Combinatorial Characterizations of Algebraic Manipulation Detection Codes Involving Generalized Difference Families

Maura B. Paterson[1] and Douglas R. Stinson[*2]

[1]Department of Economics, Mathematics and Statistics, Birkbeck, University of London,
Malet Street, London WC1E 7HX, UK
[2]David R. Cheriton School of Computer Science, University of Waterloo, Waterloo,
Ontario, N2L 3G1, Canada

March 8, 2016

Abstract

This paper provides a mathematical analysis of optimal algebraic manipulation detection (AMD) codes. We prove several lower bounds on the success probability of an adversary and we then give some combinatorial characterizations of AMD codes that meet the bounds with equality. These characterizations involve various types of generalized difference families. Constructing these difference families is an interesting problem in its own right.

1 Introduction

Algebraic manipulation detection (AMD) codes were defined in 2008 by Cramer *et al.* [3, 4] as a generalization and abstraction of techniques that were previously used in the study of robust secret sharing schemes [14, 15, 17]. AMD codes are studied further in [1, 5, 6]. Several interesting and useful applications of these structures are described in these papers, including applications to robust fuzzy extractors, secure multiparty computation, non-malleable codes, etc. Various construction methods for AMD codes are also presented in these papers.

We begin by providing some motivating examples as well as some historical context from the point of view of authentication codes. AMD codes can be considered as a variation of the classical *unconditionally secure authentication codes* [16], which we will refer to as *A-codes* for short. An A-code has the form $(\mathcal{S}, \mathcal{T}, \mathcal{K}, \mathcal{E})$ where $\mathcal{S}$ is a set of plaintext *sources*, $\mathcal{T}$ is a set of *tags*, $\mathcal{K}$ is a set of *keys* and $\mathcal{E}$ is a set of *encoding functions*. For each $K \in \mathcal{K}$, there is a (possibly randomized) encoding function $E_K : \mathcal{S} \rightarrow \mathcal{T}$. A secret key $K \in \mathcal{K}$ is chosen randomly. Later a source $s \in \mathcal{S}$ is selected and the tag $t = E_K(s)$ is computed. The tag t is authenticated by verifying that $t = E_K(s)$; this can be done only with knowledge of the key K. Having seen a valid pair (s, t), an active adversary may create a bogus pair (s', t') (where $s' \neq s$), hoping that it will be accepted as authentic (this process is called *substitution*). The adversary is trying to maximize the *success probability* of such an attack. One main objective is to design A-codes that will minimize the success probability of the adversary.

[*]D. Stinson's research is supported by NSERC discovery grant 203114-11.

Example 1.1. *Let p be prime and define $\mathcal{S} = \mathcal{T} = \mathbb{Z}_p$. Define $\mathcal{K} = \mathbb{Z}_p \times \mathbb{Z}_p$. For every $K = (c, d) \in \mathcal{K}$, define the function E_K by the rule $s \mapsto cs + d \bmod p$ for all $s \in \mathbb{Z}_p$. (That is, the encoding functions consist of all linear functions from $\mathbb{Z}_p$ to $\mathbb{Z}_p$.) Any observed source-tag pair (s, t) is valid under exactly p of the p^2 keys. Then, any substitution (s', t') $(s' \neq s)$ is valid under exactly 1 of the p "possible" keys. Therefore, the adversary's success probability is $1/p$.*

There are two types of AMD codes. The first type is a *weak AMD code*. Here there is no key, so there is only one encoding function E. Further, the tag is an element of a finite additive abelian group, say $\mathcal{G}$. The adversary is required to commit to a specific substitution of the form $g \mapsto g + \Delta$, where $\Delta \in \mathcal{G} \setminus \{0\}$ is fixed. Later, a source $s \in S$ is chosen randomly and encoded to $g = E(s)$. Then g is replaced by $g' = g + \Delta$. The adversary wins if $g' = E(s')$ for some $s \neq s'$. Again, the objective in designing such a code is to minimize the adversary's success probability.

