{B}$ satisfies

$$\bigcup_{1 \le i \ne j \le m} D(B_i, \widetilde{B}_j) = \lambda \boxtimes (G \backslash \{0\}),$$

then it is named as a standard weighted external difference family, also denoted as an (n, m, K, a, λ)-SWEDF for short.

For BSWEDFs and SWEDFs, we have the following facts on their parameters.

Lemma 3.4.2. *Let $\mathcal{B}$ be an (n, m, K, a, λ)-BSWEDF. Then we have*

$$\lambda \ge \left\lceil \frac{\widetilde{k}a(m-1)}{n-1} \right\rceil. \tag{3.23}$$

Specially, if $\mathcal{B}$ is an (n, m, K, a, λ)-SWEDF, then $(n-1) \mid (\widetilde{k}a(m-1))$ and

$$\lambda = \frac{\widetilde{k}a(m-1)}{n-1}. \tag{3.24}$$

Proof. Let $\mathcal{B} = \{B_i : 1 \le i \le m\}$. The fact

$$\bigcup_{1 \le i \ne j \le m} D(B_i, \widetilde{B}_j) = \bigcup_{1 \le i \ne j \le m} \bigcup_{b \in B_i} D(\{b\}, \widetilde{B}_j)$$

means that

$$\begin{aligned} \left| \bigcup_{1 \le i \ne j \le m} D(B_i, \widetilde{B}_j) \right| &= \sum_{1 \le i \le m} \sum_{\substack{1 \le j \le m \\ j \ne i}} \sum_{b \in B_i} |D(\{b\}, \widetilde{B}_j)| \\ &= \sum_{1 \le i \le m} \sum_{\substack{1 \le j \le m \\ j \ne i}} \sum_{b \in B_i} \widetilde{k} \\ &= \widetilde{k}a(m-1). \end{aligned} \tag{3.25}$$

Thus, we have $\lambda \geq \lceil \frac{\widetilde{k}a(m-1)}{n-1} \rceil$.

Similarly, for the case of SWEDFs, by Definition 3.4.1 and (3.25), we have $\lambda(n-1) = \widetilde{k}a(m-1)$, i.e., $\lambda = \frac{\widetilde{k}a(m-1)}{n-1}$, which also means $(n-1) \mid (\widetilde{k}a(m-1))$. □

Definition 3.4.3. *An (n, m, K, a, λ)-BSWEDF is said to be optimal if λ takes the smallest possible value for given n, m, and K.*

Specially, an (n, m, K, a, λ)-SWEDF is optimal if λ achieves the lower bound given by (3.24) with equality, i.e., $\lambda = \frac{\widetilde{k}a(m-1)}{n-1}$.

3.5 Constructions of bounded standard weighted external difference families

In this section, we are going to consider constructions for BSWEDFs and SWEDFs.

3.5.1 From external difference families to bounded standard weighted external difference families

We begin with the relationship among EDFs, SEDFs, PEDFs, SWEDFs, and BSWEDFs.

In general, an EDF is not necessarily an SWEDF. However, in the following cases, an EDF is always an SWEDF. First of all, we consider the regular case.

Lemma 3.5.1. *A regular (n, m, k, λ)-EDF forms an $(n, m, K = (k, k, \ldots, k), a = mk, \lambda)$-SWEDF.*

The lemma follows directly from the definitions of EDF and SWEDF.

For the case of GSEDFs we have the following result.

Lemma 3.5.2. *If $\{B_i : 1 \leq i \leq m\}$ is an $(n, m; k_1, k_2, \ldots, k_m; \lambda_1, \lambda_2, \ldots, \lambda_m)$-GSEDF, then $\{B_i : 1 \leq i \leq m\}$ is an $(n, m, (k_1, k_2, \ldots, k_m), a, \lambda)$-SWEDF, where $\lambda = \sum_{1 \leq i \leq m} \frac{\lambda_i \widetilde{k}}{k_i}$.*

Proof. Let $\{B_i : 1 \leq i \leq m\}$ be an $(n, m; k_1, k_2, \ldots, k_m; \lambda_1, \lambda_2, \ldots, \lambda_m)$-GSEDF, by (3.1),

$$\bigcup_{\{j:1\leq j\leq m, j\neq i\}} D(B_i, B_j) = \lambda_i \boxtimes (G \backslash \{0\}),$$

which means

$$\bigcup_{\{j:1\leq j\leq m, j\neq i\}} D(B_j, \widetilde{B}_i) = \frac{\lambda_i \widetilde{k}}{k_i} \boxtimes (G \backslash \{0\}).$$

Thus, we have

$$\begin{aligned}\bigcup_{1\leq i\leq m}\bigcup_{\{j:1\leq j\leq m,\, j\neq i\}} D(B_j,\widetilde{B}_i)=&\left(\sum_{1\leq i\leq m}\lambda_i\frac{\widetilde{k}}{k_i}\right)\boxtimes(G\backslash\{0\})\\ =&\lambda\boxtimes(G\backslash\{0\}),\end{aligned}$$

i.e., $\{B_i\,:\,1\leq i\leq m\}$ is an $(n,m,(k_1,k_2,\ldots,k_m),a,\lambda)$-SWEDF with $\lambda=\sum_{1\leq i\leq m}\frac{\lambda_i\widetilde{k}}{k_i}$. □

Similarly, the relationship between PEDFs and SWEDFs can be given by the following lemma.

Lemma 3.5.3. *If $\{B_i\,:\,1\leq i\leq m\}$ is an $(n,m;c_1,c_2,\ldots,c_l;w_1,w_2,\ldots,w_l;\lambda_1,\lambda_2,\ldots,\lambda_l)$-PEDF, then $\{B_i\,:\,1\leq i\leq m\}$ is an $(n,m,K=(|B_1|,|B_2|,\ldots,|B_m|),a,\lambda)$-SWEDF, where $\widetilde{k}=lcm(w_1,w_2,\ldots,w_l)$ and $\lambda=\sum_{1\leq t\leq l}\frac{\lambda_t\widetilde{k}}{w_t}$.*

Proof. Since $\{B_i\,:\,1\leq i\leq m\}$ is an $(n,m;c_1,c_2,\ldots,c_l;w_1,w_2,\ldots,w_l;\lambda_1,\lambda_2,\ldots,\lambda_l)$-PEDF, by (3.3),

$$\bigcup_{\{i\,:\,|B_i|=w_t\}}\bigcup_{\{j:1\leq j\leq m,\, j\neq i\}} D(B_i,B_j)=\lambda_t\boxtimes(G\backslash\{0\})$$

for $1\leq t\leq l$. By Definition 3.1.18, $|B_i|\in\{w_j\,:\,1\leq j\leq l\}$ for $1\leq i\leq m$. Thus, for $K=(|B_1|,|B_2|,\ldots,|B_m|)$, we have $\widetilde{k}=\mathrm{lcm}(|B_1|,|B_2|,\ldots,|B_m|)=\mathrm{lcm}(w_1,w_2,\ldots,w_l)$. Thus, we have

$$\begin{aligned}\bigcup_{1\leq t\leq l}\bigcup_{\{i\,:\,|B_i|=w_t\}}\bigcup_{\{j:1\leq j\leq m,\, j\neq i\}} D(B_j,\widetilde{B}_i)=&\left(\sum_{1\leq t\leq l}\lambda_t\frac{\widetilde{k}}{w_t}\right)\boxtimes(G\backslash\{0\})\\ =&\lambda\boxtimes(G\backslash\{0\}),\end{aligned}$$

i.e., $\{B_i\,:\,1\leq i\leq m\}$ is an $(n,m,K=(|B_1|,|B_2|,\ldots,|B_m|),a,\lambda)$-SWEDF, where $\lambda=\sum_{1\leq t\leq l}\frac{\lambda_t\widetilde{k}}{w_t}$. □

In what follows, we recall an example of SWEDF which is not an EDF, or an GSEDF, or a PEDF.

Example 3.5.4 ([81])**.** *Let $G=(\mathbb{Z}_{10},+)$ and $\mathcal{B}=\{B_1=\{0\},B_2=\{5\},B_3=\{2,3\},B_4=\{6,4\}\}$. Then $\widetilde{B}_1=\{0,0\},\widetilde{B}_2=\{5,5\},\widetilde{B}_3=\{2,3\},\widetilde{B}_4=\{6,4\}$. It is easy to check*

$$\bigcup_{1\leq i\leq 4}\bigcup_{\substack{1\leq j\leq 4,\\ j\neq i}} D(B_i,\widetilde{B}_j)=4\boxtimes(G\backslash\{0\}),$$

$$\bigcup_{1\leq i\leq 4}\bigcup_{\substack{1\leq j\leq 4,\\ j\neq i}} D(B_i,B_j)\neq\lambda\boxtimes(G\backslash\{0\}),$$

$$\bigcup_{2\le j\le 4} D(B_1,B_j)=\{5,8,7,4,6\}\neq \lambda\boxtimes(G\backslash\{0\}),$$

and

$$\bigcup_{3\le i\le 4}\bigcup_{\substack{1\le j\le 4,\\ j\neq i}} D(B_i,B_j)\neq \lambda\boxtimes(G\backslash\{0\}),$$

for any positive integer λ. Thus, $\mathcal{B}$ is an SWEDF which does not form an EDF, or a GSEDF, or a PEDF.

Similarly, a BEDF is not necessarily a BSWEDF in general and we have the following relationship between regular BEDFs and BSWEDFs.

Lemma 3.5.5. *The regular (n,k,λ)-BEDF forms an $(n,m,K=(k,k,\dots,k),a=mk,\lambda_1)$-BSWEDF, where $\lambda_1\le\lambda$.*

Lemma 3.5.6. *If $\mathcal{B}=\{B_i\,:\,1\le i\le m\}$ is an $(n,m;k_1,k_2,\dots,k_m;\lambda_1,\lambda_2,\dots,\lambda_m)$-BGSEDF, then $\mathcal{B}$ is an $(n,m,(k_1,k_2,\dots,\ k_m),a=\sum_{1\le i\le m}k_i,\lambda)$-BSWEDF, where $\lambda\le\sum_{1\le i\le m}\frac{\lambda_i\widetilde{k}}{k_i}$.*

Proof. Since $\mathcal{B}=\{B_i\,:\,1\le i\le m\}$ is an $(n,m;k_1,k_2,\dots,k_m;\lambda_1,\lambda_2,\dots,\lambda_m)$-BGSEDF, by (3.2),

$$\bigcup_{\substack{1\le j\le m,\\ j\neq i}} D(B_i,B_j)\subseteq \lambda_i\boxtimes(G\backslash\{0\}),$$

which means

$$\bigcup_{\substack{1\le j\le m,\\ j\neq i}} D(B_j,\widetilde{B}_i)\subseteq \lambda_i\frac{\widetilde{k}}{k_i}\boxtimes(G\backslash\{0\}). \tag{3.26}$$

Let λ be the smallest positive integer such that

$$\bigcup_{1\le i\le m}\bigcup_{\substack{1\le j\le m,\\ j\neq i}} D(B_j,\widetilde{B}_i)\subseteq \lambda\boxtimes(G\backslash\{0\}).$$

Thus, by (3.26), we have $\lambda\le\sum_{1\le i\le m}\frac{\lambda_i\widetilde{k}}{k_i}$, i.e., $\mathcal{B}$ is an $(n,m,(k_1,k_2,\dots,k_m),a=\sum_{1\le i\le m}k_i,\lambda)$-BSWEDF. □

In Fig. 3.2, there is the relationship among various external difference families.

3.5.2 Explicit constructions of optimal BSWEDFs

In what follows, we consider explicit constructions of optimal BSWEDFs. Recall the following well-known construction of difference families. Let $q=4k+1$ be a prime power. Let α be a primitive element of $\mathbb{F}_q$,

$$D_i^2=\{\alpha^{i+2j}\,:\,0\le j\le 2k-1\},\ \text{for } i=0,1 \tag{3.27}$$

$$\begin{tikzcd}
& \text{Regular } BEDF \arrow[d, "\text{Lemma 3.5.5}"] & \\
& BSWEDF & BGSEDF \arrow[l, "\text{Lemma 3.5.6}"'] \\
& SWEDF \arrow[u, "\text{Def. 3.4.1}"'] & \\
\text{Regular } EDF \arrow[ur, "\text{Lemma 3.5.1}"'] & PEDF \arrow[u, "\text{Lemma 3.5.3}"'] & GSEDF \arrow[uu, "\text{Def. 3.1.17}"'] \arrow[ul, "\text{Lemma 3.5.2}"]
\end{tikzcd}$$

Figure 3.2: Summary of the relationship among regular EDF, $PEDF$, regular $BEDF$, $SWEDF$, $GSEDF$, $BGSEDF$, and, $BSWEDF$.

and

$$D_i^4 = \{\alpha^{i+4j} \,:\, 0 \le j \le k-1\}, \text{ for } 0 \le i \le 3. \tag{3.28}$$

It is well-known that $\{D_0^2, D_1^2\}$ is a $(q, 2k, 2k-1)$-DF over the additive group of $\mathbb{F}_q$.

Construction 3.5.7. *Let $\mathcal{S} = \{S_1, S_2, S_3\}$ be the family of disjoint subsets of $\mathbb{Z}_2 \times \mathbb{F}_q$ defined as*

$$S_1 = \{(0,0), (1,0)\}, \ S_2 = \{0\} \times D_0^4 \cup \{1\} \times D_2^4,$$

and

$$S_3 = \{0\} \times D_1^4 \cup \{0\} \times D_3^4.$$

Theorem 3.5.8. *Let $\mathcal{S} = \{S_1, S_2, S_3\}$ be the family defined in Construction 3.5.7. If k is odd, then $\mathcal{S}$ is an optimal $(n = 2q, m = 3, (2, 2k, 2k), a = 4k+2, \lambda = 2k+1)$-BSWEDF.*

Before the proof we list a well-known result about D_0^2 and D_1^2.

Lemma 3.5.9. *If k is odd, then the family $\{D_0^2, D_1^2\}$ satisfies*

$$D\left(D_0^2, D_1^2\right) \cup D\left(D_1^2, D_0^2\right) = 2k \boxtimes (\mathbb{F}_q \backslash \{0\})$$

and

$$\begin{aligned} &D\left(D_0^4, D_1^4\right) \cup D\left(D_0^4, D_3^4\right) \cup D\left(D_1^4, D_0^4\right) \cup D\left(D_3^4, D_0^4\right) \\ =&k \boxtimes (\mathbb{F}_q \backslash \{0\}). \end{aligned}$$

Proof. By (3.27) and (3.28), we have

$$D_0^2 = D_0^4 \cup D_2^4 = D_0^4 \cup (-D_0^4)$$

and

$$D_1^2 = D_1^4 \cup D_3^4 = D_1^4 \cup (-D_1^4),$$

where $\alpha^{2k} = -1$. The fact $\{D_0^2, D_1^2\}$ is a $(q, 2k, 2k-1)$-PDF means that

$$D\left(D_0^2, D_1^2\right) \cup D\left(D_1^2, D_0^2\right) = 2k \boxtimes (\mathbb{F}_q \backslash \{0\}).$$

The preceding equality can be rewritten as

$$\begin{aligned} &2k \boxtimes (\mathbb{F}_q \backslash \{0\}) \\ =&D\left(D_0^2, D_1^2\right) \cup D\left(D_1^2, D_0^2\right) \\ =&D\left(D_0^4 \cup (-D_0^4), D_1^4 \cup D_3^4\right) \cup D\left(D_1^4 \cup D_3^4, D_0^4 \cup (-D_0^4)\right) \\ =&2 \boxtimes \left(D\left(D_0^4, D_1^4\right) \cup D\left(D_0^4, D_3^4\right) \cup D\left(D_1^4, D_0^4\right) \cup D\left(D_3^4, D_0^4\right)\right), \end{aligned}$$

where for the last equality we use the facts

$$\begin{aligned} D(-D_0^4, D_1^4 \cup D_3^4) =& D(-(D_1^4 \cup D_3^4), D_0^4) \\ =& D(D_3^4 \cup D_1^4, D_0^4) \end{aligned}$$

and

$$\begin{aligned} D\left(D_1^4 \cup D_3^4, -D_0^4\right) =& D\left(D_0^4, -(D_1^4 \cup D_3^4)\right) \\ =& D\left(D_0^4, D_3^4 \cup D_1^4\right). \end{aligned}$$

This completes the proof. □

Proof of Theorem 3.5.8: By Definition 3.4.1, in this case, $\widetilde{k} = \mathrm{lcm}(2k, 2) = 2k$, $\widetilde{S}_1 = k \boxtimes \{(0,0),(1,0)\}$, $\widetilde{S}_2 = S_2$, and $\widetilde{S}_3 = S_3$. Thus, $D(S_2, \widetilde{S}_3) = D(S_2, S_3)$ and $D(S_3, \widetilde{S}_2) = D(S_3, S_2)$. Recall that $S_2 = \{0\} \times D_0^4 \cup \{1\} \times (-D_0^4)$, which implies

$$\begin{aligned} & D(S_2, \widetilde{S}_3) \cup D(S_3, \widetilde{S}_2) \\ =& D(\{0\} \times D_0^4 \cup \{1\} \times (-D_0^4), \{0\} \times D_1^4 \cup \{0\} \times D_3^4) \\ & \cup D(\{0\} \times D_1^4 \cup \{0\} \times D_3^4, \{0\} \times D_0^4 \cup \{1\} \times (-D_0^4)) \\ =& \bigcup_{i=0,1} \{i\} \times \left(D\left(D_0^4, D_1^4\right) \cup D\left(D_0^4, D_3^4\right) \cup D\left(D_1^4, D_0^4\right) \cup D\left(D_3^4, D_0^4\right)\right) \\ =& k \boxtimes \left(\mathbb{Z}_2 \times (\mathbb{F}_q \backslash \{0\})\right), \end{aligned} \tag{3.29}$$

where we use the fact $D_1^4 = -D_3^4$ and the last equality holds by Lemma 3.5.9. By the fact $\bigcup_{0 \le i \le 3} D_i^4 = \mathbb{F}_q \backslash \{0\}$, we have

$$\begin{aligned} & D(S_1, \widetilde{S}_2) \cup D(S_2, \widetilde{S}_1) \\ =& \{0\} \times D_2^4 \cup \{1\} \times D_0^4 \cup \{1\} \times D_2^4 \cup \{0\} \times D_0^4 \\ & \cup k \boxtimes \left(\{0\} \times D_2^4 \cup \{1\} \times D_0^4 \cup \{1\} \times D_2^4 \cup \{0\} \times D_0^4\right) \\ =& (k+1) \boxtimes \left(\mathbb{Z}_2 \times D_0^2\right), \end{aligned}$$

and

$$\begin{aligned} & D(S_1, \widetilde{S}_3) \cup D(S_3, \widetilde{S}_1) \\ =& \{0\} \times D_1^2 \cup \{1\} \times D_1^2 \cup k \boxtimes \left(\{0\} \times D_1^2 \cup \{1\} \times D_1^2\right) \\ =& (k+1) \boxtimes \left(\mathbb{Z}_2 \times D_1^2\right), \end{aligned}$$

where we use the facts $D_i^2 = D_i^4 \cup D_{i+2}^4$ and $D_i^4 = -D_{i+2}^4$ for $i = 0, 1$. The above two equalities imply that

$$\bigcup_{i=2,3} \left(D(S_1, \widetilde{S}_i) \cup D(S_i, \widetilde{S}_1)\right) = (k+1) \boxtimes \left(\mathbb{Z}_2 \times (\mathbb{F}_q \backslash \{0\})\right). \tag{3.30}$$

Therefore, by (3.29) and (3.30),

$$\begin{aligned}\bigcup_{1\le i\ne j\le 3} D(S_i,\widetilde{S_j}) =&(2k+1)\boxtimes(\mathbb{Z}_2\times(\mathbb{F}_q\backslash\{0\}))\\ \subseteq&(2k+1)\boxtimes((\mathbb{Z}_2\times\mathbb{F}_q)\backslash\{(0,0)\}),\end{aligned}$$

i.e., $S=\{S_1,S_2,S_3\}$ is an $(n=2q,m=3,(2,2k,2k),a=4k+2,\lambda=2k+1)$-BSWEDF. By Lemma 3.4.2, we have

$$\begin{aligned}\lambda\ge&\left\lceil\frac{\widetilde{k}a(m-1)}{n-1}\right\rceil=\left\lceil\frac{2k(4k+2)2}{2q-1}\right\rceil=\left\lceil\frac{2k(8k+1)+6k}{8k+1}\right\rceil\\ =&2k+1.\end{aligned}$$

Thus, S is an optimal $(n=2q,m=3,(2,2k,2k),a=4k+2,\lambda=2k+1)$-BSWEDF. □

It is easily seen from the proof of Theorem 3.5.8 that the above BSWEDFs are not EDFs, or GSEDFs, or PEDFs.

Example 3.5.10. *Let $n=2q=26$. By Construction 3.5.7, the family of sets $\mathcal{S}=\{S_1,S_2,S_3\}$ over $\mathbb{Z}_{26}$ can be listed as*

$$S_1=\{0,13\},\ S_2=\{14,16,22,17,25,23\},\ \textit{and}$$

$$S_3=\{2,6,18,8,24,20\}.$$

It is easy to check that

$$\bigcup_{1\le i\ne j\le 3} D(S_i,\widetilde{S_j})=7\boxtimes(\mathbb{Z}_{26}\backslash\{0,13\}),$$

which means that $\mathcal{S}$ is an optimal $(26,3,(2,6,6),14,7)$-BSWEDF.

Let $n_1=2k+1$ and $\{\{0\},E_1,E_2\}$ be an $(n_1,k,k-1)$-PDF over an Abelian group G of order n_1. Such kinds of PDFs exist, for example, when n_1 is a prime power, and $E_1=D_0^2$, $E_2=D_1^2$. Based on $\{\{0\},E_1,E_2\}$ we can construct a BSWEDF as follows.

Construction 3.5.11. *Let $\mathcal{W}=\{W_1,W_2,W_3\}$ be the family of disjoint subsets of $\mathbb{Z}_2\times G$, defined as $W_1=\{(1,0)\}$, $W_2=\{0\}\times E_1$, and $W_3=\{0\}\times E_2$.*

Theorem 3.5.12. *The family $\mathcal{W}=\{W_1,W_2,W_3\}$ generated by Construction 3.5.11 is an optimal $(n=2n_1,3,(1,k,k),2k+1,k+1)$-BSWEDF.*

Proof. The fact that $\{\{0\},E_1,E_2\}$ is an $(n_1=2k+1,k,k-1)$-PDF over G means that $D(E_1,E_2)\cup D(E_2,E_1)=k\boxtimes(G\backslash\{0\})$. Thus, we have

$$\begin{aligned}D(W_2,\widetilde{W_3})\cup D(W_3,\widetilde{W_2}) =&D(W_2,W_3)\cup D(W_3,W_2)\\ =&k\boxtimes(\{0\}\times(G\backslash\{0\})),\end{aligned}$$

where we apply the fact $\widetilde{k} = \mathrm{lcm}(1, k, k) = k = |W_2| = |W_3|$. Note that

$$\begin{aligned}
&D(W_1, \widetilde{W_2}) \cup D(W_1, \widetilde{W_3}) \cup D(W_3, \widetilde{W_1}) \cup D(W_2, \widetilde{W_1}) \\
=&\{1\} \times (-E_1) \cup \{1\} \times (-E_2) \cup D(\{0\} \times E_1, k \boxtimes \{(1,0)\}) \\
&\cup D(\{0\} \times E_2, k \boxtimes \{(1,0)\}) \\
=&(k+1) \boxtimes (\{1\} \times (G \backslash \{0\})).
\end{aligned}$$

Based on the above two equalities,

$$\bigcup_{1 \le i \ne j \le 3} D(W_i, \widetilde{W_j}) \subseteq (k+1) \boxtimes ((\mathbb{Z}_2 \times G) \backslash \{(0,0)\}),$$

i.e., $\mathcal{W}$ is an $(n = 2n_1, m = 3, (1, k, k), a = 2k+1, \lambda = k+1)$-BSWEDF.

By Lemma 3.4.2, we have

$$\begin{aligned}
\lambda \ge &\left\lceil \frac{\widetilde{k}a(m-1)}{n-1} \right\rceil = \left\lceil \frac{k(2k+1)2}{2n_1 - 1} \right\rceil = \left\lceil \frac{k(4k+1)+k}{4k+1} \right\rceil \\
=&k+1.
\end{aligned}$$

Thus, $\mathcal{W}$ is an optimal $(2n_1 = 4k+2, 3, (1, k, k), 2k+1, k+1)$-BSWEDF.

□

It is easily seen from the proof of Theorem 3.5.12 that the above BSWEDFs are not EDFs, or GSEDFs, or PEDFs.

Example 3.5.13. *Let $n = 2n_1 = 22$. By Construction 3.5.11, the family of sets $\mathcal{W} = \{W_1, W_2, W_3\}$ over $\mathbb{Z}_{22}$ can be listed as*

$$W_1 = \{11\},\ W_2 = \{12, 4, 16, 20, 14\},\ \textit{and}$$

$$W_3 = \{2, 8, 10, 18, 6\}.$$

It is easy to check that

$$\bigcup_{1 \le i \ne j \le 3} D(W_i, \widetilde{W_j}) \subseteq 6 \boxtimes (Z_{22} \backslash \{0\}),$$

which means that $\mathcal{W}$ is an optimal $(22, 3, (1, 5, 5), 11, 6)$-BSWEDF.

Construction 3.5.14. *Let $q = 4k+1$ be a prime power and let $\mathcal{U} = \{U_1, U_2, U_3, U_4\}$ be the family of disjoint subsets of $\mathbb{Z}_3 \times \mathbb{F}_q$, defined as $U_1 = \{(1,0)\}$, $U_2 = \{(2,0)\}$, $U_3 = \{0\} \times D_0^2$, and $U_4 = \{0\} \times D_1^2$.*

Theorem 3.5.15. *The family $\mathcal{U} = \{U_1, U_2, U_3, U_4\}$ in Construction 3.5.14 is an optimal $(3q = 12k+3, 4, (1, 1, 2k, 2k), 4k+2, 2k+1)$-BSWEDF.*

Proof. Note that $\widetilde{k} = \mathrm{lcm}(1, 1, 2k, 2k) = 2k$, which implies $\widetilde{U}_3 = U_3$ and $\widetilde{U}_4 = U_4$. Lemma 3.5.9 shows that $D(D_0^2, D_1^2) \cup D(D_1^2, D_0^2) = 2k \boxtimes (\mathbb{F}_q \backslash \{0\})$. Thus, we have

$$\begin{aligned} D(U_3, \widetilde{U}_4) \cup D(U_4, \widetilde{U}_3) =& D(U_3, U_4) \cup D(U_3, U_4) \\ =& 2k \boxtimes (\{0\} \times (\mathbb{F}_q \backslash \{0\})). \end{aligned}$$

Recall that

$$\begin{aligned} & D(U_1, \widetilde{U}_3) \cup D(U_1, \widetilde{U}_4) \cup D(U_3, \widetilde{U}_1) \cup D(U_4, \widetilde{U}_1) \\ =& (\{1\} \times D_0^2) \cup (\{1\} \times D_1^2) \cup D(\{0\} \times D_0^2, 2k \boxtimes \{(1, 0)\}) \\ & \cup D(\{0\} \times D_1^2, 2k \boxtimes \{(1, 0)\}) \\ =& (\{1\} \times (\mathbb{F}_q \backslash \{0\})) \cup 2k \boxtimes (\{2\} \times (\mathbb{F}_q \backslash \{0\})) \end{aligned}$$

and

$$\begin{aligned} & D(U_2, \widetilde{U}_3) \cup D(U_2, \widetilde{U}_4) \cup D(U_3, \widetilde{U}_2) \cup D(U_4, \widetilde{U}_2) \\ =& \{2\} \times D_0^2 \cup \{2\} \times D_1^2 \cup D(\{0\} \times D_0^2, 2k \boxtimes \{(2, 0)\}) \\ & \cup D(\{0\} \times D_1^2, 2k \boxtimes \{(2, 0)\}) \\ =& (\{2\} \times (\mathbb{F}_q \backslash \{0\})) \cup 2k \boxtimes (\{1\} \times (\mathbb{F}_q \backslash \{0\})). \end{aligned}$$

For the differences between U_1 and U_2, we have

$$D(U_1, \widetilde{U}_2) \cup D(U_2, \widetilde{U}_1) = 2k \boxtimes \{(1, 0), (2, 0)\}.$$

Therefore, the above four equalities mean that

$$\begin{aligned} & \bigcup_{1 \le i \ne j \le 4} D(U_i, \widetilde{U}_j) \\ =& (2k \boxtimes \{(1, 0), (2, 0)\}) \cup (2k \boxtimes \{0\} \times (\mathbb{F}_q \backslash \{0\})) \\ & \cup ((2k + 1) \boxtimes \{1, 2\} \times (\mathbb{F}_q \backslash \{0\})) \\ \subseteq& (2k + 1) \boxtimes ((\mathbb{Z}_3 \times \mathbb{F}_q) \backslash \{(0, 0)\}), \end{aligned}$$

i.e., $\mathcal{U}$ is an $(n = 3q, m = 4, (1, 1, 2k, 2k), a = 4k + 2, \lambda = 2k + 1)$-BSWEDF.

By Lemma 3.4.2, we have

$$\begin{aligned} \lambda \ge & \left\lceil \frac{\widetilde{k}a(m - 1)}{n - 1} \right\rceil = \left\lceil \frac{2k(4k + 2)3}{3q - 1} \right\rceil = \left\lceil \frac{2k(12k + 2) + 8k}{12k + 2} \right\rceil \\ =& 2k + 1. \end{aligned}$$

Thus, $\mathcal{U}$ is an optimal $(3q, 4, (1, 1, 2k, 2k), 4k + 2, 2k + 1)$-BSWEDF.

□

It is easily seen from the proof of Theorem 3.5.15 that the above BSWEDFs are not EDFs, or GSEDFs, or PEDFs.

Table 3.4: Some known PDFs with parameters $(n, \mathcal{W} = (k^{\frac{n-k+1}{k}}, (k-1)^1), k-1)$

Parameters	Constraints	Ref.
$\Big(2v,\ (3^{\frac{2v-2}{3}}, 2^1),\ 2\Big)$,	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, and $3 \mid (p_t - 1)$ for $1 \le t \le r$	[9]
$\Big(sv,\ ((s+1)^{\frac{sv-s}{s+1}}, s^1),\ s\Big)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $2(s+1) \mid (p_t - 1)$ for $1 \le t \le r, s = 4, 5$	[9]
$\Big(6v,\ (7^{\frac{6v-6}{7}}, 6^1),\ 6\Big)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $28 \mid (p_t - 1)$ for $1 \le t \le r$	[9]
$\Big(7v,\ (8^{\frac{7v-7}{8}}, 7^1),\ 7\Big)$	$v = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$, $2 < p_1 < p_2 < \cdots < p_r$, $8 \mid (p_t - 1)$ for $1 \le t \le r, v \notin \{17, 89\}$	[9]
$(q-1, (\frac{q}{d}^{d-1}, (\frac{q}{d}-1)^1), \frac{q-d}{d})$	$d \mid q$, $\gcd(\frac{q}{d} - 1, (q-1)/(\frac{q}{d} - 1)) = 1$	[36]

Herein p_i's are primes. t, s, r and m are positive integers. q is a prime power.

Example 3.5.16. *Let $n = 3q = 39$. By Construction 3.5.14, the family of sets $\mathcal{U} = \{U_1, U_2, U_3, U_4\}$ over $\mathbb{Z}_{39}$ can be listed as*

$$U_1 = \{13\},\ U_2 = \{26\},\ U_3 = \{27, 30, 3, 12, 9, 36\},\ \textit{and}$$

$$U_4 = \{15, 21, 6, 24, 18, 33\}.$$

It is easy to check that

$$\bigcup_{1 \le i \ne j \le 4} D(U_i, \widetilde{U_j}) \subseteq 7 \boxtimes (\mathbb{Z}_{39} \backslash \{0\}),$$

which means that $\mathcal{U}$ is an optimal $(39, 4, (1, 1, 6, 6), 14, 7)$-BSWEDF.

3.5.3 A construction of cyclic SWEDFs

In this subsection, we are going to construct cyclic SWEDFs, which are not regular EDFs, or GSEDFs, or PEDFs. A *cyclic* SWEDF means an SWEDF over a cyclic additive group.

A well-studied kind of PDFs $\mathcal{R} = \{R_1, R_2, \dots, R_l\}$ are those with parameters $(n = (k-1)(tk+1), (k, \dots, k, k-1), k-1)$ over $\mathbb{Z}_n = \mathbb{Z}_{k-1} \times \mathbb{Z}_{tk+1}$ where $\gcd(k-1, tk+1) = 1$, $R_l = \mathbb{Z}_{k-1} \times \{0\}$, $|R_l| = k-1$ and $l = t(k-1)+1$. In Table 3.4, we list such PDFs which can be applied in the following construction.

Construction 3.5.17. *Let $\mathcal{V} = \{V_1, V_2, \ldots, V_{t(k-1)+k-2}\}$ be the family of disjoint subsets of $\mathbb{Z}_n$, defined as*

$$V_i = R_i \quad \textit{for } 1 \leq i \leq t(k-1)$$

and

$$V_{t(k-1)+j} = \{(j,0)\} \textit{ for } 1 \leq j \leq k-2.$$

Theorem 3.5.18. *Let $\mathcal{V}$ be the family in Construction 3.5.17. Then $\mathcal{V}$ is a cyclic $(n, t(k-1)+k-2, K = (k, \ldots, k, 1, 1, \ldots, 1), n-1, (t+1)k^2 - (t+3)k)$-SWEDF, where the element 1 appears $k-2$ times and the element k appears $t(k-1)$ times in K.*

Proof. Since $\mathcal{R}$ is an $(n = (k-1)(tk+1), (k, \ldots, k, k-1), k-1)$-PDF, we can conclude that

$$\bigcup_{1 \leq i \neq j \leq l} D(R_i, R_j) = (n-k+1) \boxtimes ((\mathbb{Z}_{k-1} \times \mathbb{Z}_{tk+1}) \backslash \{(0,0)\}).$$

Recall that $R_l = \mathbb{Z}_{k-1} \times \{0\}$, which means

$$\bigcup_{1 \leq i \leq l-1} (D(R_i, R_l) \cup D(R_l, R_i)) = (2k-2) \boxtimes (\mathbb{Z}_{k-1} \times (\mathbb{Z}_{tk+1} \backslash \{0\})).$$

Thus, by Construction 3.5.17, we have

$$\begin{aligned}
&\bigcup_{1 \leq i \neq j \leq l-1} D(V_i, \widetilde{V_j}) \\
=& \bigcup_{1 \leq i \neq j \leq l-1} D(V_i, V_j) \\
=& \bigcup_{1 \leq i \neq j \leq l-1} D(R_i, R_j) \\
=& \left(\bigcup_{1 \leq i \neq j \leq l} D(R_i, R_j) \right) \backslash \left(\bigcup_{1 \leq i \leq l-1} (D(R_i, R_l) \cup D(R_l, R_i)) \right) \\
=& \left((n-k+1) \boxtimes ((\mathbb{Z}_{k-1} \backslash \{0\}) \times \{0\})\right) \\
&\cup \left((n-3k+3) \boxtimes (\mathbb{Z}_{k-1} \times (\mathbb{Z}_{tk+1} \backslash \{0\}))\right),
\end{aligned} \tag{3.31}$$

where we use the fact $\widetilde{k} = k$.

Note that for any $1 \leq j \leq k-2$,

$$\begin{aligned}
&\bigcup_{1 \leq i \leq l-1} (D(V_i, \widetilde{V}_{l-1+j}) \cup D(V_{l-1+j}, \widetilde{V_i})) \\
=& \bigcup_{1 \leq i \leq l-1} (D(R_i, k \boxtimes \{(j,0)\}) \cup D(\{(j,0)\}, R_i)) \\
=&(k+1) \boxtimes (\mathbb{Z}_{k-1} \times (\mathbb{Z}_{tk+1} \backslash \{0\})).
\end{aligned}$$

Thus, we have

$$\begin{aligned}&\bigcup_{1\leq j\leq k-2}\bigcup_{1\leq i\leq l-1}(D(V_i,\widetilde{V}_{l-1+j})\cup D(V_{l-1+j},\widetilde{V}_i))\\ =&((k+1)(k-2))\boxtimes(\mathbb{Z}_{k-1}\times(\mathbb{Z}_{tk+1}\backslash\{0\})).\end{aligned}\tag{3.32}$$

For the last part of external differences, we have

$$\begin{aligned}\bigcup_{1\leq i\neq j\leq k-2}D(V_{l-1+i},\widetilde{V}_{l-1+j})=&\bigcup_{1\leq i\neq j\leq k-2}D(\{(i,0)\},k\boxtimes\{(j,0)\})\\ =&k\boxtimes\left(\bigcup_{1\leq i\neq j\leq k-2}D(\{(i,0)\},\{(j,0)\})\right)\\ =&k(k-3)\boxtimes((\mathbb{Z}_{k-1}\backslash\{0\})\times\{0\}).\end{aligned}\tag{3.33}$$

Combining (3.31), (3.32), and (3.33),

$$\begin{aligned}&\bigcup_{1\leq i\neq j\leq l+k-3}D(V_i,\widetilde{V}_j)\\ =&\left(\bigcup_{1\leq i\neq j\leq l-1}D(V_i,\widetilde{V}_j)\right)\cup\left(\bigcup_{1\leq i\neq j\leq k-2}D(V_{l-1+i},\widetilde{V}_{l-1+j})\right)\\ &\cup\left(\bigcup_{1\leq j\leq k-2}\bigcup_{1\leq i\leq l-1}(D(V_i,\widetilde{V}_{l-1+j})\cup D(V_{l-1+j},\widetilde{V}_i))\right)\\ =&((n-k+1+k(k-3))\boxtimes(\mathbb{Z}_{k-1}\backslash\{0\})\times\{0\})\\ &\cup((n-3k+3+(k+1)(k-2))\boxtimes(\mathbb{Z}_{k-1}\times(\mathbb{Z}_{tk+1}\backslash\{0\})))\\ =&((t+1)k^2-tk-3k)\boxtimes((\mathbb{Z}_{k-1}\times\mathbb{Z}_{tk+1})\backslash\{(0,0)\}),\end{aligned}$$

where $n=(k-1)(tk+1)$.

Therefore, $\mathcal{V}$ is a cyclic $(n,t(k-1)+k-2,(k,k,\ldots,k,1,1,\ldots,1),n-1,(t+1)k^2-(t+3)k)$-SWEDF, where the element 1 occurs $k-2$ times in K and the element k appears $t(k-1)$ times in K. This completes the proof. □

In [55], Huczynska and Paterson introduced some constructions of SWEDFs with the so-called bimodal property.