Example 1.2. *Let $\mathcal{S} = \{1, 2, 3, 4, 5\}$ and let $\mathcal{G} = \mathbb{Z}_{21}$. The encoding function E is defined by $E(1) = 3$, $E(2) = 6$, $E(3) = 12$, $E(4) = 7$ and $E(5) = 14$. It turns out that the adversary's success probability is $1/5$, independent of his choice of $\Delta \neq 0$. This follows because $\{3, 6, 12, 7, 14\}$ is a difference set in $\mathbb{Z}_{21}$ (for the definition of difference set, see Section 2).*

The second type of AMD code is a *strong AMD code*. It is basically the same as a weak AMD code, except that the adversary is given the source (but not the encoded version of the source) before choosing Δ.

Example 1.3. *This example is based on Example 2.7. Let $\mathcal{S} = \{1, 2, 3, 4\}$ and let $\mathcal{G} = \mathbb{Z}_7$. The encoding function E is defined by $E(1) = 1$, $E(2) = 2$, $E(3) = 4$ and $E(4) \in_R \{0, 3, 5, 6\}$ (the notation "$\in_R$" denotes that the given encoding is to be chosen uniformly at random from the given set). If the source $s = 1, 2$ or 3, then the adversary succeeds with probability 1 by choosing Δ such that $E(s) + \Delta = E(s')$ for some $s' \neq s$. However, if the source $s = 4$, it can be verified that the adversary's success probability is $1/2$. To see this, observe for any $\Delta \neq 0$ that $E(4) + \Delta \in \{E(1), E(2), E(3)\}$ for precisely two of the four possible values of $E(4)$.*

1.1 Notation

In this section, we present relevant notation that we will use in the rest of the paper.

- There is a set $\mathcal{S}$ of plaintext messages which is termed the *source space*, where $|\mathcal{S}| = m$. There will be a probability distribution on $\mathcal{S}$, which is assumed to be public. We will normally assume $\mathbf{Pr}[s] = 1/m$ for all $s \in \mathcal{S}$, so we have *equiprobable sources*.

- The *encoded message space* (or more simply, *message space*) is a set $\mathcal{G}$, where $|\mathcal{G}| = n$ (note: $\mathcal{G}$ will usually be an additive abelian group with identity 0).

- For every source $s \in \mathcal{S}$, let $A(s) \subseteq \mathcal{G}$ denote the set of *valid* encodings of s. We require that $A(s) \cap A(s') = \emptyset$ if $s \neq s'$; this ensures that any message can be correctly decoded. Denote $\mathcal{A} = \{A(s) : s \in \mathcal{S}\}$.

- Let $a_s = |A(s)|$ for any $s \in \mathcal{S}$. Define

$$\mathcal{G}_0 = \bigcup_{s \in \mathcal{S}} A(s)$$

and denote

$$a = \sum_{s \in \mathcal{S}} a_s.$$

If a_s is constant, say k, then the code is *k-uniform*. In this case, $a = km$.

- $E : \mathcal{S} \rightarrow G$ is a (possibly randomized) *encoding function* that maps a source $s \in \mathcal{S}$ to some $g \in A(s)$ according to a certain probability distribution defined on $A(s)$:

  $$\mathbf{Pr}[E(s) = g] = \mathbf{Pr}[g \mid s].$$

  The encoding function E, as well as the probability distributions $\mathbf{Pr}[E(s) = g]$, are assumed to be public. Observe that, for equiprobable sources, the induced probability distribution on $\mathcal{G}_0$ is given by

  $$\mathbf{Pr}[g] = \frac{1}{m} \times \mathbf{Pr}[E(s) = g]$$

  for all $s \in \mathcal{S}$ and all $g \in A(s)$.

- Formally, we can define the AMD code as a 4-tuple $(\mathcal{S}, \mathcal{G}, \mathcal{A}, E)$.

- If $\mathbf{Pr}[E(s) = g] = 1/a_s$ for every $s \in \mathcal{S}$ and every $g \in A(s)$, then the code has *equiprobable encoding*. Such a code can be denoted as a 3-tuple $(\mathcal{S}, \mathcal{G}, \mathcal{A})$. In a code with equiprobable sources and equiprobable encoding, we have

  $$\mathbf{Pr}[g] = \frac{1}{a_s m}$$

  for all $s \in \mathcal{S}$ and all $g \in A(s)$.

- A k-uniform code that has equiprobable sources and equiprobable encoding is said to be *k-regular*. In a k-regular code, we have

  $$\mathbf{Pr}[g] = \frac{1}{km}$$

  for all $g \in \mathcal{G}_0$.

- A 1-regular code is said to be *deterministic* because the source uniquely determines the encoding. In a deterministic code with equiprobable sources, we have

  $$\mathbf{Pr}[g] = \frac{1}{m}$$

  for all $g \in \mathcal{G}_0$.

1.2 Formal Definitions of Weak and Strong AMD Codes

We formally define the notion of *weak security* for an AMD code $(\mathcal{S}, \mathcal{G}, \mathcal{A}, E)$ by considering a certain game incorporating an adversary. The adversary has complete information about the AMD code that is being used. Based on this information, the adversary will adopt a *strategy* σ which he will use to choose a value Δ in the game described below. A strategy is allowed to be randomized.

Definition 1.1 (Weak AMD code)**.**

Suppose $(\mathcal{S}, \mathcal{G}, \mathcal{A}, E)$ is an AMD code.

1. *The value $\Delta \in \mathcal{G} \setminus \{0\}$ is chosen according to the adversary's strategy.*
2. *The source $s \in \mathcal{S}$ is chosen uniformly at random by the encoder (i.e., we have equiprobable sources).*
3. *The source is encoded into $g \in A(s)$ using the encoding function E.*
4. *The adversary wins if and only if $g + \Delta \in A(s')$ for some $s' \neq s$.*

The success probability *of the strategy σ, denoted ϵ_σ, is the probability that the adversary wins this game using the specific strategy σ.*

We will say that the code $(\mathcal{S}, \mathcal{G}, \mathcal{A}, E)$ is a ***weak*** *$(m, n, \hat{\epsilon})$**-AMD code** where $\hat{\epsilon}$ denotes the success probability of the adversary's optimal strategy. That is,*

$$\hat{\epsilon} = \max_{\sigma}\{\epsilon_\sigma\}.$$

We now turn to the stronger security model. The following concept of *strong security* is also defined as a game involving an adversary. In this model, the *strategy* σ used to choose Δ will depend on the source s.

Definition 1.2 (Strong AMD code)**.**

1. *The source $s \in \mathcal{S}$ is given to the adversary (here there is no probability distribution defined on $\mathcal{S}$).*
2. *The value $\Delta \in \mathcal{G} \setminus \{0\}$ is chosen according to the adversary's strategy.*
3. *The source is encoded into $g \in A(s)$ using the encoding function E.*
4. *The adversary wins if and only if $g + \Delta \in A(s')$ for some $s' \neq s$.*

For a given source s the success probability *of the strategy σ, denoted $\epsilon_{\sigma,s}$, is the probability that the adversary wins this game using the specific strategy σ.*

We will say that the code $(\mathcal{S}, \mathcal{G}, \mathcal{A}, E)$ is a ***strong*** *$(m, n, \hat{\epsilon})$**-AMD code** where $\hat{\epsilon}$ denotes the maximum success probability of any strategy over all sources s. That is,*

$$\hat{\epsilon} = \max_{\sigma,s}\{\epsilon_{\sigma,s}\}.$$

As we mentioned earlier, the difference between a weak and strong AMD code is that, in a weak code, the adversary chooses Δ before he sees s, while in a strong code, the adversary is given s and then he chooses Δ.

Remark: We have defined AMD codes using the idea of a "game" involving an adversary. We think that this approach is convenient for the presentation of the proofs of various bounds, as well as to compare and contrast AMD codes with "classical" authentication codes. We also note that some previous papers on AMD codes, such as [5, 6], also make mention of an adversary, much as we have done, but it is not part of the formal definition of an AMD code. The definitions we are using are easily seen to be equivalent to the definitions used in other papers such as [3, 5].