Definition 3.5.19 ([55]). *Let G be a finite Abelian group and $\mathcal{B}$ be a collection $B_1,B_2,\ldots,B_m$ of disjoint subsets of G with sizes $k_1,k_2,\ldots,k_m$, respectively. We say that $\mathcal{B}$ has the bimodal property if for each $\delta\in G\backslash\{0\}$ we have $N_i(\delta)\in\{0,k_i\}$ for $1\leq i\leq m$, where*

$$N_i(\delta)\triangleq\left|\left\{(b_i,b_j):b_i\in B_i,\ b_j\in\bigcup_{1\leq t\neq i\leq m}B_t,\ \textit{and }b_j-b_i=\delta\right\}\right|.$$

Remark 3.5.20. *The SWEDF generated by Construction 3.5.17 does not have the bimodal property. Let $\mathcal{V}$ be the SWEDF generated by Construction 3.5.17. For any $v \in V_i$ with $|V_i| = k$, we have $0 \in D(V_i, \{v\})$ and $|D(V_i, \{v\})| = |V_i| = k$. However, by Construction 3.5.17, 0 is not an element of V_j for $1 \leq j \leq l+k-3$. Thus, the number of solutions for $a - b = v$ for $a \in V_i$ and $b \in V_j$ for $1 \leq j \leq l+k-3$ and $j \neq i$ is at most $k-1$, since $\bigcup_{1\leq j \leq l+k-3} V_j = \mathbb{Z}_n \backslash \{0\}$, i.e., $N_i(v) \leq k-1$. Next, we show that there exists V_i with $|V_i| = k$ satisfying $N_i(v) \neq 0$. If $a - b \neq v$ for all $a \in V_i$ and $b \in V_j$ for $1 \leq j \leq l+k-3$ and $j \neq i$, then $a \in V_i$ means that $(a + \langle v \rangle)\backslash\{0\} \in V_i$. This is to say that V_i is the union of some cosets of $\langle v \rangle$ besides the element 0 and $k = \tau|\langle v \rangle| - 1$ for some integer $\tau \geq 1$. This is impossible since there are elements v with $|\langle v \rangle| > k+1$ in $\mathbb{Z}_n\backslash\{0\}$. Thus, the SWEDF generated by Construction 3.5.17 is not bimodal. For more details about SWEDFs with bimodal property the reader may refer to [55, 56].*

Remark 3.5.21. *The only known SWEDFs without the bimodal property have parameters $(k_1k_2 + 1, m = 2, (k_1, k_2), a = k_1 + k_2, \lambda = 2)$, where $k_1 > 2$ and $k_2 > 2$ are integers. Thus, compared with the constructions in [55], Theorem 3.5.18 can generate new SWEDFs without bimodal property. Our construction also present the first class of SWEDFs without bimodal with $m > 2$.*

Example 3.5.22. *Let $G = (\mathbb{Z}_{15}, +)$ and*

$$\mathcal{R} = \{R_1 = \{6, 9, 2, 8\}, R_2 = \{11, 14, 7, 13\}, R_3 = \{1, 4, 12, 3\}, R_4 = \{0, 5, 10\}\}.$$

It is easy to check that $\mathcal{R}$ is a PDF with parameters $(15, (4, 4, 4, 3), 3)$. By Construction 3.5.17, we generate a family of subsets of $\mathbb{Z}_{15}$ as

$$\begin{aligned}\mathcal{V} = \{&V_1 = \{6, 9, 2, 8\}, V_2 = \{11, 14, 7, 13\}, V_3 = \{1, 4, 12, 3\},\\ &V_4 = \{5\}, V_5 = \{10\}\}.\end{aligned}$$

It is easy to check that

$$\bigcup_{1\leq i \neq j \leq 5} D(V_i, \widetilde{V_j}) = 16 \boxtimes (\mathbb{Z}_{15}\backslash\{0\}),$$

i.e., $\mathcal{V}$ is a $(15, 5, (4, 4, 4, 1, 1), 14, 16)$-SWEDF. Note that $N_3(6) = 3 \notin \{0, 4\}$, which means the SWEDF does not have the bimodal property by Definition 3.5.19.

CHAPTER 4

Weak Algebraic Manipulation Detection Codes

In this chapter, we study weak algebraic manipulation detection (AMD) codes, i.e., under the assumption that the adversary has no knowledge about the source. We investigate the relationships between weak AMD codes and external difference families, and between systematic weak AMD codes and highly nonlinear functions. By means of these characterizations, we construct infinite families of weak AMD codes and systematic weak AMD codes, respectively.

4.1 Known weak AMD codes

In this section, we recall some known results about AMD codes under the weak model, i.e., weak AMD codes.

4.1.1 G-optimal weak AMD codes

We begin with the concept of G-optimal weak AMD codes.

Theorem 4.1.1 ([81]). *For any weak $(n, m, K = (a_1, a_2, \cdots, a_m), \rho)$-AMD code, the parameters satisfy that*

$$\rho \geq \frac{1}{a}, \tag{4.1}$$

where $a = \sum_{1 \leq i \leq m} a_i$.

Definition 4.1.2. *A weak (n, m, K, ρ)-AMD code is said to be G-optimal if it achieves the bound in (4.1) with equality, where "G" means that guessing the most likely encoding message is an optimal strategy for the adversary.*

Theorem 4.1.3 ([81]). *A G-optimal weak $(n, m, K = (k, k, \cdots, k), \rho = \frac{1}{a})$-AMD code is equivalent with an $(n, m, k, 1)$-BEDF, where $a = km$.*

Remark 4.1.4. *Generally speaking, for G-optimal weak AMD codes, we still do not have enough explicit constructions. This problem is still widely open.*

4.1.2 R-optimal weak AMD codes

In this subsection, we recall known results for R-optimal weak AMD codes. We begin with the formal definition of R-optimal weak AMD codes.

Theorem 4.1.5 ([81]). *For any weak $(n, m, K = (a_1, a_2, \cdots, a_m), \rho)$-AMD code, the probability ρ satisfies*

$$\rho \geq \frac{a(m-1)}{m(n-1)},$$

where $a = \sum_{1 \leq i \leq m} a_i$.

Definition 4.1.6 ([81]). *A weak AMD code that meets the bound of Theorem 4.1.5 with equality is said to be R-optimal with respect to the bound in Theorem 4.1.5, where "R" is used to indicate that random choosing Δ is an optimal strategy for adversaries.*

In [55], Huczynska and Paterson characterized R-optimal weak AMD codes $(S, G, \mathcal{A}, E_u)$ by reciprocally-weighted external difference families, which can be defined as follows.

Definition 4.1.7 ([55]). *Let $\mathcal{B} = \{B_i : 1 \leq i \leq m\}$ be a family of disjoint subsets of G. Let $K = (k_1, k_2, \cdots, k_m)$ with $k_i = |B_i|$ for $1 \leq i \leq m$. Then $\mathcal{B}$ is said to be an $(n, m, (k_1, k_2, \cdots, k_m), d)$ reciprocally-weighted external difference family (RWEDF) if*

$$d = \sum_{1 \leq i \leq m} \frac{N_i(\alpha)}{k_i} \text{ for each } \delta \in G \backslash \{0\},$$

where

$$N_i(\alpha) \triangleq \left| \left\{ (b_i, b_j) : b_i \in B_i,\ b_j \in \bigcup_{1 \leq t \neq i \leq m} B_t,\ \text{and } b_i - b_j = \alpha \right\} \right|.$$

Theorem 4.1.8 ([55]). *A weak $(n, m, K = (a_1, a_2, \cdots, a_m), \rho)$-AMD code $(S, G, \mathcal{A}, E_u)$ is R-optimal with respect to the bound in Theorem 4.1.5 if and only if there exists an (n, m, K, d)-RWEDF, where $a = \sum_{1 \leq i \leq m} a_i$, $\rho = \frac{a(m-1)}{m(n-1)}$, and $d = \frac{a(m-1)}{n-1}$.*

Remark 4.1.9. *RWEDFs with $m = 2$ were investigated in [55]. However, not much is known for RWEDFs with $m \geq 3$. Explicit constructions of RWEDFs with $m \geq 3$ are still open.*

4.1.3 A lower bound on systematic weak AMD codes

Chen et al. [23] established a lower bound on the smallest possible cheating probability ρ for systematic weak AMD codes in terms of nonlinearity of functions.

Lemma 4.1.10 ([23]). *For any systematic weak AMD code* $(G_1, G = G_1 \times G_2 \times B, \{(s, x, f(s, x)) : s \in G_1\}, E)$ *with parameters* $(|G|, |G_1|, (|G_2|, |G_2|, \cdots, |G_2|), \rho)$, *we have*

$$\rho \geq \frac{|G_1||G_2| - 1 - N_f(|G_1| - 1)|B|}{(|G_1| - 1)|G_2||B|},$$

where N_f *is the nonlinearity of* $f : G_1 \times G_2 \to B$.

4.2 Between weak AMD codes and weighted external difference families

In this section, we further consider the relationship between weak AMD codes and weighted external difference families.

For $\Delta \in G\backslash\{0\}$, let ρ_Δ denote the probability that the adversary wins by modifying $g \in A_s$ into $g + \Delta \in A_{s'}$ for some $s' \neq s$. Thus, we have $\rho = \max\{\rho_\Delta : \Delta \in G\backslash\{0\}\}$.

Theorem 4.2.1. *There exists a weak* (n, m, K, ρ)*-AMD code* $(S, G, \mathcal{A}, E_u)$ *if and only if there exists an* (n, m, K, a, λ)*-BSWEDF, where* $|G| = n$, $a = \sum_{1 \leq i \leq m} |A_{s_i}|$, $K = (|A_{s_1}|, |A_{s_2}|, \cdots, |A_{s_m}|)$, $s_i \in S$, *and* $\rho = \frac{\lambda}{\widetilde{k}m}$.

Proof. If $(S, G, \mathcal{A}, E_u)$ is a weak (n, m, K, ρ)-AMD code, then for any $\Delta \in G\backslash\{0\}$, we have

$$\rho_\Delta \leq \rho = \frac{\lambda}{\widetilde{k}m},$$

that is,

$$\begin{aligned}
\frac{\lambda}{\widetilde{k}m} \geq \rho_\Delta =& \sum_{s \in S} Pr(s) \sum_{g \in A_s} Pr(E_u(s) = g) \left(\sum_{s' \neq s, s' \in S} Pr(g + \Delta \in A_{s'}) \right) \\
=& \sum_{s \in S} \frac{1}{m} \sum_{g \in A_s} \frac{1}{|A_s|} \left(\sum_{s' \neq s, s' \in S} Pr(g + \Delta \in A_{s'}) \right) \\
=& \sum_{s \in S} \frac{1}{m} \frac{1}{|A_s|} \left(\sum_{s' \neq s, s' \in S} \sum_{g \in A_s} Pr(g + \Delta \in A_{s'}) \right),
\end{aligned} \tag{4.2}$$

where the second equality holds by the fact that E_u encodes s to elements of A_s with uniform probability. Note that for given Δ, s, $g \in A_s$ and $s' \neq s$,

$$Pr(g + \Delta \in A_{s'}) = \begin{cases} 1, & \Delta \in D(A_{s'}, \{g\}), \\ 0, & \Delta \notin D(A_{s'}, \{g\}). \end{cases}$$

Thus, inequality (4.2) implies that

$$
\begin{aligned}
\frac{\lambda}{m} \geq \widetilde{k}\rho_{\Delta} =& \sum_{s\in S}\frac{1}{m}\frac{\widetilde{k}}{|A_s|}\left(\sum_{s'\neq s, s'\in S}\sum_{g\in A_s} Pr(g+\Delta\in A_{s'})\right)\\
=& \sum_{s\in S}\frac{1}{m}\frac{\widetilde{k}}{|A_s|}\left(\sum_{s'\neq s, s'\in S}\#\left(\Delta, D(A_{s'}, A_s)\right)\right)\\
=& \sum_{s\in S}\frac{1}{m}\left(\sum_{s'\neq s, s'\in S}\frac{\widetilde{k}}{|A_s|}\#\left(\Delta, D(A_{s'}, A_s)\right)\right)\\
=& \sum_{s\in S}\frac{1}{m}\left(\sum_{s'\neq s, s'\in S}\#\left(\Delta, D(A_{s'}, \widetilde{A}_s)\right)\right)\\
=& \frac{1}{m}\#\left(\Delta, \bigcup_{\substack{s,s'\in S,\\ s'\neq s}} D(A_{s'}, \widetilde{A}_s)\right),
\end{aligned}
\tag{4.3}
$$

where $\#(\Delta, B)$ denotes the number of times that Δ appears in the multi-set B. This means that any $\Delta \in G\backslash\{0\}$ appears at most λ times in the multi-set $\bigcup_{\substack{s,s'\in S,\\ s'\neq s}} D(A_{s'}, \widetilde{A}_s)$, i.e.,

$$
\bigcup_{\substack{s,s'\in S,\\ s'\neq s}} D(A_{s'}, \widetilde{A}_s) \subseteq \lambda \boxtimes (G\backslash\{0\}).
$$

Note that $\rho = \max\{\rho_{\Delta} : \Delta \in G\backslash\{0\}\}$ means there exists at least one $\Delta \in G\backslash\{0\}$ such that the equality in inequality (4.3) holds. Then $\{A_s : s \in S\}$ forms an $(n, m, (|A_{s_1}|, |A_{s_2}|, \cdots, |A_{s_m}|), a, \lambda)$-BSWEDF by Definition 3.4.1.

Conversely, suppose that there exists an (n, m, K, a, λ)-BSWEDF $\mathcal{B} = \{B_i : 1 \leq i \leq m\}$ over G. Let $S = \{s_i : 1 \leq i \leq m\}$ and $A_{s_i} = B_i$ for $1 \leq i \leq m$. Then we can define a weak AMD code, where $E_u(s_i) = g \in B_i$ with equiprobability. For any

$\Delta \in G\backslash\{0\}$, similarly as (4.2), we have

$$\begin{aligned}
\rho_\Delta =& \sum_{s\in S}\frac{1}{m}\frac{1}{|A_s|}\left(\sum_{s'\neq s, s'\in S}\sum_{g\in A_s} Pr(g+\Delta\in A_{s'})\right)\\
=& \sum_{1\leq i\leq m}\frac{1}{m}\frac{1}{|B_i|}\left(\sum_{\substack{1\leq j\leq m\\ j\neq i}}\#(\Delta, D(B_j, B_i))\right)\\
=& \sum_{1\leq i\leq m}\frac{1}{m\widetilde{k}}\left(\sum_{\substack{1\leq j\leq m\\ j\neq i}}\#(\Delta, D(B_j, \widetilde{B}_i))\right)\\
=&\frac{1}{m\widetilde{k}}\left(\sum_{1\leq j\neq i\leq m}\#(\Delta, D(B_j, \widetilde{B}_i))\right)\\
\leq&\frac{\lambda}{m\widetilde{k}},
\end{aligned}$$

where the last inequality holds by the fact that $\mathcal{B}$ is an (n, m, K, a, λ)-BSWEDF. According to Definition 3.4.1, the equality is achieved for at least one $\Delta \in G\backslash\{0\}$ in the preceding inequality. Thus, the weak (n, m, K, ρ)-AMD code defined based on the BSWEDF $\mathcal{B}$ satisfies

$$\rho = \max\{\rho_\Delta : \Delta \in G\backslash\{0\}\} = \frac{\lambda}{\widetilde{k}m},$$

which completes the proof. □

When we consider the optimality of BSWEDF, the size-distribution $K = (k_1, k_2, \ldots, k_m)$ is given. However, the R-optimality of weak AMD only relates with $a = \sum_{1\leq i\leq m} k_i$ as defined in [81] but disregards the exact size-distribution K of $\mathcal{A}$. This is to say that the optimality of AMD codes try to find out minimum ρ for given parameters n, m, and a. According to Theorem 4.2.1, there may exist several BSWEDFs with different K which correspond to weak AMD codes with exactly the same parameter a. Thus, although the BSWEDF gives a characterization of the weak AMD code, in general, the optimal BSWEDF for a given K does not necessarily corresponds to an R-optimal weak AMD code for a given a.

Definition 4.2.2. *For given n, m and a, an (n, m, K, a, λ)-BSWEDF is said to be strongly optimal if $\frac{\lambda}{\widetilde{k}m} = \rho_{(n,m,a)}$, where*

$$\rho_{(n,m,a)} = \min_{K'}\left\{\frac{\lambda'}{\widetilde{k}'m} : \exists\,(n, m, K', a, \lambda')\text{-}BSWEDF\, s.t. \sum_{1\leq i\leq m} k_i' = a\right\}. \tag{4.4}$$

By Theorem 4.2.1 and Lemma 3.4.2, we have

Corollary 4.2.3. *For any weak (n, m, a, ρ)-AMD code $(S, G, \mathcal{A}, E_u)$, we have*

$$\rho \geq \rho_{(n,m,a)} \geq \min_{K} \left\{ \left\lceil \frac{\widetilde{k}a(m-1)}{n-1} \right\rceil \frac{1}{m\widetilde{k}} : \sum_{1 \leq i \leq m} k_i = a \right\},$$

where $|A_i| = k_i$ for any $A_i \in \mathcal{A}$.

Proof. Let $(S, G, \mathcal{A}, E_u)$ be a weak (n, m, a, ρ)-AMD code. By Theorem 4.2.1, there exists an (n, m, K, a, λ)-BSWEDF with $\lambda = \rho m\widetilde{k}$. Then by Lemma 3.4.2 and (4.4),

$$\rho = \frac{\lambda}{m\widetilde{k}} \geq \rho_{(n,m,a)} \geq \min_{K} \left\{ \left\lceil \frac{\widetilde{k}a(m-1)}{n-1} \right\rceil \frac{1}{m\widetilde{k}} : \sum_{1 \leq i \leq m} k_i = a \right\}.$$

□

Especially, for the t-regular AMD codes we have

Corollary 4.2.4. *For any t-regular weak (n, m, tm, ρ)-AMD code, we have*

$$\rho \geq \left\lceil \frac{t^2 m(m-1)}{n-1} \right\rceil \frac{1}{tm}. \tag{4.5}$$

Definition 4.2.5. *A weak AMD code (resp. t-regular AMD code) with $\rho = \rho_{(n,m,a)}$ is said to be optimal with respect to the bound in Corollary 4.2.3 (resp. Corollary 4.2.4).*

When $(n-1) \mid (\widetilde{k}a(m-1))$, the bound in Corollary 4.2.3 is exactly the same as the one given in Theorem 4.1.5. However, when $(n-1) \nmid (\widetilde{k}a(m-1))$, our bound in Corollary 4.2.3 can improve the known one in Theorem 4.1.5. The following is an easy example.

Corollary 4.2.6. *For any weak (n, m, a, ρ)-AMD code $(S, G, \mathcal{A}, E_u)$, if $n-1$ is a prime and $a < n-1$, then we have*

$$\rho \geq \min_{K} \left\{ \left\lceil \frac{\widetilde{k}a(m-1)}{n-1} \right\rceil \frac{1}{m\widetilde{k}} : \sum_{1 \leq i \leq m} k_i = a \right\} > \frac{a(m-1)}{m(n-1)}.$$

Proof. The corollary follows from the facts that $k_i \leq a < n-1$ for $1 \leq i \leq m$, $m \leq a < n-1$, and $n-1$ is a prime. In this case, $(n-1) \nmid (\widetilde{k}a(m-1))$. □

A more concrete example is listed below.

Example 4.2.7. *Let $n = 10$, $m = 3$, and $a = 5$. Let $B = \{\{5\}, \{2\}, \{0, 4, 6\}\}$ be a family of disjoint subsets of $\mathbb{Z}_{10}$, which corresponding to a weak $(10, 3, 5, \rho)$-AMD code, where $\rho = \frac{1}{3} \cdot \frac{1}{1} \cdot 1 + \frac{1}{3} \cdot \frac{1}{1} \cdot 0 + \frac{1}{3} \cdot \frac{1}{3} \cdot 1 = \frac{4}{9}$. According to Theorem*

4.1.5 and Definition 4.1.6, this is not an R-optimal weak AMD code. However, R-optimality should mean that random choosing Δ is an optimal strategy for the adversary. Clearly, according to Corollary 4.2.3, the parameter ρ cannot be smaller then

$$\begin{aligned}&\min_{K}\left\{\left\lceil\frac{\widetilde{k}5(3-1)}{10-1}\right\rceil\frac{1}{3\widetilde{k}}:\sum_{1\le i\le 3}k_i=5\right\}\\ =&\min\left\{\left\lceil\frac{\operatorname{lcm}(1,1,3)\cdot 5\cdot 2}{9}\right\rceil\frac{1}{3\operatorname{lcm}(1,1,3)},\left\lceil\frac{\operatorname{lcm}(1,2,2)\cdot 5\cdot 2}{9}\right\rceil\frac{1}{3\operatorname{lcm}(1,2,2)}\right\}\\ =&\min\left\{\frac{4}{9},\frac{1}{2}\right\}=\frac{4}{9}.\end{aligned}$$

Therefore, this example should be an optimal weak $(10,3,5,\rho)$-AMD code. This trouble is due to the fact that the known bound in Theorem 4.1.5 is not always tight.

Relationships between optimal weak AMD codes and optimal BSWEDFs are described below.

Theorem 4.2.8. *Let n and m be positive integers. There exists a G-optimal weak (n,m,a,ρ)-AMD code $(S,G,\mathcal{A},E_u)$ if and only if its corresponding BSWEDF with parameters $(n,m,K,a,\lambda=\frac{m\widetilde{k}}{a})$ is optimal, where $S=\{s_i:1\le i\le m\}$, $\mathcal{A}=\{A_{s_i}:1\le i\le m\}$, $k_i=|A_{s_i}|$ for $1\le i\le m$, $K=(k_1,k_2,\dots,k_m)$, and $a=\sum_{1\le i\le m}k_i$.*

Proof. If there exists a G-optimal weak (n,m,a,ρ)-AMD code $(S,G,\mathcal{A},E_u)$, i.e., $\rho=\frac{1}{a}$, then by Theorem 4.2.1, there exists a BSWEDF with parameters $(n,m,K,a,\lambda=\frac{m\widetilde{k}}{a})$. Now by Theorem 4.2.1 the optimality of weak $(n,m,a,\rho=\frac{1}{a})$-AMD code enhances that $\rho_{n,m,K}=\frac{1}{a}$, otherwise there will exist a weak $(n,m,a,\rho<\frac{1}{a})$-AMD code, which contradicts Theorem 4.1.1.

If we have an optimal BSWEDF with parameters $(n,m,K,a,\lambda=\frac{m\widetilde{k}}{a})$, then by Theorem 4.2.1, the corresponding weak AMD code has parameters $(n,m,a,\rho=\frac{1}{a})$, which is G-optimal by Theorem 4.1.1. □

Theorem 4.2.9. *Let n and m be positive integers.*

(I) For given $K=(k_1,k_2,\dots,k_m)$, let $\rho_{(n,m,K)}$ denote the the smallest possible ρ for weak (n,m,K,ρ)-AMD codes. Then a weak (n,m,K,ρ)-AMD code $(S,G,\mathcal{A},E_u)$ has the smallest ρ, i.e., $\rho=\rho_{(n,m,K)}$ if and only if its corresponding BSWEDF with parameters $(n,m,K,a,\lambda=m\widetilde{k}\rho)$ is optimal, where $S=\{s_i:1\le i\le m\}$, $\mathcal{A}=\{A_{s_i}:1\le i\le m\}$, $k_i=|A_{s_i}|$ for $1\le i\le m$, $K=(k_1,k_2,\dots,k_m)$, and $a=\sum_{1\le i\le m}k_i$.

(II) For given a, there exists an R-optimal weak (n,m,a,ρ)-AMD code $(S,G,\mathcal{A},E_u)$ with respect to the bound in Corollary 4.2.3 if and only if there exists a strongly optimal (n,m,K,a,λ)-BSWEDF, where $|G|=n$, $a=\sum_{s\in S}|A_s|$, $\rho=\rho_{(n,m,a)}$, and $\lambda=\rho_{(n,m,a)}\widetilde{k}m$.

(III) *There exists an R-optimal weak (n, m, a, ρ)-AMD code $(S, G, \mathcal{A}, E_u)$ with respect to the bound in Lemma 4.1.5 if and only if there exists an (n, m, K, a, λ)-SWEDF, where $\rho = \frac{a(m-1)}{m(n-1)}$, and $\lambda = \frac{\widetilde{k}a(m-1)}{n-1}$.*

Proof. By Theorem 4.2.1, for given n, m, K (or a, resp.), a weak AMD code with the smallest ρ is equivalent to a BSWEDF with the smallest λ, i.e., an optimal (or strongly optimal, resp.) BSWEDF. The third part of the result follows directly from Theorem 4.2.1 and Lemma 3.4.2. □

Example 4.2.10. *Let $n = 10$, $m = 3$, and $a = 5$. Let*

$$\mathcal{B}^{(1)} = \{B_1^{(1)} = \{5\}, B_2^{(1)} = \{4, 6\}, B_3^{(1)} = \{2, 8\}\}$$

and

$$\mathcal{B}^{(2)} = \{B_1^{(2)} = \{5\}, B_2^{(2)} = \{2\}, B_3^{(2)} = \{0, 4, 6\}\}$$

be two families of disjoint subsets of $\mathbb{Z}_{10}$. It is easy to verify that

$$\bigcup_{1 \le i \le 3} D\left(B_i^{(1)}, \widetilde{B}_j^{(1)}\right) \subseteq 3 \boxtimes (\mathbb{Z}_{10} \backslash \{0\})$$

and

$$\bigcup_{1 \le i \le 3} D\left(B_i^{(2)}, \widetilde{B}_j^{(2)}\right) \subseteq 4 \boxtimes (\mathbb{Z}_{10} \backslash \{0\}).$$

According to Lemma 3.4.2, $\mathcal{B}^{(1)}$ is an optimal $(10, 3, (1, 2, 2), 5, 3)$-BSWEDF and $\mathcal{B}^{(2)}$ is an optimal $(10, 3, (1, 1, 3), 5, 4)$-BSWEDF. By Corollary 4.2.3,

$$\rho_{(10,3,5)} \ge \min_K \left\{ \left\lceil \frac{\widetilde{k}5(3-1)}{10-1} \right\rceil \frac{1}{3\widetilde{k}} : \sum_{1 \le i \le 3} k_i = 5 \right\} = \frac{4}{9}.$$

Thus, by Definition 4.2.2, $\mathcal{B}^{(2)}$ is in fact not only an optimal, but a strongly optimal BSWEDF. By Corollary 4.2.9. (II), we can obtain a corresponding R-optimal weak AMD code with respect to the bound in Corollary 4.2.3 from $\mathcal{B}^{(2)}$.

Although the weak $(n, m, a, \rho_{(n,m,K)} = \frac{\lambda}{\widetilde{k}m})$-AMD code $(S, G, \mathcal{A}, E_u)$ based on an optimal (n, m, K, a, λ)-BSWEDF may sometimes not correspond to an optimal weak AMD code with parameters $(n, m, a, \rho_{(n,m,a)})$, the difference $\rho_{(n,m,K)} - \rho_{(n,m,a)}$ is not big.

Lemma 4.2.11. *Let $a = \sum_{A \in \mathcal{A}} |A| = \sum_{1 \le i \le m} k_i$. Let $(S, G, \mathcal{A}, E_u)$ be the weak $(n, m, a, \rho = \frac{\lambda}{\widetilde{k}m})$-AMD code based on an optimal (n, m, K, a, λ)-BSWEDF with $\lambda = \lceil \frac{\widetilde{k}a(m-1)}{n-1} \rceil$, and let $(S, G, \mathcal{A}', E_u)$ be the R-optimal weak $(n, m, a, \rho_{(n,m,a)})$-AMD code with respect to the bound in Corollary 4.2.3. Then we have*

$$0 \le \rho_{(n,m,K)} - \rho_{(n,m,a)} \le \frac{1}{m\widetilde{k}}.$$

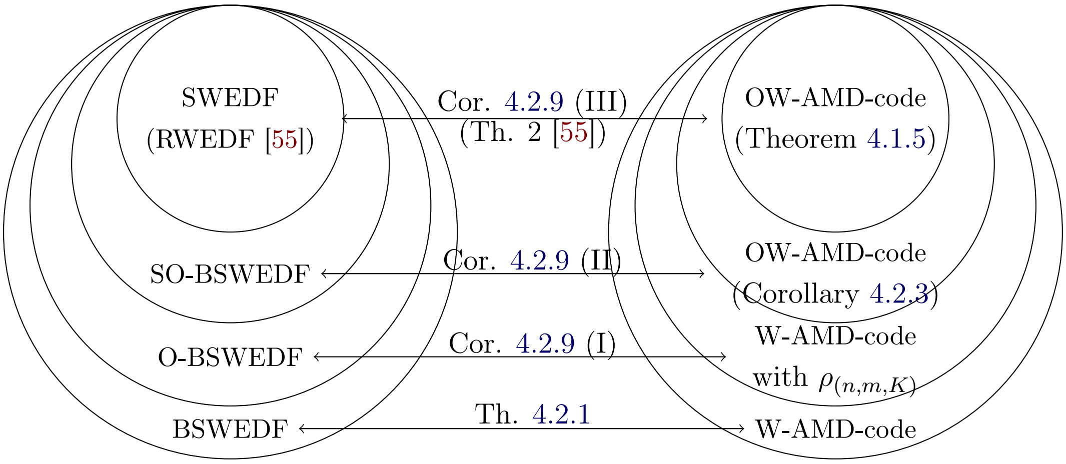

Figure 4.1: The relationships between AMD codes and BSWEDFs

Proof. The lemma follows directly from the fact that

$$\begin{aligned}
0 \leq & \rho_{(n,m,K)} - \rho_{(n,m,a)} \\
= & \left\lceil \frac{\widetilde{k}a(m-1)}{n-1} \right\rceil \frac{1}{m\widetilde{k}} - \rho_{(n,m,a)} \\
\leq & \left\lceil \frac{\widetilde{k}a(m-1)}{n-1} \right\rceil \frac{1}{m\widetilde{k}} - \frac{a(m-1)}{m(n-1)} \leq \frac{1}{m\widetilde{k}}.
\end{aligned}$$

□

In [55], Huczynska and Paterson characterized R-optimal AMD codes by R-WEDFs by using $N_i(\alpha)$. Clearly, $N_i(\alpha) = \#\left(\alpha, \bigcup_{\substack{1\leq j\leq m\\ j\neq i}} D(B_i, B_j)\right)$ for $1 \leq i \leq m$, and by Theorem 4.1.8 and Corollary 4.2.9 or Definitions 3.4.1 and 4.1.7, we know that an (n, m, K, a, d)-RWEDF is essentially the same as an (n, m, K, a, λ)-SWEDF, where $d = \frac{\lambda}{\widetilde{k}}$.

Therefore, Theorem 4.2.1 and Corollary 4.2.9 provide more combinatorial characterizations for various weak AMD codes $(S, G, \mathcal{A}, E_u)$. These results can be viewed as a generalization of Theorem 4.1.8.

In Figure 4.1, we summarize the relationships between weak AMD codes and BSWEDFs, where SO-BSWEDF, O-BSWEDF, and OW-AMD-code denote strongly optimal BSWEDF, optimal BSWEDF, and optimal weak AMD-code, respectively.

4.2.1 Weak algebraic manipulation detection codes from BSWEDFs

By Theorem 4.2.9, BSWEDFs correspond to AMD codes. Thus, according to Theorem 4.2.9 (I) and Theorem 3.5.8, we directly have the following corollary.

Corollary 4.2.12. *If k is odd and $4k+1$ is a prime power, then there exists a weak $(8k+2, 3, (2, 2k, 2k), \frac{2k+1}{6k})$-AMD code.*

Example 4.2.13. *Let $n = 2q = 26$. By Construction 3.5.7, the family of sets $\mathcal{A} = \{A_1, A_2, A_3\}$ over $\mathbb{Z}_{26}$ can be listed as*

$$A_1 = \{0, 13\},\ A_2 = \{14, 16, 22, 17, 25, 23\},\ \textit{and}$$

$$A_3 = \{2, 6, 18, 8, 24, 20\}.$$

It is easy to check that

$$\bigcup_{1 \le i \ne j \le 3} D(A_i, \widetilde{A}_j) = 7 \boxtimes (\mathbb{Z}_{26} \backslash \{0, 13\}),$$

which means that $\mathcal{A}$ is an optimal $(26, 3, (2, 6, 6), 14, 7)$-BSWEDF. Define the encoding map E_u from $S = \{s_1, s_2, s_3\}$ to $\mathbb{Z}_{26}$ as

$$E_u(s_i) = x \in_R A_i,$$

where $x \in_R A_i$ for $1 \le i \le 3$ means that x is chosen from A_i uniformly at random. Then it is easy to check that $(S, \mathbb{Z}_{26}, \mathcal{A}, E_u)$ is a weak AMD code with parameters $(26, 3, (2, 6, 6), \frac{7}{18})$.

Similarly, by Theorem 4.2.9 (I) and Theorem 3.5.12, we have

Corollary 4.2.14. *If there exists an $(n_1, k, k-1)$-PDF with $n_1 = 2k+1$, then there exists a weak $(n = 4k+2, m = 3, (1, k, k), \rho = \frac{k+1}{3k})$-AMD code.*

Example 4.2.15. *Let $n = 2n_1 = 22$. By Construction 3.5.11, the family of sets $\mathcal{W} = \{W_1, W_2, W_3\}$ over $\mathbb{Z}_{22}$ can be listed as*

$$W_1 = \{11\},\ W_2 = \{12, 4, 16, 20, 14\},\ \textit{and}\ W_3 = \{2, 8, 10, 18, 6\}.$$

It is easy to check that

$$\bigcup_{1 \le i \ne j \le 3} D(W_i, \widetilde{W}_j) \subseteq 6 \boxtimes (Z_{22} \backslash \{0\}),$$

which means that $\mathcal{W}$ is an optimal $(22, 3, (1, 5, 5), 11, 6)$-BSWEDF. Define the encoding map E_u from $S = \{s_1, s_2, s_3\}$ to $\mathbb{Z}_{22}$ as

$$E_u(s_i) = x \in_R W_i,$$

where $x \in_R W_i$ for $1 \le i \le 3$ means that x is chosen from W_i uniformly at random. Then it is easy to check that $(S, \mathbb{Z}_{22}, \mathcal{W}, E_u)$ is a weak AMD code with parameters $(22, 3, (1, 5, 5), \frac{2}{5})$.

From Theorem 4.2.9 (I) and Theorem 3.5.15, we have

Corollary 4.2.16. *If $4k+1$ is a prime power, then there exists a weak $(12k+3, 4, (1,1,2k,2k), \frac{2k+1}{8k})$-AMD code.*

Example 4.2.17. *Let $n = 3q = 39$. By Construction 3.5.14, the family of sets $\mathcal{U} = \{U_1, U_2, U_3, U_4\}$ over $\mathbb{Z}_{39}$ can be listed as*

$$U_1 = \{13\},\ U_2 = \{26\},\ U_3 = \{27, 30, 3, 12, 9, 36\},\ \textit{and}$$

$$U_4 = \{15, 21, 6, 24, 18, 33\}.$$

It is easy to check that

$$\bigcup_{1 \le i \ne j \le 4} D(U_i, \widetilde{U}_j) \subseteq 7 \boxtimes (\mathbb{Z}_{39} \backslash \{0\}),$$

which means that $\mathcal{U}$ is an optimal $(39, 4, (1,1,6,6), 14, 7)$-BSWEDF. Define the encoding map E_u from $S = \{s_1, s_2, s_3, s_4\}$ to $\mathbb{Z}_{39}$ as

$$E_u(s_i) = x \in_R U_i,$$

where $x \in_R U_i$ for $1 \le i \le 4$ means that x is chosen from U_i uniformly at random. Then it is easy to check that $(S, \mathbb{Z}_{39}, \mathcal{U}, E_u)$ is a weak AMD code with parameters $(39, 4, (1,1,6,6), \frac{7}{24})$.

Directly from Theorem 4.2.9 (III) and Theorem 3.5.18, we have

Corollary 4.2.18. *Let $\gcd(k-1, tk+1) = 1$. If there exists a cyclic $((k-1)(tk+1), (k, \cdots, k, k-1), k-1)$-PDF over $\mathbb{Z}_{k-1} \times \mathbb{Z}_{tk+1}$ where one of the blocks is of form $\mathbb{Z}_{k-1} \times \{0\}$, then there exists an R-optimal weak $((k-1)(tk+1), K = (k, k, \cdots, k, 1, \cdots, 1), (k-1)(tk+1) - 1, \frac{t(k-1)+k-3}{t(k-1)+k-2})$-AMD code, where 1 appears $k-2$ times in K.*

By Remark 3.5.20, the AMD codes generated by Corollary 4.2.18 based on SWEDFs (or RWEDFs) without bimodal property. However, the only known R-WEDFs without the bimodal property have parameters $(k_1k_2 + 1, m = 2, a = k_1 + k_2, \lambda = 2)$, where $k_1 > 2$ and $k_2 > 2$ are integers. Thus, compared with the constructions in [55], Corollary 4.2.18 can generate new R-optimal weak AMD codes with flexible parameters based on SWEDFs without bimodal property.

Example 4.2.19. *Let $G = (\mathbb{Z}_{15}, +)$ and*

$$\mathcal{R} = \{R_1 = \{6, 9, 2, 8\}, R_2 = \{11, 14, 7, 13\}, R_3 = \{1, 4, 12, 3\}, R_4 = \{0, 5, 10\}\}.$$

It is easy to check that $\mathcal{R}$ is a PDF with parameters $(15,(4,4,4,3),3)$. *By Construction 3.5.17, we generate a family of subsets of* $\mathbb{Z}_{15}$ *as*

$$\begin{aligned}\mathcal{V} =&\{V_1 = \{6,9,2,8\}, V_2 = \{11,14,7,13\}, V_3 = \{1,4,12,3\},\\ &V_4 = \{5\}, V_5 = \{10\}\}.\end{aligned}$$

It is easy to check that

$$\bigcup_{1\leq i\neq j\leq 5} D(V_i,\widetilde{V_j}) = 16 \boxtimes (\mathbb{Z}_{15}\backslash\{0\}),$$

i.e., $\mathcal{V}$ *is a* $(15,5,(4,4,4,1,1),14,16)$*-SWEDF (or* $(15,5,(4,4,4,1,1),14,4)$*-RWEDF). Note that* $N_3(6) = 3 \notin \{0,4\}$*, which means the SWEDF does not have the bimodal property by Definition 3.5.19. Define the encoding map from* $S = \{s_1,s_2,s_3,s_4,s_5\}$ *to* $\mathbb{Z}_{15}$ *as*

$$E_u(s_i) = x \in_R V_i,$$

where $x \in_R V_i$ $1 \leq i \leq 5$ *means that* x *is chosen from* V_i *uniformly at random. Then it is easy to check that* $(S,\mathbb{Z}_{15},\mathcal{V},E_u)$ *is an R-optimal weak AMD code with parameters* $(15,5,(4,4,4,1,1),\frac{4}{5})$.

4.3 Weak algebraic manipulation detection codes from highly nonlinear functions

In this section, we propose a construction for systematic weak AMD codes via highly nonlinear functions, which was implicitly described in [23]. Perfect nonlinear functions were used [20] to analyse deterministic systematic weak AMD codes, that is, their encoding map is a deterministic one such that $E_f : A_1 \to G = A_1 \times B$ as $E_f(S_1) = (S_1, f(S_1))$. Here we give a systematic investigation on systematic weak AMD codes via the partial nonlinearity of the function f as follows.

Construction 4.3.1. *Let* f *be a map from* $A = A_1 \times A_2$ *to* B *and let* $S = A_1$ *be a subgroup of* A*. Define a probabilistic encoding map* $E_f : A_1 \to G = A \times B = A_1 \times A_2 \times B$ *as*

$$E_f(S_1) = (S_1, S_2, f(S_1,S_2)), \tag{4.6}$$

where $S_2 \in_R A_2$.

By the probabilistic encoding map E_f and the corresponding deterministic decoding function given by (2.4), we can define a systematic AMD code (E_f, Dec) from A_1 to $G = A \times B = A_1 \times A_2 \times B$. Note that a possible successful tampering should satisfy $\Delta \in (A_1 \setminus \{0\}) \times A_2 \times B$. However, for the nonlinearity, we should consider all possible $\Delta \in A_1 \times A_2 \times B \setminus \{(0,0,0)\}$. Thus, to the convenience of analysis, we introduce the partial nonlinearity of a function, which only considers the case $\Delta \in (A_1 \setminus \{0\}) \times A_2 \times B$.