1.3 Our Contributions

In this paper, we study *optimal* AMD codes, i.e., codes in which the adversary's success probability is as small as possible. We consider bounds for both weak and strong AMD codes and investigate when these bounds can be achieved. This involves several generalizations of difference families, some of which have apparently not been studied previously.

Connections between difference sets and difference families on the one hand, and objects such as AMD codes and robust secret sharing schemes on the other hand, have been observed previously in several papers, beginning with [13] (see also [4, 5, 15, 14]). The paper [5] and other prior work is mainly concerned with codes that are "close to" optimal and/or the construction of classes of codes that have *asymptotically optimal* behaviour. This is of course desirable from the point of view of applications. In contrast, our focus is on mathematical characterizations of codes where the relevant bounds are *exactly* met with equality; this is the sense in which we are using the term "optimal".

The rest of this paper is organized as follows. In Section 2, we define all the generalizations of difference families that we will be using in the rest of the paper. We give some examples and constructions as well as prove some nonexistence results. Section 3 studies weak AMD codes. Bounds are considered in Section 3.1, where we introduce the notion of *R-optimal* and *G-optimal* AMD codes; these bound arise in the analysis of two different adversarial strategies. Conditions under which these bounds can be met with equality are presented in Section 3.2. Section 4 provides an analogous treatment of strong AMD codes. Finally, we conclude the paper in Section 5.

2 Difference Families and Generalizations

In this section, we describe several variations of difference sets and difference families. These concepts will be essential for constructions and combinatorial characterizations of optimal (strong and weak) AMD codes. Some of the definitions we give are new, and we prove some interesting connections between various types of difference families that may be of independent interest.

Let $\mathcal{G}$ be an abelian group. For any two disjoint sets $A_1, A_2 \subseteq \mathcal{G}$, define

$$\mathcal{D}(A_1, A_2) = \{x - y : x \in A_1, y \in A_2\}.$$

Note that $\mathcal{D}(A_1, A_2)$ is a multiset. Also, for any $A_1 \subseteq \mathcal{G}$, define

$$\mathcal{D}(A_1) = \{x - y : x, y \in A_1, x \neq y\}.$$

$\mathcal{D}(A_1)$ is also a multiset.

Our first two definitions—difference sets and difference families—are standard notions from combinatorial design theory. There is a large literature on these combinatorial structures.

Definition 2.1 (Difference Set)**.** *Let $\mathcal{G}$ be an additive abelian group of order n. An (n, m, λ)-**difference set** (or (n, m, λ)-**DS**) is a set $A_1 \subseteq \mathcal{G}$ with $|A_1| = m$, such that the following multiset equation holds:*

$$\mathcal{D}(A_1) = \lambda(\mathcal{G} \setminus \{0\}).$$

If an (n, m, λ)-DS exists, then $\lambda(n-1) = m(m-1)$.

Remark: We can consider any set of size 1 to be a (trivial) difference set with $\lambda = 0$.

Definition 2.2 (Difference Family)**.** *Let $\mathcal{G}$ be an additive abelian group of order n. An (n, m, k, λ)-**difference family** (or (n, m, k, λ)-**DF**) is a set of m k-subsets of $\mathcal{G}$, say $A_1, \ldots, A_m$, such that the following multiset equation holds:*

$$\bigcup_i \mathcal{D}(A_i) = \lambda(\mathcal{G} \setminus \{0\}).$$

If an (n, m, k, λ)-DF exists, then $\lambda(n-1) = mk(k-1)$. Also, an (n, m, λ)-DS is an $(n, 1, m, \lambda)$-DF.

The following definition is from [14].

Definition 2.3 (External difference family)**.** *Let $\mathcal{G}$ be an additive abelian group of order n. An (n, m, k, λ)-**external difference family** (or (n, m, k, λ)-**EDF**) is a set of m disjoint k-subsets of $\mathcal{G}$, say $A_1, \ldots, A_m$, such that the following multiset equation holds:*

$$\bigcup_{\{i,j:i \neq j\}} \mathcal{D}(A_i, A_j) = \lambda(\mathcal{G} \setminus \{0\}).$$

If an (n, m, k, λ)-EDF exists, then $n \geq mk$ and

$$\lambda(n-1) = k^2 m(m-1). \tag{1}$$

Also, an $(n, m, 1, \lambda)$-EDF is the same thing as an (n, m, λ) difference set.