Definition 4.3.2. *The partial nonlinearity* $\Psi_f(A_1)$ *of a function* f *from* $A = A_1 \times A_2$ *to* B *is defined as*

$$\begin{aligned}\Psi_f(A_1) &\triangleq \max_{a_1 \in A_1 \setminus \{0\}} \max_{a_2 \in A_2} \max_{b \in B} \Pr\left(D_{(a_1,a_2)}(f(x)) = b\right) \\ &= \max_{a_1 \in A_1 \setminus \{0\}} \max_{a_2 \in A_2} \max_{b \in B} \frac{|\{x \in A : \ D_{(a_1,a_2)}(f(x)) = b\}|}{|A|},\end{aligned} \tag{4.7}$$

where A_1 *is a subgroup of* A*,*

$$D_a(f(x)) = f(x+a) - f(x) \text{ for } a \in A$$

and $\Pr\left(D_{(a_1,a_2)}(f(x)) = b\right)$ *denotes the probability of the occurrence of the event* $D_{(a_1,a_2)}(f(x)) = b$*.*

Remark 4.3.3. *By Definitions 2.4.1 and 4.3.2, we know that* $N_f \geq \Psi_f(A_1)$ *for any subgroup* A_1 *of* A*, where* N_f *denotes the nonlinearity of* f*.*

The parameters of the constructed AMD code have the following relationship with the nonlinearity of f.

Theorem 4.3.4. *If the function* f *from* $A = A_1 \times A_2$ *to* B *with partial nonlinearity* $\Psi_f(A_1)$ *has equiprobable sources and* E_f *is equiprobable encoding, then the systematic weak AMD code* (E_f, Dec) *generated by Construction 4.3.1 has parameters* $(n_1, n_1 n_2 m, \Psi_f(A_1) \leq N_f)$*, where* $|A_1| = n_1$*,* $|A_2| = n_2$*, and* $|B| = m$*.*

Proof. By Construction 4.3.1, we only need to prove that the probability of successful tampering is upper bounded by $\Psi_f(A_1)$. For any $\Delta = (a_1, a_2, b) \in G = A_1 \times A_2 \times B$ with $a_1 \in A_1 \setminus \{0\}$, $a_2 \in A_2$ and $b \in B$,

$$\begin{aligned} & \sum_{S_1 \in A_1} \Pr(S' = S_1) \sum_{S_2 \in A_2} \Pr(E_f(S_1) = (S_1, S_2, f(S_1, S_2)) \\ & \times \Pr(\mathrm{Dec}((S_1, S_2, f(S_1, S_2)) + \Delta) \notin \{S_1, \perp\}) \\ = & \sum_{S_1 \in A_1} \Pr(S' = S_1) \sum_{S_2 \in A_2} \Pr(S^* = S_2) \\ & \times \Pr(\mathrm{Dec}((S_1, S_2, f(S_1, S_2)) + \Delta) \notin \{S_1, \perp\}) \qquad (4.8) \\ = & \sum_{S_1 \in A_1} \frac{1}{|A_1|} \sum_{S_2 \in A_2} \frac{1}{|A_2|} \Pr(f(S_1 + a_1, S_2 + a_2) = f(S_1, S_2) + b) \\ = & \frac{1}{|A_1||A_2|} \sum_{S_1 \in A_1} \sum_{S_2 \in A_2} \Pr(f(S_1 + a_1, S_2 + a_2) = f(S_1, S_2) + b) \\ = & \frac{1}{|A_1||A_2|} |\{(S', S^*) \in A : \ f(S' + a_1, S^* + a_2) - f(S', S^*) = b\}| \\ \leq & \Psi_f(A_1) \\ \leq & N_f. \qquad (4.9) \end{aligned}$$

According to Definitions 2.3.5 and 4.3.2, we know that the systematic weak AMD code (E_f, Dec) generated by Construction 4.3.1 has parameters $(n_1, n_1 n_2 m, \Psi_f(A_1) \leq N_f)$, which completes the proof. □

For deterministic systematic weak AMD codes, Theorem 4.3.4 is similar to [2, Theorem 2]. However, our construction also works for general case of systematic weak AMD codes with a random part. It is a generalization of [2, Theorem 2] for the case of a map f of two variables.

In what follows, we list some well-known highly nonlinear functions and their corresponding systematic AMD codes as applications of Construction 4.3.1.

4.3.1 Linear functions

One simple but useful way to obtain functions with high nonlinearity is to use linear functions from $(\mathbb{F}_{q^r}, +)$ to $(\mathbb{F}_q, +)$ as functions from $(\mathbb{F}_{q^r}^*, \times) \cong (\mathbb{Z}_{q^r-1}, +)$ to $(\mathbb{F}_q, +)$.

Lemma 4.3.5 ([16])**.** *Any nonzero linear function L from $(\mathbb{F}_{q^r}, +)$ to $(\mathbb{F}_q, +)$ is a function from $(\mathbb{F}_{q^r}^*, \times)$ to $(\mathbb{F}_q, +)$ with nonlinearity $N_f = \frac{1}{q} + \frac{1}{q(q^r-1)}$.*

Applying the highly nonlinear functions in Lemma 4.3.5, the following corollary follows directly from Construction 4.3.1 and Theorem 4.3.4.

Corollary 4.3.6. *Let $A = (\mathbb{Z}_{q^r-1}, +)$ and $B = (\mathbb{F}_q, +)$. Further let $A_1 = (\mathbb{Z}_{m_1}, +)$ and $A_2 = (\mathbb{Z}_{m_2}, +)$ be two subgroups of A with order m_1 and $m_2 = \frac{q^r-1}{m_1}$, respectively. If $\gcd(m_1, m_2) = 1$, then $A \cong A_1 \times A_2$. Define the probabilistic encoding map E_L from A_1 to $G = A_1 \times A_2 \times B$ as*

$$E_L(s_1) = (s_1, s_2, L(\Phi(s_1, s_2))), \tag{4.10}$$

where $s_2 \in_R A_2$, Φ is an isomorphism from $\mathbb{Z}_{q^r-1}$ to $(\mathbb{F}_{q^r}^*, \times)$, and $L(x)$ is a nonzero linear function from $(\mathbb{F}_{q^r}^*, \times)$ to $(\mathbb{F}_q, +)$. If the systematic weak AMD code (E_L, Dec) given by (4.10) and (2.4) has equiprobable sources and E_L is equiprobable encoding, then it is an m_2-regular $(m_1, (q^r-1)q, \frac{1}{q} + \frac{1}{q(q^r-1)})$-AMD code.

Corollary 4.3.7. *Let $r \in \mathbb{N}$ and m_2 be a factor of $q^r - 1$. Further let $m_1 = \frac{q^r-1}{m_2}$, $u = \lfloor \log m_1 \rfloor$, and $k = \lfloor \log \frac{q^r-1}{q^r-1} \rfloor$. If $\gcd(m_1, m_2) = 1$, then the effective tag size $\varpi^*(k, u)$ for weak AMD codes satisfies*

$$k - 1 \leq \varpi^*(k, u) < k + 1 + \log \frac{m_2 q^r}{q^r - 1}.$$

The systematic weak AMD code in Corollary 4.3.6 has an asymptotically optimal effective tag size with respect to the bound in Lemma 2.3.7, i.e.,

$$\lim_{k \to \infty} \frac{\log |G| - \log |A_1|}{k - 1} = 1.$$

Proof. By Corollary 4.3.6, there exists a systematic weak AMD code with $|A_1| = \frac{q^r-1}{m_2} \geq 2^u$, $\rho = \frac{q^{r-1}}{q^r-1} \leq 2^{-k}$, and the tag size

$$\varpi = \log|G| - \log|A_1| = \log(m_2 q) = \log m_2 + \log q.$$

Note that

$$\begin{aligned}\varpi - k =& \log m_2 + \log q - \left\lfloor \log \frac{q^r-1}{q^{r-1}} \right\rfloor \\ <& 1 + \log m_2 + \log \frac{q^r}{q^r-1}.\end{aligned}$$

The first conclusion then follows from the fact that for any $k, u \in \mathbb{N}$, we have $k-1 \leq \varpi^*(k,u) \leq \varpi$. The second conclusion can be derived by the fact that

$$\lim_{k\to\infty} \frac{\log|G| - \log|A_1|}{k-1} = \lim_{q\to\infty} \frac{1+\log m_2 + \log\frac{q^r}{q^r-1} + k}{k-1} = 1$$

by noting that $m_2 \in \mathbb{N}$ is a constant. □

Example 4.3.8. *Let L be the trace function from $\mathbb{F}_{2^4}$ to $\mathbb{F}_2$, i.e.,*

$$L(x) = \mathrm{Tr}_2^{2^4}(x) = x + x^2 + x^4 + x^8.$$

Set $m_1 = 5$ and $m_2 = 3$. Define a probabilistic encoding function from $\mathbb{Z}_5 \times \mathbb{Z}_3$ to $\mathbb{Z}_5 \times \mathbb{Z}_3 \times \mathbb{Z}_2$ as

$$E_L((x_1)) = (x_1, x_2, L(\alpha^{\phi(x_1,x_2)})),$$

where $x_2 \in_R \mathbb{Z}_3$, α is a primitive element of $\mathbb{F}_{2^4}$ and $\alpha^{\phi(x_1,x_2)} = \alpha^{(6x_1+10x_2)}$ is an isomorphism from $\mathbb{Z}_5 \times \mathbb{Z}_3$ to $\mathbb{F}_{2^4}^*$. Let $\mathcal{A} = \{A_i \ : \ i \in \mathbb{Z}_5\}$, where

$$A_i = \{(i, x_2, L(\alpha^{\phi(i,x_2)})) \ : \ x_2 \in \mathbb{Z}_3\}.$$

It is easy to check that $(\mathbb{Z}_5, \mathbb{Z}_5 \times \mathbb{Z}_3 \times \mathbb{Z}_2, \mathcal{A}, E_L)$ is a weak AMD code with parameters $(30, 5, 15, \frac{8}{15})$.

4.3.2 Maiorana-McFarland's class of functions

Let $r \in \mathbb{N}$ and q be a prime power. Define a function $f : (\mathbb{F}_q^{2r}, +) \to (\mathbb{F}_q, +)$ as

$$f(x_1, x_2, \ldots, x_{2r}) = \sum_{1\leq i\leq r} x_i x_{i+r}. \tag{4.11}$$

Lemma 4.3.9 ([70]). *The function $f(x_1, x_2, \ldots, x_{2r})$ defined by (4.11) has perfect nonlinearity $N_f = \frac{1}{q}$.*

Corollary 4.3.10. *Let $A_1 = (\mathbb{F}_{q^{2r-1}}, +)$ and $A_2 = B = (\mathbb{F}_q, +)$, where we regard an element of $\mathbb{F}_{q^{2r-1}}$ as a vector in $\mathbb{F}_q^{2r-1}$. Define the probabilistic encoding map E_f from A_1 to $G = A_1 \times A_2 \times B$ as*

$$\begin{aligned} E_f(S_1) =& (S_1, s_2, f(S_1, s_2)) \\ =& \left(x_1, x_2, \ldots, x_{2r-1}, s_2, x_r s_2 + \sum_{1 \leq i \leq r-1} x_i x_{i+r} \right), \end{aligned} \tag{4.12}$$

where $S_1 = (x_1, x_2, \ldots, x_{2r-1}) \in A_1$, $s_2 \in_R A_2$, and f is defined by (4.11). If the systematic weak AMD code (E_f, Dec) given by (4.12) and (2.4) has equiprobable sources and E_f is equiprobable encoding, then it is an R-optimal q-regular $(q^{2r-1}, q^{2r+1}, \frac{1}{q})$-AMD code with respect to the bound in Corollary 4.2.4.

Proof. The statement that the constructed AMD code (E_f, Dec) has parameters $(q^{2r-1}, q^{2r+1}, \frac{1}{q})$ directly follows from Theorem 4.3.4, Lemma 4.3.9, and the fact that it is q-regular. According to Corollary 4.2.4, we should have

$$\rho \geq \left\lceil \frac{q^2 q^{2r-1}(q^{2r-1} - 1)}{q^{2r+1} - 1} \right\rceil \frac{1}{q^{2r}} = \frac{1}{q},$$

which means that the constructed AMD code is R-optimal. □

The encoding map (4.11) appeared in [2, 20, 63] without requiring a random $s_2 \in_R A_2$, i.e., it is deterministic.

Corollary 4.3.11. *For any $k, u \in \mathbb{N}$, the effective tag size $\varpi^*(k, u)$ for weak AMD codes is bounded as follows:*

$$k - 1 \leq \varpi^*(k, u) \leq 2k.$$

Proof. For any given k and u, choose $q = 2^k$ and r to be the smallest positive integer such that $u \leq k(2r - 1)$. According to Corollary 4.3.10, there exists a systematic weak AMD code with $|A_1| = q^{2r-1} \geq 2^u$, $\rho = \frac{1}{q} \leq 2^{-k}$, and the tag size $\varpi = \log|G| - \log|A_1| = 2\log q = 2k$. Then the claim follows from the fact that $k - 1 \leq \varpi^*(k, u) \leq \varpi$. □

Example 4.3.12. *Let*

$$f(x_1, x_2, x_3, x_4) = \sum_{1 \leq i \leq 2} x_i x_{i+2}$$

be a function from $\mathbb{F}_2^4 \cong \mathbb{F}_{2^4}$ to $\mathbb{F}_2$. Define a probabilistic encoding function from $\mathbb{F}_2^3$ to $\mathbb{F}_2^5$ as

$$E_f((x_1, x_2, x_3, x_4)) = (x_1, x_2, x_3, x_4, F(x_1, x_2, x_3, x_4)),$$

where $x_4 \in_R \mathbb{F}_2$. Let $\mathcal{A} = \{A_i \ : \ i \in \mathbb{F}_2^3\}$, where

$$A_i = \{(i, x_4, f(i, x_4)) \ : \ x_4 \in \mathbb{F}_2\}.$$

It is easy to check that $(\mathbb{F}_2^3, \mathbb{F}_2^5, \mathcal{A}, E_f)$ is a weak AMD code with parameters $(32, 8, 16, \frac{1}{2})$.

4.3.3 Dillon's class of functions

In this subsection, we recall the well-known Dillon's class of functions with perfect nonlinearity. A function $g : A \to B$ is *balanced* if the size of $g^{-1}(b)$ is the same for every $b \in B$, which is $|A|/|B|$. It is known (see, for example, [16]) that g has perfect nonlinearity if and only if for every $a \in A \setminus \{0\}$, the derivative $D_a(g(x))$ is balanced, and this is possible only when $|B|$ divides $|A|$.

Lemma 4.3.13 ([30]). *For any $r \in \mathbb{N}$, let $\mathbb{F}_q^r$ be identified with the finite field $\mathbb{F}_{q^r}$ and let g be any balanced function from $\mathbb{F}_{q^r}$ to $\mathbb{F}_q$. Then the function $f : (\mathbb{F}_{q^{2r}}, +) \to (\mathbb{F}_q, +)$ defined by*

$$f(x, y) = g(xy^{q^r-2}), \quad x, y \in \mathbb{F}_{q^r}$$

has perfect nonlinearity $N_f = \frac{1}{q}$.

Corollary 4.3.14. *Let $A_1 = (\mathbb{F}_{q^{2r-1}}, +)$ and $A_2 = B = (\mathbb{F}_q, +)$, where we regard an element of $\mathbb{F}_{q^{2r-1}}$ as a vector in $\mathbb{F}_q^{2r-1}$. Let $g : \mathbb{F}_{q^r} \to \mathbb{F}_q$ be a balanced function. Define the probabilistic encoding map E_f from A_1 to $G = A_1 \times A_2 \times B$ as*

$$\begin{aligned} E_f(S_1) =& (S_1, y_r, f(X, Y)) \\ =& \left(x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_{r-1}, y_r, g(XY^{q^r-2})\right), \end{aligned} \tag{4.13}$$

where $S_1 = (x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_{r-1}) \in A_1$, $y_r \in_R A_2$, $X = (x_1, x_2, \ldots, x_r)$, and $Y = (y_1, y_2, \ldots, y_r)$. If the systematic weak AMD code (E_f, Dec) given by (4.13) and (2.4) has equiprobable sources and E_f is equiprobable encoding, then it is an R-optimal q-regular $(q^{2r-1}, q^{2r+1}, \frac{1}{q})$-AMD code with respect to the bound in Corollary 4.2.4.

The proof of Corollary 4.3.14 is similar to that of Corollary 4.3.10 so we omit it here. Note that although the parameters of the systematic weak AMD codes constructed in Corollaries 4.3.10 and 4.3.14 are the same, their probabilistic encoding maps are different.

Remark 4.3.15. *In the past few decades, highly nonlinear functions have received much attention, and many results have been reported in this topic. For more general form of functions with perfect nonlinearity, similar to f in (4.11), the interested reader is referred to [26, 58, 64, 70, 71].*

Example 4.3.16. *Let the balanced function $g(x)$ be the trace function $\mathrm{Tr}_q^{q^2} : \mathbb{F}_{q^2} \to \mathbb{F}_q$ defined by*

$$\mathrm{Tr}_q^{q^2}(x) = x + x^q,$$

then

$$f(X = (x_1, x_2), Y = (y_1, y_2)) = \mathrm{Tr}_q^{q^2}(XY^{q^2-2}),$$

is a function from $\mathbb{F}_{q^2}^2 \cong \mathbb{F}_{q^4}$ *to* $\mathbb{F}_q$*, where we regard the element of* $\mathbb{F}_q^2$ *as a vector of length* 2 *over* $\mathbb{F}_q$*. Define a probabilistic encoding map from* $\mathbb{F}_q^3$ *to* $\mathbb{F}_q^5$ *as*

$$E_f((x_1, x_2, x_3, x_4)) = (x_1, x_2, x_3, x_4, f(X = (x_1, x_2), Y = (x_3, x_4))),$$

where $x_4 \in_R \mathbb{F}_q$*. Let* $\mathcal{A} = \{A_i \ : \ i \in \mathbb{F}_q^3\}$*, where*

$$A_{i=(x_1,x_2,x_3)} = \{(x_1, x_2, x_3, x_4, f(X = (x_1, x_2), Y = (x_3, x_4))) \ : \ x_4 \in \mathbb{F}_q\}.$$

It is easy to check that $(\mathbb{F}_q^3, \mathbb{F}_q^5, \mathcal{A}, E_f)$ *is a weak AMD code with parameters* $(q^5, q^3, q^4, \frac{1}{q})$*.*

Chapter 5

Strong Algebraic Manipulation Detection Codes

In this chapter, we study strong algebraic manipulation detection (AMD) codes, i.e., under the assumption that the adversary has full knowledge about the source. We investigate the relationships between systematic strong AMD codes and highly nonlinear functions. By means of relationships, we construct infinite families of systematic strong AMD codes and highly nonlinear functions, respectively. A combinatorial construction for systematic strong AMD codes is also provided.

5.1 Known results about strong algebraic manipulation detection codes

In this section we recall some known results about strong algebraic manipulation detection codes.

5.1.1 Strong algebraic manipulation detection codes: bounds

For strong algebraic manipulation detection codes, the following theoretic bounds on the parameters are known.

Theorem 5.1.1 ([81]). *For any strong AMD code $(S, G, \mathcal{A}, E)$ with parameters (n, m, a, ρ), we have*

$$\rho \geq \frac{a - \min\{|A_i| \ : \ A_i \in \mathcal{A}\}}{n-1}. \tag{5.1}$$

Theorem 5.1.2 ([81]). *For any strong AMD code $(S, G, \mathcal{A}, E)$ with parameters (n, m, a, ρ), we have*

$$\rho_s \geq \frac{1}{|A_s|} \tag{5.2}$$

for any source $s \in S$, where ρ_s is the probability of successful tampering given the source $s \in S$ for a random chosen Δ.

Definition 5.1.3 ([81]). *A strong AMD code is G-optimal if its parameters meet the bound in Lemma 5.1.2 with equality. Here, "G" indicates that guessing the most likely encoding is an optimal strategy for the adversary.*

Theorem 5.1.4 ([28]). *For any systematic strong AMD code* $(G_1, G = G_1 \times G_2 \times B, \{\{(s, x, f(s, x)) : x \in G_2\} : s \in G_1\})$ *with parameters* $(|G|, |G_1|, (|G_2|, \cdots, |G_2|), \rho)$, *we have*

$$\rho \geq \frac{1}{|G_2|}$$

and

$$\rho \geq \frac{1}{|B|}.$$

5.1.2 Strong algebraic manipulation detection codes: algebraic constructions

For strong algebraic manipulation detection codes, Cramer *et al.* introduced a construction of strong AMD codes with nearly optimal tag size via polynomial evaluations [27].

Construction 5.1.5 ([27]). *Let* $\mathbb{F}_q$ *be a finite field with characteristic* p. *Define a probabilistic encoding mapping* $E_h : \mathbb{F}_{q^t} \to \mathbb{F}_{q^t} \times \mathbb{F}_q \times \mathbb{F}_q$ *given by*

$$E_h(S = (s_1, s_2, \cdots, s_t)) \to (S, x_2, h(S, x_2))$$

with $x_2 \in_R \mathbb{F}_q$ *and*

$$h(S, x) = x^{t+2} + \sum_{1 \leq t \leq t} s_i x^i. \tag{5.3}$$

Theorem 5.1.6 ([27]). *If* $p \nmid t + 2$, *then the systematic strong AMD code generated by Construction 5.1.5 has parameters* $(q^{t+2}, q^t, q^{t+1}, \frac{d+1}{q})$.

In [29], Cramer *et al.* introduced an explicit construction of systematic strong AMD codes based on linear error correcting codes.

Definition 5.1.7. *For a linear code* $\mathcal{C}$ *over* $\mathbb{F}_q$ *with parameters* $[n, k, d]_q$, *the codewords* $C_1 = (c_{1,1}, c_{1,2}, \ldots, c_{1,n}) \in \mathcal{C}$ *and* $C_2 = (c_{2,1}, c_{2,2}, \ldots, c_{2,n}) \in \mathcal{C}$ *is said to be AMD-equivalent if there exists a pair* $(x, y) \in \mathbb{F}_q \times \mathbb{Z}_n$ *such that*

$$L^y(C_1 + x) \triangleq (c_{1,y} + x, c_{1,y+1} + x, \ldots, c_{1,y-1} + x) = C_2.$$

Construction 5.1.8 ([29]). *For a linear code* $\mathcal{C}$ *over* $\mathbb{F}_q$ *with parameters* $[n, k, d]_q$, *let* $\overline{\mathcal{C}} = \{C_1, C_2, \ldots, C_m\} \subseteq \mathcal{C}$ *be a subset of codewords such that any two codewords of* $\overline{\mathcal{C}}$ *are not AMD-equivalent. Then there exists a strong AMD code* $(\mathbb{Z}_m, \mathbb{Z}_m \times \mathbb{Z}_n \times \mathbb{F}_q, \mathcal{A}, E_c)$, *where the encoding function* E_c *from* $\mathbb{Z}_m$ *to* $\mathbb{Z}_m \times \mathbb{Z}_n \times \mathbb{F}_q$ *can be defined as*

$$E_c(i) = (i, x, c_{i,x}),$$

with $x \in_R \mathbb{Z}_n$ *and* $\mathcal{A} = \{A_i \ : \ i \in \mathbb{Z}_n\}$

$$A_i = \{(i, x, c_{i,x}) \ : \ x \in \mathbb{Z}_n\}.$$

Lemma 5.1.9 ([29])**.** *If the linear code $\mathcal{C}$ over $\mathbb{F}_q$ has parameters $[n, k, d]_q$, then the strong AMD code $(\mathbb{Z}_m, \mathbb{Z}_m \times \mathbb{Z}_n \times \mathbb{F}_q, \mathcal{A}, E_c)$ generated by Construction 5.1.8 has parameters $(mnq, |\overline{\mathcal{C}}|, mn, \rho = \frac{n-d}{n})$.*

5.1.3 Strong algebraic manipulation detection codes: combinatorial constructions

In [81], it was proved that external difference families are important tools to construct strong algebraic manipulation detection codes. In this subsection, we recall some of those relationships between external difference families and strong algebraic manipulation detection codes.

Lemma 5.1.10 ([81])**.** *If there exists an $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-GSEDF, there exists an R-optimal AMD code with $a = \sum_{1 \leq i \leq m} k_i$.*

Lemma 5.1.11 ([81])**.** *If there exists an R-optimal strong AMD code with equiprobable encoding, then $\{A_s : s \in S\}$ forms an $(n, m, k_1, k_2, \cdots k_m; \lambda_1, \lambda_2, \cdots, \lambda_m)$-GSEDF.*

Lemma 5.1.12 ([81])**.** *A G-optimal strong AMD code is equivalent to an $(n, m; k_1, \ldots, k_m; 1, \ldots, 1)$-BGSEDF*

Remark 5.1.13. *Although GSEDFs and BGSEDFs are closely related with strong AMD codes, the problem of finding explicit constructions for GSEDFs and BGSEDFs is still widely open.*

5.1.4 From weak AMD codes to strong AMD codes

In [27], Cramer *et al.* introduced a method to modify weak AMD codes to strong AMD codes. In this subsection, we recall this method as the main known relationship between weak AMD codes and strong AMD codes. The main idea is to apply an authentication code and weak AMD codes to encoding the source messages and the random redundance at the same time.

Definition 5.1.14 ([87])**.** *A systematic authentication code is a four-tuple (S, T, K, Au), where S denotes the source space, T is the tag space, K is the key space, and $Au : S \times K \to T$ is the encode function. A transmitter and a receiver share a secret $k \in K$. The transmitter communicates a piece of information $s \in S$ to the receiver by encoding s using A_u into $(s, Au(s, k)) \in S \times T$ and transmits the message $m = (s, Au(s, k))$ by a public channel. For each received message $m^* = (s^*, t^*)$, the receive checks whether $t^* = Au(s^*, k)$. If yes, the receiver will accept the message m^*, otherwise the receiver will reject it.*

Construction 5.1.15 ([27]). *Let Au be an encoding function of a systematic authentication code from $S \times K$ to T, i.e, $(x, k) \mapsto Au(x, k) \in T$, where $k \in_R K$ denotes the private key for the source $x \in S$. Let $(K, G, \mathcal{A}, E)$ be a weak AMD code. Then we can define an encoding function E_s from S to $S \times G \times T$ as*

$$E_s(x) = (s, E(k), Au(s, k)),$$

where $k \in_R K$ is a key chosen at random. Define $\mathcal{A}' = \{A'_i \ : \ i \in S\}$,

$$A'_i = \{(i, E(k), Au(s, k)) : k \in K\}.$$

Theorem 5.1.16 ([27]). *The code $(S, G' = S \times G \times T, \mathcal{A}', E_s)$ is a systematic strong AMD code with parameters $(|G'|, |S|, |S||K|, \rho')$, where $\rho' = \rho + p_{au}$. Herein, ρ and p_{au} denote the probability of successful tampering for the weak AMD code $(S, G, \mathcal{A}, E)$ and the probability of successful substitution attack, i.e., the maximum over all $s \neq s' \in S$ of the probability of successfully substituting the authenticated s by s', for the authentication code (S, T, K, Au) in Construction 5.1.15, respectively. Furthermore, if the underlying weak AMD code is systematic then $\rho' = \max\{\rho, p_{au}\}$.*

5.2 Strong algebraic manipulation detection codes via highly nonlinear functions

In this section, we consider the systematic strong AMD codes generated by Construction 4.3.1 via highly nonlinear functions. We first analyze the relationship between the nonlinearity of the function f and the probability of successful tampering in Theorem 5.2.1. By choosing some special functions, we construct systematic strong AMD codes as examples in Corollaries 5.2.2 and 5.2.5.

By the probabilistic encoding map E_f given by (4.6) and the corresponding decoding function given by (2.4), we can define a systematic AMD code (E_f, Dec) from A_1 to $G = A \times B = A_1 \times A_2 \times B$. In what follows, we first analyze the relationship between parameters of the strong AMD code generated by Construction 4.3.1 and the nonlinearity of f.

Theorem 5.2.1. *Let f be a function from $A = A_1 \times A_2$ to B with nonlinearity N_f and partial nonlinearity $\Psi_f(A_1)$, where $|A_1| = n_1, |A_2| = n_2$, and $|B| = m$. For the equiprobable encoding case, the systematic strong AMD code (E_f, Dec) generated by Construction 4.3.1 has parameters $(n_1, n_1 n_2 m, \rho)$ if and only if for any given $S_1 \in A_1$,*

$$\frac{|\{S^* \in A_2 \ : \ f(S_1 + a_1, S^* + a_2) = f(S_1, S^*) + b\}|}{|A_2|} \leq \rho \tag{5.4}$$

holds for any $\Delta = (a_1, a_2, b) \in (A_1 \setminus \{0\}) \times A_2 \times B$. The parameter ρ satisfies that $\rho \geq \Psi_f(A_1) \geq \frac{1}{|B|}$. Furthermore, if f is a perfect nonlinear function, then we have $\rho \geq N_f$.

Proof. By Construction 4.3.1 and Definition 2.3.5, the AMD code (E_f, Dec) generated by Construction 4.3.1 has parameters (n_1, n_1n_2m, ρ) if and only if for any $S_1 \in A_1$,

$$\Pr(\mathrm{Dec}(E_f(S_1)+\Delta) \notin \{S_1, \perp\}) \le \rho \tag{5.5}$$

holds for any $\Delta = (a_1, a_2, b) \in G$ with $a \in A_1 \setminus \{0\}$, $a_2 \in A_2$, and $b \in B$. But

$$\begin{aligned} & \Pr(\mathrm{Dec}(E_f(S_1)+\Delta) \notin \{S_1, \perp\}) \\ = & \sum_{S^* \in A_2} \Pr(S_2 = S^*) \Pr(f(S_1+a_1, S^*+a_2) = f(S_1, S^*)+b) \\ = & \sum_{S^* \in A_2} \frac{1}{|A_2|} \Pr(f(S_1+a_1, S^*+a_2) = f(S_1, S^*)+b) \\ = & \frac{|\{S^* \in A_2 : \ f(S_1+a_1, S^*+a_2) = f(S_1, S^*)+b\}|}{|A_2|}, \end{aligned}$$

where the second equality follows from the fact that E_f is equiprobable encoding. Therefore, (5.5) is equivalent with (5.4).

Now we prove $\rho \ge \Psi_f(A_1)$. Clearly, there exists a fixed $\Delta = (a_1, a_2, b) \in (A_1 \setminus \{0\}) \times A_2 \times B$ such that $\Psi_f(A_1) = \Pr(D_{(a_1,a_2)}(f(S_1, S_2)) = b)$. Then

$$\begin{aligned} \Psi_f(A_1) &= \Pr(D_{(a_1,a_2)}(f(S_1, S_2)) = b) \\ &= \frac{\sum\limits_{S_1 \in A_1} |\{S^* \in A_2 : \ f(S_1+a_1, S^*+a_2) = f(S_1, S^*)+b\}|}{|A_1||A_2|} \\ &\le \frac{\sum\limits_{S_1 \in A_1} \rho}{|A_1|} \\ &= \rho. \end{aligned}$$

For any $a_1 \in A_1 \setminus \{0\}$ and $a_2 \in A_2$,

$$\begin{aligned} & \max_{b \in B} \frac{|\{(S', S^*) \in A_1 \times A_2 : D_{(a_1,a_2)}(f(S', S^*)) = b\}|}{|A_1||A_2|} \\ \ge & \frac{\sum\limits_{b \in B} \frac{|\{(S', S^*) \in A_1 \times A_2 : \ D_{(a_1,a_2)}(f(S', S^*)) = b\}|}{|A_1||A_2|}}{|B|} \\ = & \frac{1}{|B|} \end{aligned}$$

implies that $\Psi_f(A_1) \ge \frac{1}{|B|}$.

At last, if f is a perfect nonlinear function, then we have $\frac{1}{|B|} = N_f \ge \Psi_f(A_1) \ge \frac{1}{|B|}$ according to Remark 4.3.3. Thus, we have $\rho \ge \Psi_f(A_1) = N_f = \frac{1}{|B|}$. □

In what follows, we list a few systematic strong AMD codes with $\rho = \frac{1}{|B|}$. Especially, we include the classes of Maiorana-McFarland functions and Dillon functions to construct such AMD codes.

Based on Theorem 5.2.1 and Lemma 4.3.9, we have the following corollary.

Corollary 5.2.2. *Let $A_1 = A_2 = \mathbb{F}_{q^r}$ and $B = \mathbb{F}_q$, where we regard an element of $\mathbb{F}_{q^r}$ as a vector in $\mathbb{F}_q^r$. Define the probabilistic encoding map E_f from A_1 to $G = A_1 \times A_2 \times B$ as*

$$\begin{aligned} E_f(S_1) =& (S_1, S_2, f(S_1, S_2)) \\ =& \left(x_1, x_2, \dots, x_r, x_{r+1}, \dots, x_{2r}, \sum_{1 \le i \le r} x_i x_{i+r} \right), \end{aligned}$$

where $S_1 = (x_1, x_2, \dots, x_r) \in A_1$, $S_2 = (x_{r+1}, x_{r+2}, \dots, x_{2r}) \in_R A_2$, and f is defined by (4.11). Then, for the equiprobable encoding case, the systematic strong AMD code given by E_f has parameters $(q^r, q^{2r+1}, \frac{1}{q})$, where $\rho = \frac{1}{q}$ is minimum with respect to Theorem 5.2.1. Especially, when $r = 1$, the q-regular AMD code is optimal with respect to the bound in Lemma 5.1.4.

Proof. We first prove $\rho = N_f = \frac{1}{q}$. By Theorem 5.2.1, we only need to prove that for any $S_1 \in \mathbb{F}_{q^r}$, $a_1 \in \mathbb{F}_{q^r} \setminus \{0\}$, $a_2 \in \mathbb{F}_{q^r}$, and $b \in \mathbb{F}_q$,

$$\frac{|\{S_2 \in \mathbb{F}_{q^r} : \ f(S_1 + a_1, S_2 + a_2) = f(S_1, S_2) + b\}|}{q^r} \le \frac{1}{q}, \tag{5.6}$$

i.e., $f(S_1 + a_1, S_2 + a_2) = f(S_1, S_2) + b$ has at most q^{r-1} solutions for $S_2 \in \mathbb{F}_{q^r}$. Since $a_1 \neq 0$, without loss of generality, we may assume $a_1 = (a_{11}, a_{12}, \dots, a_{1r})$ with $a_{1r} \neq 0$. Note that

$$\begin{aligned} & f(S_1 + a_1, S_2 + a_2) - f(S_1, S_2) - b \\ =& h(S + \Delta) - h(S) + (x_r + a_{1r})(x_{2r} + a_{2r}) - x_r x_{2r} - b \\ =& h(S + \Delta) - h(S) + a_{1r} x_{2r} + a_{2r} x_r + a_{1r} a_{2r} - b, \end{aligned} \tag{5.7}$$

where

$$\begin{aligned} h(S) = h(S_1, S_2) =& h(x_1, x_2, \dots, x_{2r}) \\ \triangleq& \begin{cases} \sum_{1 \le i \le r-1} x_i x_{i+r}, & r \ge 2, \\ 0, & r = 1, \end{cases} \end{aligned}$$

$$\Delta = (a_1, a_2) = (a_{11}, a_{12}, \dots, a_{1r}, a_{21}, a_{22}, \dots a_{2r}),$$

and $a_2 = (a_{21}, a_{22}, \dots, a_{2r})$. For any given $(x_1, x_2, \dots, x_{2r-1}) \in \mathbb{F}_q^{2r-1}$, $a_1 \in \mathbb{F}_{q^r} \backslash \{0\}$, and $a_2 \in \mathbb{F}_{q^r}$, the fact $h(S+\Delta) - h(S) + a_{1r} x_{2r} + a_{2r} x_r + a_{1r} a_{2r} - b = 0$ has at most one solution $x_{2r} \in \mathbb{F}_q$ implies that (5.7) has at most q^{r-1} solutions for all possible $S_2 \in \mathbb{F}_{q^r}$, i.e., (5.6) holds. Then $\rho = N_f = \frac{1}{q}$ by Theorem 5.2.1.

The second assertion is obvious from the definitions. □

Remark 5.2.3. *(1) When q is a power of 2, the AMD codes above based on Maiorana-McFarland's class of functions are special cases of the codes based on Reed-Muller codes [60].*

(2) For more general form of functions with perfect nonlinearity, similar to f in (4.11), the interested reader is referred to [26, 58, 64, 70].

Example 5.2.4. *Let*

$$f(x_1, x_2, x_3, x_4) = \sum_{1 \leq i \leq 2} x_i x_{i+2}$$

be a function from $\mathbb{F}_2^4 \cong \mathbb{F}_{2^4}$ to $\mathbb{F}_2$. Define a probabilistic encoding function from $\mathbb{F}_2^2$ to $\mathbb{F}_2^5$ as

$$E_f((x_1, x_2, x_3, x_4)) = (x_1, x_2, x_3, x_4, F(x_1, x_2, x_3, x_4)),$$

where $(x_3, x_4) \in_R \mathbb{F}_2^2$. Let $\mathcal{A} = \{A_i \ : \ i \in \mathbb{F}_2^2\}$, where

$$A_i = \{(i, x_3, x_4, f(i, x_4)) \ : \ (x_3, x_4) \in \mathbb{F}_2^2\}.$$

It is easy to check that $(\mathbb{F}_2^2, \mathbb{F}_2^5, \mathcal{A}, E_f)$ is a weak AMD code with parameters $(32, 4, 16, \frac{1}{2})$.

Recalling the well-known Dillon's class of functions with perfect nonlinearity in Lemma 4.3.13, we have the following corollary. Note that the trace function $\mathrm{Tr}_q^{q^r} : \mathbb{F}_{q^r} \to \mathbb{F}_q$ defined by $\mathrm{Tr}_q^{q^r}(x) = \sum_{0 \leq i \leq r-1} x^{q^i}$ is a balanced function.

Corollary 5.2.5. *Let $A_1 = A_2 = \mathbb{F}_{q^r}$ and $B = \mathbb{F}_q$, where we regard an element of $\mathbb{F}_{q^r}$ as a vector in $\mathbb{F}_q^r$. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ and $\{\beta_1, \beta_2, \ldots, \beta_r\}$ be a pair of dual bases of $\mathbb{F}_{q^r}$ over $\mathbb{F}_q$, that is,*

$$\mathrm{Tr}_q^{q^r}(\alpha_i \beta_j) = \begin{cases} 1, & i = j, \\ 0, & \textit{otherwise}. \end{cases} \tag{5.8}$$

Define $f : (\mathbb{F}_{q^{2r}}, +) \to (\mathbb{F}_q, +)$ as $f(x, y) = \mathrm{Tr}_q^{q^r}(\hat{x}^{q^r - 2}\hat{y})$, where $x = (x_1, x_2, \ldots, x_r) \in \mathbb{F}_{q^r}$, $y = (y_1, y_2, \ldots, y_r) \in \mathbb{F}_{q^r}$, $\hat{x} = \sum_{1 \leq i \leq r} x_i \alpha_i$, and $\hat{y} = \sum_{1 \leq i \leq r} y_i \beta_i$. Define the probabilistic encoding map E_f from A_1 to $G = A_1 \times A_2 \times B$ as

$$E_f(S_1) = (S_1, S_2, f(S_1, S_2)),$$

where $S_1 = (x_1, x_2, \ldots, x_r) \in A_1$ and $S_2 = (y_1, y_2, \ldots, y_r) \in_R A_2$. Then, for the equiprobable encoding case, the systematic strong AMD code given by E_f has parameters $(q^r, q^{2r+1}, \frac{1}{q})$, where $\rho = \frac{1}{q}$ is minimum with respect to Theorem 5.2.1. Especially, when $r = 1$, the q-regular AMD code is optimal with respect to the bound in Lemma 5.1.4.