There are several papers giving construction methods for external difference families, e.g., [2, 7, 8, 9, 10, 11, 18]. Here is an example of one infinite class of external difference families due to Tonchev [18]; it was later rediscovered in [10].

Theorem 2.1. *[18, 10] Suppose that $q = 2u\ell + 1$ is a prime power, where u and ℓ are odd. Then there exists a $(q, u, \ell, (q - 2\ell - 1)/4)$-EDF in $\mathbb{F}_q$.*

Proof. Let $\alpha \in \mathbb{F}_q$ be a primitive element. Let C be the subgroup of ${\mathbb{F}_q}^*$ having order u and index 2ℓ. The ℓ cosets $\alpha^{2i}C$ $(0 \leq i \leq \ell - 1)$ form the EDF. □

Example 2.1. *We give an example to illustrate Theorem 2.1. Let $\mathcal{G} = (\mathbb{Z}_{19}, +)$. Then $\alpha = 2$ is a primitive element and $C = \{1, 7, 11\}$ is the (unique) subgroup of order 3 in ${\mathbb{Z}_{19}}^*$. A $(19, 3, 3, 3)$-EDF is given by the three sets $\{1, 7, 11\}$, $\{4, 9, 6\}$ and $\{16, 17, 5\}$.*

We refer to [9, Table II] for a list of known external difference families.

Remark: The related but more general concept of a *difference system of sets* was defined much earlier, by Levenshtein, in [12]. This is similar to the definition of an external difference family, except that every difference $x - y$ $(x \in A_i, y \in A_j, i \neq j)$ is required to occur *at least* λ times. However, we note that a *perfect, regular difference system of sets* is equivalent to an external difference family.

As we will discuss later, for the applications to AMD codes we will be considering, it is sufficient that every difference occurs *at most* λ times. This motivates the following definition.

Definition 2.4 (Bounded external difference family)**.** *Let $\mathcal{G}$ be an additive abelian group of order n. A (n, m, k, λ)-**bounded external difference family** (or (n, m, k, λ)-**BEDF**) is a set of m disjoint k-subsets of $\mathcal{G}$, say $A_1, \ldots, A_m$, such that the following condition holds for every $g \in \mathcal{G} \setminus \{0\}$:*

$$|\{x - y : x - y = g, x \in A_i, y \in A_j, i \neq j\}| \leq \lambda.$$

It is obvious that an (n, m, k, λ)-EDF is an (n, m, k, λ)-BEDF.

Definition 2.5 (Strong external difference family)**.** *Let $\mathcal{G}$ be an additive abelian group of order n. An $(n, m, k; \lambda)$-**strong external difference family** (or $(n, m, k; \lambda)$-**SEDF**) is a set of m disjoint k-subsets of $\mathcal{G}$, say $A_1, \ldots, A_m$, such that the following multiset equation holds for every i, $1 \leq i \leq m$:*

$$\bigcup_{\{j : j \neq i\}} \mathcal{D}(A_i, A_j) = \lambda(\mathcal{G} \setminus \{0\}). \tag{2}$$

It is easy to see that a (n, m, k, λ)-SEDF is an $(n, m, k, m\lambda)$-EDF. Therefore, from (1), if an (n, m, k, λ)-SEDF exists, then

$$\lambda(n - 1) = k^2(m - 1). \tag{3}$$

Example 2.2. *Let $\mathcal{G} = (\mathbb{Z}_{k^2+1}, +)$, $A_1 = \{0, 1, \ldots, k - 1\}$ and $A_2 = \{k, 2k, \ldots, k^2\}$. This is a $(k^2 + 1, 2; k; 1)$-SEDF.*

Example 2.3. *Let $\mathcal{G} = (\mathbb{Z}_n, +)$ and $A_i = \{i\}$ for $0 \leq i \leq n - 1$. This is a $(n, n; 1; 1)$-SEDF.*