Proof. To prove $\rho = \frac{1}{q}$, according to Theorem 5.2.1, it suffices to prove that for any $S_1 \in \mathbb{F}_{q^r}$, $a_1 = (a_{11}, \ldots, a_{1r}) \in \mathbb{F}_{q^r} \setminus \{(0, 0, \ldots, 0)\}$, $a_2 = (a_{21}, \ldots, a_{2r}) \in \mathbb{F}_{q^r}$, and $b \in \mathbb{F}_q$,

$$\frac{|\{S_2 \in \mathbb{F}_{q^r} : \ f(S_1 + a_1, S_2 + a_2) = f(S_1, S_2) + b\}|}{q^r} \leq \frac{1}{q},$$

i.e., $f(S_1+a_1,S_2+a_2)=f(S_1,S_2)+b$ has at most q^{r-1} solutions for $S_2\in\mathbb{F}_{q^r}$. Let

$$\hat{S_1}^{q^r-2}=\sum_{1\le i\le r}x'_{1i}\alpha_i, \tag{5.9}$$

$$(\hat{S_1}+\sum_{1\le i\le r}a_{1i}\alpha_i)^{q^r-2}=\sum_{1\le i\le r}x^*_{1i}\alpha_i, \tag{5.10}$$

and

$$a'=(a'_{11}=x^*_{11}-x'_{11},\ldots,a'_{1r}=x^*_{1r}-x'_{1r}). \tag{5.11}$$

Since $a_1\neq(0,0,\ldots,0)$ and x^{q^r-2} is a non-identity permutation of $\mathbb{F}_{q^r}$, we have $a'\neq(0,0,\ldots,0)$. Without loss of generality, we may assume $a'_{11}\neq 0$. By (5.8)-(5.11),

$$\begin{aligned}&f(S_1+a_1,S_2+a_2)-f(S_1,S_2)-b\\ &=y_1\left(\mathrm{Tr}_q^{q^r}\left(\beta_1\left(\hat{S_1}+\sum_{1\le i\le r}a_{1i}\alpha_i\right)^{q^r-2}\right)-\mathrm{Tr}_q^{q^r}\left(\beta_1\hat{S_1}^{q^r-2}\right)\right)\\ &\quad+C(S,a_1,a_2,b)\\ &=a'_{11}y_1+C(S,a_1,a_2,b),\end{aligned}$$

where $S=(x_1,x_2,\ldots,x_r,y_2,\ldots,y_r)\in\mathbb{F}_{q^{2r-1}}$ and $C(S,a_1,a_2,b)$ is a constant determined by S, a_1, a_2, and b. Thus, the fact $a'_{11}\neq 0$ means that

$$f(S_1+a_1,S_2+a_2)-f(S_1,S_2)-b=0$$

has at most q^{r-1} solutions for $S_2\in\mathbb{F}_{q^r}$, which completes the proof. □

5.3 Highly nonlinear functions from systematic AMD codes

By Theorems 4.3.4 and 5.2.1, we can construct systematic AMD codes from known highly nonlinear functions for both weak and strong attack models. In this section, we further analyze the relationship between AMD codes and highly nonlinear functions. Especially, we try to construct highly nonlinear functions from given systematic AMD codes. Note that a strong (m,n,ρ)-AMD code is always a weak (m,n,ρ)-AMD code. Thus, throughout this section, we only consider the functions derived from systematic weak AMD codes.

Let A_1, A_2 and B be Abelian groups. For a given systematic AMD code with probabilistic encoding map $E:A_1\to A_1\times A_2\times B$,

$$E(s_1)=(s_1,s_2,t_{s_1,s_2}),\quad s_1\in A_1,\ s_2\in_R A_2,$$

define a function f_E from $A_1 \times A_2$ to B as

$$f_E(s_1, s_2) = t_{s_1,s_2}. \tag{5.12}$$

Theorem 5.3.1. *Let $E : A_1 \to A_1 \times A_2 \times B$ be the probabilistic encoding map of a systematic weak AMD code with parameters (m, n, ρ), where $m = |A_1|$ and $n = |A_1||A_2||B|$. Then the map $f_E : A_1 \times A_2 \to B$ has nonlinearity*

$$N_{f_E} \leq \max\{\{\rho\} \cup \{N_{f_{E,s'}} :\ s' \in A_1\}\},$$

where $f_{E,s'}(x) \triangleq f_E(s', x)$ is a map from A_2 to B defined by f_E and $s' \in A_1$, and $N_{f_{E,s'}}$ denotes the nonlinearity of $f_{E,s'}$.

Proof. Let $\rho_{(\Delta_1,\Delta_2,\Delta_3)}$ denote the probability of successful tampering $(\Delta_1, \Delta_2, \Delta_3) \in (A_1 \setminus \{0\}) \times A_2 \times B$. Then

$$\begin{aligned}
\rho \geq& \max\left\{\rho_{(\Delta_1,\Delta_2,\Delta_3)} :\ (\Delta_1, \Delta_2, \Delta_3) \in (A_1 \setminus \{0\}) \times A_2 \times B\right\} \\
=& \max_{\Delta_1 \in A_1 \setminus \{0\}} \max_{\Delta_2 \in A_2} \max_{\Delta_3 \in B} \left\{ \sum_{s' \in A_1} \Pr(s_1 = s') \sum_{s^* \in A_2} \Pr(s_2 = s^*) \right. \\
& \qquad \left. \times \Pr(f_E(s' + \Delta_1, s^* + \Delta_2) = f_E(s', s^*) + \Delta_3) \right\} \\
=& \max_{\Delta_1 \in A_1 \setminus \{0\}} \max_{\Delta_2 \in A_2} \max_{\Delta_3 \in B} \left\{ \sum_{s' \in A_1} \frac{1}{|A_1|} \sum_{s^* \in A_2} \frac{1}{|A_2|} \right. \\
& \qquad \left. \times \Pr(f_E(s' + \Delta_1, s^* + \Delta_2) = f_E(s', s^*) + \Delta_3) \right\} \\
=& \max_{\Delta_1 \in A_1 \setminus \{0\}} \max_{\Delta_2 \in A_2} \max_{\Delta_3 \in B} \\
& \left\{ \frac{|\{(s', s^*) \in A_1 \times A_2 : f_E(s' + \Delta_1, s^* + \Delta_2) = f_E(s', s^*) + \Delta_3\}|}{|A_1||A_2|} \right\}.
\end{aligned}$$

Meanwhile, for $\Delta_1 = 0$ and $\Delta_2 \in A_2 \setminus \{0\}$, we define

$$\rho_{(0,\Delta_2,\Delta_3)} \triangleq \frac{|\{(s', s^*) \in A_1 \times A_2 :\ f_E(s', s^* + \Delta_2) = f_E(s', s^*) + \Delta_3\}|}{|A_1||A_2|}.$$

Then

$$\begin{aligned}
& \max_{\Delta_2 \in A_2 \setminus \{0\}} \max_{\Delta_3 \in B} \rho_{(0,\Delta_2,\Delta_3)} \\
=& \max_{\Delta_2 \in A_2 \setminus \{0\}} \max_{\Delta_3 \in B} \sum_{s' \in A_1} \frac{|\{s^* \in A_2 :\ f_{E,s'}(s^* + \Delta_2) = f_{E,s'}(s^*) + \Delta_3\}|}{|A_1||A_2|} \\
\leq& \max\{N_{f_{E,s'}} :\ s' \in A_1\},
\end{aligned}$$

where the last inequality comes from the fact that

$$N_{f_{E,s'}} = \max_{\Delta_2 \in A_2 \setminus \{0\}} \max_{\Delta_3 \in B} \frac{|\{s^* \in A_2 :\ f_{E,s'}(s^* + \Delta_2) = f_{E,s'}(s^*) + \Delta_3\}|}{|A_2|}.$$

Therefore,

$$\begin{aligned}N_{f_E} = &\max_{(\Delta_1,\Delta_2)\in A_1\times A_2\setminus\{(0,0)\}}\max_{\Delta_3\in B}\\ &\frac{|\{(s',s^*)\in A_1\times A_2 : f(s'+\Delta_1, s^*+\Delta_2) = f(s',s^*)+\Delta_3\}|}{|A_1||A_2|}\\ =&\max\{\max\{\rho_{(\Delta_1,\Delta_2,\Delta_3)} : \ \Delta_1\in A_1\setminus\{0\}, \Delta_2\in A_2, \Delta_3\in B\},\\ &\qquad\qquad \max\{\rho_{(0,\Delta_2,\Delta_3)} : \ \Delta_2\in A_2\setminus\{0\}, \Delta_3\in B\}\}\\ \le&\max\{\{\rho\}\cup\{N_{f_{E,s'}} : \ s'\in A_1\}\}.\end{aligned}$$

□

Generally speaking, from a systematic AMD code we can not determine the nonlinearity of the function f_E directly. This is mainly because that in an AMD code, we do not regard the case $\mathrm{Dec}(E(s)+\Delta) = s$ as an adversary's successful tampering, as shown in Theorem 5.3.1. However, for a stronger setting [28,60,94] an adversary succeeds even when producing a new encoding of the original source, that is, the case $\mathrm{Dec}(E(s)+\Delta) \neq \perp$ is regarded as an adversary's successful tampering. In this setting, we directly have the following result. The proof is similar, so we omit it here.

Theorem 5.3.2. *Let $E : A_1 \to A_1\times A_2\times B$ be the probabilistic encoding map of a systematic weak AMD code. If*

$$\Pr(\mathrm{Dec}(E(s)+\Delta)\neq\perp)\le\rho,$$

then the function $f_E : A_1\times A_2 \to B$ defined as (5.12) has nonlinearity $N_{f_E}\le\rho$.

As an application of Theorem 5.3.1, we analyse the functions derived from the systematic q-regular strong AMD codes in [27, Theorem 2].

Corollary 5.3.3. *Let q be a power of a prime p, and $t > 0$ be an integer such that $p \nmid (t+2)$. Let (E_h, Dec) be the systematic strong AMD codes in [27, Theorem 2] with parameters $(q^t, q^{t+2}, \frac{t+1}{q})$, where the probabilistic encoding map $E_h : \mathbb{F}_{q^t} \to \mathbb{F}_{q^t}\times\mathbb{F}_q\times\mathbb{F}_q$ is given by*

$$E_h(S = (s_1, s_2, \ldots, s_t)) = (S, x, h(S,x))$$

with $x \in_R \mathbb{F}_q$ and

$$h(S,x) = x^{t+2} + \sum_{1\le i\le t} s_i x^i. \tag{5.13}$$

Then the function $h(S,x)$ can be viewed as a function from $(\mathbb{F}_{q^{t+1}}, +)$ to $(\mathbb{F}_q, +)$ with nonlinearity $N_h \le \frac{t+1}{q}$, where we regard elements of $\mathbb{F}_{q^{t+1}}$ as vectors in $\mathbb{F}_q^{t+1}$.

Proof. According to Theorem 5.3.1, it suffices to prove that for any given $S_1 \in \mathbb{F}_{q^t}$, $N_{h_{E,S_i}} \leq \frac{t+1}{q}$ holds, where $h_{E,S_1}(x) = h(S_1, x)$ is a function from $\mathbb{F}_q$ to $\mathbb{F}_q$. By (5.13), for any given $S_1 = (s_1, s_2, \ldots, s_t) \in \mathbb{F}_{q^t}$ and $\Delta \in \mathbb{F}_q \setminus \{0\}$, we have

$$\begin{aligned}
&h_{E,S_1}(x+\Delta) - h_{E,S_1}(x)\\
=&(x+\Delta)^{t+2} - x^{t+2} + \sum_{1\leq i\leq t} s_i((x+\Delta)^i - x^i)\\
=&R_{(S_1,\Delta)}(x),
\end{aligned}$$

where $\deg(R_{(S_1,\Delta)}(x)) = t+1$, for the reason that $p \nmid (t+2)$. Thus, for any $S_1 \in \mathbb{F}_{q^t}$,

$$\begin{aligned}
N_{h_{E,S_1}} =& \max_{\Delta\in\mathbb{F}_q\setminus\{0\}} \max_{b\in\mathbb{F}_q} \frac{|\{x \in \mathbb{F}_q : R_{(S_1,\Delta)}(x) = b\}|}{q}\\
\leq&\frac{t+1}{q},
\end{aligned}$$

which completes the proof. □

Remark 5.3.4. *By Corollary 5.3.3, we know that the systematic AMD codes in Theorem 2 of [27] can also be explained by means of highly nonlinear functions.*

5.4 A combinatorial construction of strong AMD codes

In this section, we are going to construct AMD codes via combinatorial methods.

For positive integers m, e and $u > 2$, let p be a prime and $p^m = eu+1$. Let $\mathbb{F}_{p^m}$ be the finite field with p^m elements and α be a primitive element of $\mathbb{F}_{p^m}$. Define

$$V_0 = \{\alpha^{je} \ : \ 0 \leq j \leq u-1\}.$$

It is easy to check that V_0 is a subgroup of $(\mathbb{F}^*_{p^m}, *)$ with order u. Thus, there exists e elements $a_i = \alpha^i$ with $0 \leq i \leq e-1$ such that

$$V_i = a_i V_0 = \{\alpha^{i+je} \ : \ 0 \leq j \leq u-1\}$$

are exactly e cosets of V_0 in $(\mathbb{F}^*_{p^m}, *)$.

Construction 5.4.1. *Let* $\mathcal{V} = \{V_i \ : \ 0 \leq i \leq e-1\}$. *Define* $\mathcal{A} \triangleq \{A_i \ : \ 0 \leq i \leq e-1\}$ *such that*

$$A_i \triangleq \{(i, j, \alpha^{i+je}) \ : \ 0 \leq j \leq u-1\}. \tag{5.14}$$

Let E_v *be the encoding function from* $\mathbb{Z}_e$ *to* $\mathbb{Z}_e \times \mathbb{Z}_u \times \mathbb{F}_{p^m}$

$$E_v(i) \in_R A_i \quad \text{for } 0 \leq i \leq e-1, \tag{5.15}$$

where $E_v(i) \in_R A_i$ *implies that* $E_v(i)$ *is chosen from* A_i *uniformly at random.*

Theorem 5.4.2. *For positive integers m, e, and $u > 2$, let p be a prime and $p^m = eu + 1$. Then, considering the equiprobable encoding case, there exists a G-optimal systematic strong AMD code with parameters $(uep^m = (p^m - 1)p^m, e, \frac{1}{u})$, where the probabilistic encoding map $E_v : \mathbb{Z}_e \to \mathbb{Z}_e \times \mathbb{Z}_u \times \mathbb{F}_{p^m}$ is given by (5.15).*

Proof. By Construction 5.4.1, we only need to prove $\rho \leq \frac{1}{u}$, i.e., for any given $s \in \mathbb{Z}_m$, the inequality

$$\Pr(\mathrm{Dec}(E(s) + \Delta) \notin \{s, \perp\}) \leq \frac{1}{u} \tag{5.16}$$

holds for any $\Delta = (\Delta_1, \Delta_2, \Delta_3)$ with $\Delta_1 \in \mathbb{Z}_e \setminus \{0\}$, $\Delta_2 \in \mathbb{Z}_u$, and $\Delta_3 \in \mathbb{F}_{p^m}$.

By (5.14) and (5.15), we have

$$\begin{aligned} &\Pr(\mathrm{Dec}(E(s) + \Delta) \notin \{s, \perp\}) \\ =&\Pr(\{\alpha^{s+\Delta_1+(t+\Delta_2)e} = \alpha^{s+te} + \Delta_3 \,:\, t \in_R \mathbb{Z}_u\}) \\ =&\frac{|\{t \in \mathbb{Z}_u \,:\, \alpha^{s+\Delta_1+(t+\Delta_2)e} = \alpha^{s+te} + \Delta_3\}|}{u} \end{aligned} \tag{5.17}$$

where the second equality holds by the fact that E is equiprobable encoding. Since $\Delta_1 \neq 0$, we have $A_s \cap A_{s+\Delta_1} = \emptyset$. For any given $\Delta = (\Delta_1, \Delta_2, \Delta_3)$ with $\Delta_1 \in \mathbb{Z}_e \setminus \{0\}$, $\Delta_2 \in \mathbb{Z}_u$, and $\Delta_3 \in \mathbb{F}_{p^m}$, note that $\alpha^{s+\Delta_1+(t+\Delta_2)e} - \alpha^{s+te} = \alpha^{s+te}(\alpha^{\Delta_1+\Delta_2 e} - 1)$ run through the set $\mathbb{F}_{p^m}^*$, when (s, t) run through the set $\mathbb{Z}_e \times \mathbb{Z}_u$. Therefore, we have for any given $\Delta = (\Delta_1, \Delta_2, \Delta_3)$ with $\Delta_1 \in \mathbb{Z}_e \setminus \{0\}$, $\Delta_2 \in \mathbb{Z}_u$, and $\Delta_3 \in \mathbb{F}_{p^m}$, we have

$$|\{t \in \mathbb{Z}_u \,:\, \alpha^{s+\Delta_1+(t+\Delta_2)e} = \alpha^{s+te} + \Delta_3\}| \leq 1.$$

Recall (5.17), we have

$$\Pr(\mathrm{Dec}(E(s) + \Delta) \notin \{s, \perp\}) \leq \frac{1}{u}.$$

This completes the proof. □

Chapter 6

Conclusions and Open Problems

In this chapter, we draw a brief conclusion of new results obtained in this dissertation, and also propose several interesting open problems.

6.1 Conclusions

In this dissertation, we studied algebraic manipulation detection codes for both strong/weak attack model and related structures such as external difference families and highly nonlinear functions. In the following, we briefly list the new results in this dissertation.

Weak algebraic manipulation detection codes

- We defined a new type of weighted external difference families, which are proved equivalent with weak algebraic manipulation detection codes. In this way, the combinatorial characterization of weak algebraic manipulation detection codes was proposed.

- We improved the known lower bound, i.e., the R-bound by Paterson and Stinson [81] on the maximum probability of successful tampering for the adversary's all possible strategies;

- We derived a necessary condition for the R-bound to be achieved;

- We determined the exact combinatorial structure for a weak algebraic manipulation detection code with the minimum possible probability of successful tampering, when the R-bound is not achievable.

- We constructed some new R-optimal weak algebraic manipulation detection codes based on weighted external difference families.

- We constructed some new weak algebraic manipulation detection codes with asymptotically optimal effective tag size.

Strong algebraic manipulation detection codes

- We constructed optimal strong algebraic manipulation detection codes via highly nonlinear functions.
- We constructed G-optimal strong algebraic manipulation detection codes by an explicit combinatorial construction.
- We constructed highly nonlinear functions based on strong algebraic manipulation detection codes.

External difference family

- We proposed three constructions of optimal bounded standard weighted external difference families.
- We proposed a construction of optimal cyclic standard weighted external difference families.
- We constructed optimal partitioned difference families with new parameters.
- We constructed optimal difference system of sets with positive rate based on partitioned difference families.

6.2 Open problems

In this section, we list several interesting open problems related with the topics in this dissertation.

- How to construct strong algebraic manipulation detection codes with flexible parameters via highly nonlinear functions?
- We proved that the R-bound for weak algebraic manipulation detection codes is not always tight. How about the R-bound for strong algebraic manipulation detection codes? If it is also not tight in some special cases, how to improve the R-bound for strong algebraic manipulation detection codes?
- How to explicitly construct standard weighted external difference families for the case $m \geq 3$, i.e., R-optimal weak algebraic manipulation detection codes?
- The general relationship between error correcting codes and algebraic manipulation detection codes is known. However, how to find explicit construction that yields better parameters than the known ones?
- How to explicitly construct bounded standard weighted external difference families?
- How to construct strong external difference families for the case $m \geq 5$?