Theorem 2.2. *There does not exist an $(n, m, k, 1)$-SEDF with $m \geq 3$ and $k > 1$.*

Proof. Suppose $A_1, \ldots, A_m$ is an $(n, m, k, 1)$-SEDF with $m \geq 3$ and $k > 1$. From (2), it follows that

$$\bigcup_{\{i,j : 1 \leq i \leq m, 1 \leq j \leq m, i \neq j\}} \mathcal{D}(A_i, A_j) = m(\mathcal{G} \setminus \{0\}). \tag{4}$$

Then, from (2) and (4), we have

$$\bigcup_{\{i,j : 2 \leq i \leq m, 2 \leq j \leq m, i \neq j\}} \mathcal{D}(A_i, A_j) = (m - 2)(\mathcal{G} \setminus \{0\}). \tag{5}$$

Suppose $x, y \in A_1$, $x \neq y$ (note that $k > 1$ so we have two distinct elements in A_1). Now, from (5), since $m > 2$, there exists $u \in A_i$, $v \in A_j$ such that $i, j > 1$, $i \neq j$ and $u - v = x - y$. Then $u - x = v - y$, which contradicts (2). □

Theorem 2.3. *There exists an $(n, m, k, 1)$-SEDF if and only if $m = 2$ and $n = k^2 + 1$, or $k = 1$ and $m = n$.*

Proof. From Theorem 2.2, we only need to consider the cases $m = 2$ and $k = 1$. If $m = 2$, then from (3), we must have $n = k^2 + 1$, and the relevant SEDF exists from Example 2.2. If $k = 1$, then from (3) we must have $m = n$, and the relevant SEDF exists from Example 2.3. □

Next, we consider generalizations of external difference families and strong external difference families in which the subsets $A_1, \ldots, A_m$ are allowed to be of possibly different sizes.

Definition 2.6 (Generalized external difference family)**.** *Let $\mathcal{G}$ be an additive abelian group of order n. An $(n, m; k_1, \ldots, k_m; \lambda)$-**generalized external difference family** (or $(n, m; k_1, \ldots, k_m; \lambda)$-**GEDF**) is a set of m disjoint subsets of $\mathcal{G}$, say $A_1, \ldots, A_m$, such that $|A_i| = k_i$ for $1 \leq i \leq m$ and the following multiset equation holds:*

$$\bigcup_{\{i,j:i \neq j\}} \mathcal{D}(A_i, A_j) = \lambda(\mathcal{G} \setminus \{0\}).$$

Clearly, an (n, m, k, λ)-EDF is an $(n, m; k, \ldots, k; \lambda)$-GEDF.

Example 2.4. *Let $\mathcal{G} = (\mathbb{Z}_{13}, +)$, $A_1 = \{0, 1\}$ and $A_2 = \{2, 4, 6\}$. This is a $(13, 2; 2, 3; 1)$-GEDF.*

Example 2.5. *Let $\mathcal{G} = (\mathbb{Z}_{11}, +)$, $A_1 = \{0\}$, $A_2 = \{1\}$, and $A_3 = \{3, 5\}$. This is a $(11, 3; 1, 1, 2; 1)$-GEDF.*

Remark: A generalized external difference family is also known as a *perfect difference system of sets.*

Definition 2.7 (Generalized strong external difference family)**.** *Let $\mathcal{G}$ be an additive abelian group of order n. An $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-**generalized strong external difference family** (or $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-**GSEDF**) is a set of m disjoint subsets of $\mathcal{G}$, say $A_1, \ldots, A_m$, such that $|A_i| = k_i$ for $1 \leq i \leq m$ and the following multiset equation holds for every i, $1 \leq i \leq m$:*

$$\bigcup_{\{j:j \neq i\}} \mathcal{D}(A_i, A_j) = \lambda_i(\mathcal{G} \setminus \{0\}).$$

It is obvious that an (n, m, k, λ)-SEDF is an $(n, m; k, \ldots, k; \lambda, \ldots, \lambda)$-GSEDF.