Bibliography

[1] H. Ahmadi and R. Safavi-Naini, "Detection of algebraic manipulation in the presence of leakage," *ICITS 2013,* LNCS, vol. 8317, pp. 238-258, 2013.

[2] K. D. Akdemir, Z. Wang, M. Karpovsky, and B. Sunar, "Design of cryptographic devices resilient to fault injection attacks using nonlinear robust codes," *Fault analysis in cryptography,* pp. 171-199, Springer, Berlin, Heidelberg, 2012.

[3] T. M. Apostol, *Introduction to Analytic Number Theory,* Springer-Verlag, New York, 1976.

[4] J. Bao, L. Ji, R. Wei, and Y. Zhang, "New existence and nonexistence results for strong external difference families," *Discr. Math.,* vol. 341, no. 6, pp. 1798-1805, 2018.

[5] E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," *J. Cryptology,* vol. 4, no. 1, pp. 3-72, 1991.

[6] G. Brassard, A. Broadbent, J. Fitzsimons, S. Gambs, and A. Tapp, "Anonymous quantum communication," *Advances in Cryptology-Asiacrypt2007*, LNCS, vol. 4833, pp. 460-473, 2007.

[7] A. Broadbent and A. Tapp, "Information-theoretic security without an honest majority.," "Anonymous quantum communication," *Advances in Cryptology-Asiacrypt2007*, LNCS, vol. 4833, pp. 410-426, 2007.

[8] M. Buratti and D. Ghinelli, "On disjoint $(3t, 3, 1)$ cyclic difference families," *J. Statist. Plann. Inference,* vol. 140, no. 7, pp. 1918-1922, 2010.

[9] M. Buratti, J. Yan, and C. Wang, "From a 1-rotational RBIBD to a partitioned difference family," *Electron. J. Comb.,* vol. 17, R139, 2010.

[10] M. Buratti, "Hadamard partitioned difference families and their descendants," *Cryptogr. Commun.,* to appear.

[11] M. Buratti, "On disjoint $(v, k, k-1)$ difference families," *Des. Codes Cryptogr.,* to appear.

[12] S. Cabello, C. Padró, and G. Sáez, "Secret sharing schemes with detection of cheaters for a general access structure," *Des. Codes Cryptogr.,* vol. 25, no. 2, pp. 175-188, 2002.

[13] H. Cai, X. Zeng, T. Helleseth, X. Tang, and Y. Yang, "A new construction of zero-difference balanced functions and its applications," *IEEE Trans. Inf. Theory,* vol. 59, no. 8, pp. 5008-5015, 2013.

[14] H. Cai, Z. Zhou, X. Tang, and Y. Miao, "Zero-difference balanced functions with new parameters and their applications," *IEEE Trans. Inf. Theory,* vol. 63, no. 7, pp. 4379-4387, 2017.

[15] A. Canteaut, P. Charpin, and H. Dobbertin, "Weight divisibility of cyclic codes, highly nonlinear functions on $\mathbb{F}_{2^m}$, and cross correlation of maximum-length sequences," *SIAM J. Discr. Math.* vol 13, no. 1, pp. 105-138, 2000.

[16] C. Carlet and C. Ding, "Highly nonlinear mappings," *Journal of Complexity,* vol. 20, no. 2-3, pp. 205-244, 2004.

[17] C. Carlet and C. Ding, "Authentication schemes from highly nonlinear functions," *Des. Codes Cryptogr,* vol. 40, no. 1, pp. 71-79, 2006.

[18] C. Carlet and C. Ding, "Nonlinearities of S-boxes," *Finite Fields and Their Applications,* vol. 13, no. 1, pp. 121-135, 2007.

[19] C. Carlet, G. Gong, and Y. Tan, "Quadratic zero-difference balanced functions, APN functions and strongly regular graphs," *Des. Codes Cryptogr.*, vol. 78, no. 3, pp. 629-654, 2016.

[20] C. Carlet, A. B. Levina, and S. V. Taranov, "Algebraic manipulation detection codes with perfect nonlinear functions under non-uniform distribution," *Scientific and Technical Journal of Information Technologies, Mechanics and Optics,* vol. 17, no. 6, pp. 1052-1062, 2017.

[21] Y. Chang and C. Ding, "Constructions of external difference families and disjoint difference families," *Des. Codes Cryptogr.,* vol. 40, no. 2, pp. 167-185, 1997.

[22] Y. M. Chee, A. C. H. Ling, and J. Yin, "Optimal partitioned cyclic difference packings for frequency hopping and code synchronization," *IEEE Trans. Inf. Theory,* vol. 56, no. 11, pp. 5738-5746, 2010.

[23] Y. Chen, Y. Tan, and G. Gong, "New bounds and constructions of weak systematic algebraic modification detection codes," in *Sequences and Their Applications(SETA)*, 2016.

[24] W. Chu and C. J. Colbourn, "Optimal frequency-hopping sequences via cyclotomy," *IEEE Trans. Inf. Theory,* vol. 51, no. 3, pp. 1139-1141, Mar. 2005.

[25] F. R. K. Chung, J.A. Salehi, and V.K. Wei, "Optical orthogonal codes: design, analysis, and applications," *IEEE Trans. Inf. Theory,* vol. 35, no. 3, pp. 595-604, 1989.

[26] C. J. Colbourn and J. H. Dinitz, *Handbook of Combinatorial Designs*, Chapman & Hall/CRC, vol. 42, 2006.

[27] R. Cramer, Y. Dodis, S. Fehr, C. Padró, and D. Wichs, "Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors," *Eurocrypt 2008*, LNCS, vol 4965, pp. 471-488, 2008.

[28] R. Cramer, S. Fehr, and C. Padró, "Algebraic manipulation delection codes," *Science China Mathematics*, vol. 56, no. 7, pp. 1349-1358, 2013.

[29] R. Cramer. C. Padró, and C. Xing, "Optimal algebraic manipulation detection codes in the constant-error model," *TCC2015*, LNCS, vol. 9014, pp. 481-501, 2015.

[30] J. F. Dillon, "Elementary Hadamard Difference Sets," Ph.D. Thesis, University of Maryland, 1974.

[31] C. Ding, D. Pei, and A. Salomaa, *Chinese Remainder Theorem: Applications in Computing, Coding, Cryptography,* World Scientific Publishing, 1996.

[32] C. Ding, T. Helleseth, H. M. Martinsen, "New families of binary sequences with optimal three-level autocorrelation," *IEEE Trans. Inf. Theory,* vol. 47, no. 1, pp. 428-433, 2001.

[33] C. Ding and H. Niederreiter, "Systematic authentication codes from highly nonlinear functions," *IEEE Trans. Inf. Theory,* vol. 50, no. 10, pp. 2421-2428, 2004.

[34] C. Ding and J. Yin, "Combinatorial constructions of optimal constant-composition codes," *IEEE Trans. Inf. Theory,* vol. 51, no. 10, pp. 3671-3674, 2005.

[35] C. Ding and J. Yin, "A construction of optimal constant composition codes," *Des. Codes Cryptgr.,* vol. 40, no. 2, pp. 157-165, 2006.

[36] C. Ding, "Optimal constant composition codes from zero-difference balanced functions," *IEEE Trans. Inf. Theory,* vol. 54, no. 12, pp. 5766-5770, 2008.

[37] C. Ding, "Optimal and perfect difference systems of sets," *Journal of Combinatorial Theory, Series A,* vol. 116, no. 1, pp. 109-119, 2009.

[38] C. Ding and Y. Tan, "Zero-difference balanced functions with applications," *Journal of Statistical Theory and Practice*, vol. 6, no. 1, pp. 3-19, 2012.

[39] C. Ding, Q. Wang, and M. Xiong, "Three new families of zero-difference balanced functions with applications," *IEEE Trans. Inf. Theory,* vol. 60, no. 4, pp. 2407-2413, 2014.

[40] J. H. Dinitz and P. Rodney, "Disjoint difference families with block size 3," *Util. Math.,* vol. 52, pp. 153-160, 1997.

[41] J. H. Dinitz and N. Shalaby, "Block disjoint difference families for Steiner triple systems: $v \equiv 3 \bmod 6$," *J. Statist. Plann. Inference,* vol. 106, vol. 1-2, pp. 77-86, 2002.

[42] Y. Dodis, J. Katz, L. Reyzin, and A. Smith, "Robust fuzzy extractors and authenticated key agreement from close secrets," *Advance in Cryptology-Crypto'06*, LNCS, vol. 4117, pp. 232-250, 2006.

[43] S. Dziembowski, K. Pietrzak, and D. Wichs, "Non-malleable codes," *Journal of the ACM*, vol. 65, no. 4, pp. 1-32, 2018.

[44] C. Fan, J. Lei and Y. Chang, "Constructions of difference systems of sets and disjoint difference families," *IEEE Trans. Inf. Theory,* vol. 54, no. 7, pp. 3195-3201, 2008.

[45] Y. Fujiwara and V. D. Tonchev, "High-rate self-synchronizing codes," *IEEE Trans. Inf. Theory,* vol. 59, no. 4, pp. 2328-2335, 2013.

[46] R. Fuji-Hara, Y. Miao, and M. Mishima, "Optimal frequency hopping sequences: a combinatorial approach," *IEEE Trans. Inf. Theory,* vol. 50, no. 10, pp. 2408-2420, 2004.

[47] S. Furino, "Difference families from rings," *Discr. Math.,* vol. 97, no. 1-3, pp. 177-190, 1991.

[48] G. Ge, R. Fuji-Hara, and Y. Miao, "Further combinatorial constructions for optimal frequency-hopping sequences," *J. Comb. Theory, Ser. A,* vol. 113, no. 8, pp. 1699-1718, 2006.

[49] G. Ge, Y. Miao, and Z. Yao, "Optimal frequency hopping sequences: auto- and cross-correlation properties," *IEEE Trans. Inf. Theory,* vol. 55, no. 2, pp. 867-879, 2009.

[50] S. Ge, Z. Wang, M. Karpovsky, and P. Luo, "Reliable and secure memories based on algebraic manipulation detection codes and robust error correction," in *Proc. Int. Depend Symp.*, 2013.

[51] S. W. Golomb, B. Gordon, and L. R. Welch, "Comma-free codes," *Canad. J. Math.,* vol. 10, no. 2, pp. 202-209, 1958.

[52] V. Guruswami and A. Smith, "Codes for Computationally Simple Channel-s: Explicit Constructions with Optimal Rate," in: Foundation of Computer Science (FOCS), 51th Annual IEEE symposium on pp. 723-732, 2010.

[53] X. Hou, "q-Ary bent functions constructed from chain rings," *Finite Fields and Their Applications*, vol. 4, no. 1, pp. 55-61 1998.

[54] S. Huczynska and M. B. Paterson, "Existence and non-existence results for strong external difference families," *Discr. Math.,* vol. 341, no. 1, pp. 87-95, 2018.

[55] S. Huczynska and M. B. Paterson, "Weighted external difference families and R-optimal AMD codes," *Discr. Math.,* vol 342, no. 3, pp. 855-867, 2019.

[56] S. Huczynska and M. B. Paterson, "Characterising bimodal collections of sets in finite groups," *arXiv:* 1903.11620, 2019.

[57] J. Jedwab and S. Li, "Construction and nonexistence of strong external difference families," *Journal of Algebraic Combinatorics,* vol. 49, no .1, pp. 21-48, 2019.

[58] W. Jia, X. Zeng, T. Helleseth, and C. Li, "A class of binomial bent functions over the finite fields of odd characteristic," *IEEE Trans. Inf. Theory,* vol. 58, no. 9, pp. 6054-6063, 2012.

[59] M. Jimbo and S. Kuriki, "On a composition of cyclic 2-designs," *Discr. Math.,* vol. 46, no. 3, pp. 249-255, 1983.

[60] M. Karpovsky and Z. Wang. "Design of strongly secure communication and computation channels by nonlinear error detecting codes," *IEEE Transactions on Computers,* vol. 63, no. 11, pp. 2716-2728, 2013.

[61] V. I. Levenshtein, "One method of constructing quasilinear codes providing synchronization in the presence of errors," *Prob. Inf. Transmiss.,* vol. 7, no. 3, pp. 215-222, 1971.

[62] V. I. Levenshtein, "Combinatorial problems motivated by comma-free codes," *J. Combin. Designs,* vol. 12, no. 3, pp. 184-196, 2004.

[63] A. Levina and S. Taranov, "New construction of algebraic manipulation detection codes based on wavelet transform," Proceeding of the 18th conference of FRUCT Association, pp. 187-192, 2016.

[64] N. Li, X. Tang, and T. Helleseth, "New constructions of quadratic bent functions in polynomial form," *IEEE Trans. Inf. Theory,* vol. 60, no. 9, pp. 5760-5767, 2014.

[65] S. Li, H. Wei, and G. Ge, "Generic constructions for partitioned difference families with applications: a unified combinatorial approach," *Des. Codes Cryptogr.*, vol. 82, no. 3, pp. 583-599, 2017.

[66] X. Lu, X. Niu, and H. Cao, "Some results on generalized strong external difference families," *Des. Codes Cryptogr,* vol. 86, no. 12, pp. 2857-2868, 2018.

[67] P. Luo, Z. Wang, and M. Karpovsky, "Secure NAND flash architecture resilient to strong fault-injection attacks using algebraic manipulation detection code," in Proceedings of the International Conference on Security and Management (SAM), 2013.

[68] W. J. Martin and D. R. Stinson, "Some nonexistence results for strong external difference families using character theory," *Bull. Inst. Combin. Appl.,* vol. 80, pp. 79-92, 2017.

[69] M. Matsui, "Linear cryptanalysis method for DES cipher," *Advances in Cryptology – Eurocrypt'93*, LNCS, vol. 765. Berlin: Springer, pp. 386-397, 1994.

[70] S. Mesnager, *Bent Functions: Fundamentals and Results,* Springer, 2016.

[71] S. Mesnager, Z. Zhou, and C. Ding, "On the nonlinearity of Boolean functions with restricted input," *Cryptography and Communications,* vol. 11, no. 1, pp. 63-76, 2019.

[72] K. Momihara, "Disjoint difference families from Galois rings," *Electron. J. Combin.,* vol. 24, no. 3, P3.23, 2017.

[73] S.-L. Ng and M. B. Paterson, "Disjoint difference families and their applications", *Des. Codes Cryptogr.,* vol. 78, no. 1, pp. 103-127, 2016.

[74] K. Nyberg, "Perfect nonlinear S-boxes," in *Advance in Cryptology-Eurocrypt'91,* (Brighton, 1991), vol. 547, LNCS, pp. 378-386, Berlin: Springer, 1991.

[75] S. Obana and T. Araki, "Almost optimum secret sharing schemes secure against cheating for arbitrary secret distribution," *Advances in Cryptology-Asiacrypt2006*, LNCS, vol. 4284, pp. 364-379, 2006.

[76] W. Ogata and K. Kurosawa, "Optimum secret sharing scheme secure against cheating," *Advance in Cryptology-Eurocrypt'96*, LNCS, vol. 1070, pp. 200-211, 1996.

[77] W. Ogata, K. Kurosawa, D. R. Stinson, and H. Saido, "New combinatorial designs and their applications to authentication codes and secret sharing schemes," *Discr. Math.*, vol. 279, pp. 383-405, 2004.

[78] J. D. Olsen, R. A. Scholtz, L. R. Welch, "Bent function sequences," *IEEE Trans. Inf. Theory,* vol. 28, no. 6, pp. 858-864, 1982.

[79] R. Omrani and S. V. Maric, "A new construction of multipleet sonar and extended Costas arrays with perfect correlation," 2006 40th Annual Conference on Information Sciences and Systems, New Jersey, USA, Mar. 22-26, 2006, pp. 512-517.

[80] C. Padró, G. Sáez, and J. L. Villar, "Detection of cheaters in vector space secret sharing schemes," *Des. Codes Cryptogr.*, vol. 16, no. 1, pp.75-85, 1999.

[81] M. B. Paterson and D. R. Stinson, "Combinatorial characterizations of algebraic manipulation detection codes involving generalized difference families," *Discr. Math.,* vol. 339, no. 12, pp. 2891-2906, 2016.

[82] K. T. Phelps, "Isomorphism problems for cyclic block designs," *Annals of Discrete Mathematic*, vol. 37, pp. 385-392, 1987.

[83] O. S. Rothaus, "On bent functions," *Journal of Combinatorial Theory, Series A,* vol. 20, no. 3, pp. 300 - 305, 1976.

[84] J. A. Salehi, "Code division multiple-access techniques in optical fiber networks-part I: fundamental principles," *IEEE Trans. Commun.,* vol. 37, no. 8, pp. 824-833, 1989.

[85] J. A. Salehi and C. A. Brackett, "Code division multiple-access techniques in optical fiber networks-part II: systems performance analysis," *IEEE Trans. Commun.,* vol. 37, no. 8, pp. 834-850, 1989.

[86] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612-613, 1979.

[87] G.J. Simmons, "Authentication theory/coding theory," in: *Advances in Cryptology— Crypto'84*, LNCS, vol. 196, pp. 411-431, 1984.

[88] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook.* New York: McGraw-Hill, 2002.

[89] M. Tompa and H. Woll, "How to share a secret with cheaters," *J. Cryptology*, vol. 1, no. 3, pp. 133-138, 1989.

[90] H. Wang, "A new bound for difference systems of sets," *J. Combin. Math. Combin. Comput.,* vol. 58, pp. 161-167, 2006.

[91] Q. Wang and Y. Zhou, "Sets of zero-difference balanced functions and their applications," *Adv. Math. Commun.,* vol. 8, no. 1, pp. 83-101, 2014.

[92] X. Wang and J. Wang, "Partitioned difference families and almost difference sets,"*J. Stat. Plan. Inference,* vol. 141, no. 5, pp. 1899-1909, 2011.

[93] Z. Wang and M. Karpovsky "Algebraic manipulation detection codes and their applications for design of secure cryptographic devices, *IEEE 17th International On-Line Testing Symposium*, pp. 234-239, 2011.

[94] H. Wee, "Public key encryption against related key attacks," *International Workshop on Public Key Cryptography,* Springer, Berlin, Heidelberg, 2012.

[95] J. Wen, M. Yang, and K. Feng, "The (n, m, k, λ)-strong external difference family with $m \geq 5$ exists," *arXiv:* 1612.09495v1, 2016.

[96] J. Wen, M. Yang, F. Fu, and K. Feng, "Cyclotomic construction of strong external difference families in finite fields," *Des. Codes Cryptogr,* vol. 86, no. 5, pp. 1149-1159, 2018.

[97] R. M. Wilson, "Cyclotomy and difference families in elementary Abelian groups," *J. Number Theory,* vol. 4, pp. 17-47, 1972.

[98] J. Yin, "Some combinatorial constructions for optical orthogonal codes," *Discr. Math.,* vol. 185, pp. 201-219, 1998.

[99] J. Yin, X. Shan, and Z. Tian, "Constructions of partitioned difference families," *Eur. J. Comb.,* vol. 29, no. 6, pp. 1507-1519, 2008.

[100] Z. Zha and L. Hu, "Constructions of zero-difference balanced functions with applications," *IEEE Trans. Inf. Theory,* vol. 61, no. 3, pp. 1491-1495, 2015.

[101] Z. Zhou, X. Tang, D. Wu, and Y. Yang, "Some new classes of zero-difference balanced functions," *IEEE Trans. Inf. Theory,* vol. 58, no. 1, pp. 139-145, 2012.

List of Publications

- **M. Shao** and Y. Miao, "On optimal weak algebraic manipulation detection codes and weighted external difference families," *Designs, Codes and Cryptography*, vol. 88, no. 7, pp. 1349-1369, 2020.
- **M. Shao** and Y. Miao, "Algebraic manipulation detection codes via highly nonlinear functions," submitted to *Cryptography and Communications*, *preprint arXiv:2002.03724*, (2020).