Example 2.6. *Let $\mathcal{G} = (\mathbb{Z}_n, +)$, $A_1 = \{0\}$ and $A_2 = \{1, 2, \ldots, n-1\}$. This is a $(n, 2; 1, n-1; 1, 1)$-GSEDF.*

Example 2.7. *Let $\mathcal{G} = (\mathbb{Z}_7, +)$, $A_1 = \{1\}$, $A_2 = \{2\}$, $A_3 = \{4\}$, and $A_4 = \{0, 3, 5, 6\}$. This is a $(7, 4; 1, 1, 1, 4; 1, 1, 1, 2)$-GSEDF.*

A $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-GSEDF is *maximal* if $\sum k_i = n$. Here is a nice characterization of maximal GSEDF.

Theorem 2.4. *Suppose $A_1, \ldots, A_m$ is a partition of $\mathcal{G}$ (where $|\mathcal{G}| = n$) with $|A_i| = k_i$ for $1 \leq i \leq m$. Then $A_1, \ldots, A_m$ is a (maximal) $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-GSEDF if and only if A_i is an $(n, k_i, k_i - \lambda_i)$-DS in $\mathcal{G}$, for $1 \leq i \leq m$.*

Proof. Fix a value i, $1 \leq i \leq m$. It is clear that

$$\begin{aligned}
\bigcup_{\{j:j\neq i\}} \mathcal{D}(A_i, A_j) &= \mathcal{D}(A_i, \mathcal{G} \setminus A_i) \\
&= \bigcup_{x \in A_i} \mathcal{D}(x, \mathcal{G} \setminus A_i) \\
&= \bigcup_{x \in A_i} (\mathcal{D}(x, \mathcal{G} \setminus \{x\}) \setminus \mathcal{D}(x, A_i \setminus \{x\})) \\
&= \left(\bigcup_{x \in A_i} \mathcal{D}(x, \mathcal{G} \setminus \{x\}) \right) \setminus \left(\bigcup_{x \in A_i} \mathcal{D}(x, A_i \setminus \{x\}) \right) \\
&= \left(\bigcup_{x \in A_i} \mathcal{G} \setminus \{0\}) \right) \setminus \mathcal{D}(A_i) \\
&= (k_i(\mathcal{G} \setminus \{0\})) \setminus \mathcal{D}(A_i),
\end{aligned}$$

where all operations are multiset operations. Therefore,

$$\bigcup_{\{j:j\neq i\}} \mathcal{D}(A_i, A_j) = \lambda_i(\mathcal{G} \setminus \{0\})$$

if and only if

$$\mathcal{D}(A_i) = (k_i - \lambda_i)(\mathcal{G} \setminus \{0\}).$$

□

Theorem 2.5. *Suppose there exists an $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-GSEDF where $k_i = 1$. Then $\lambda_i = 1$ and $\sum_{i=1}^{m} k_i = n$ (i.e., the GSEDF is maximal).*

Proof. We have $k_i(a - k_i) = a - 1 = \lambda_i(n - 1)$, where $a = \sum_{i=1}^{m} k_i$. Since $a \leq n$ and $\lambda_i \geq 1$, it must be the case that $a = n$ and $\lambda_i = 1$. □

Definition 2.8 (Bounded generalized strong external difference family)**.** *Let $\mathcal{G}$ be an additive abelian group of order n. An $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-**bounded generalized strong external difference family** (or $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-**BGSEDF**) is a set of m disjoint subsets of $\mathcal{G}$, say $A_1, \ldots, A_m$, such that $|A_i| = k_i$ for $1 \leq i \leq m$ and the following multiset equation holds for every j, $1 \leq j \leq m$, and for every $g \in \mathcal{G} \setminus \{0\}$:*

$$|\{x - y : x - y = g, x \in A_i, y \in A_j, i \neq j\}| \leq \lambda_j.$$

Remark: A BGSEDF is equivalent to the notion of a *differential structure*, as defined, e.g., in [5].

Definition 2.9 (Partitioned external difference family)**.** *Let $\mathcal{G}$ be an additive abelian group of order n. An $(n, m; c_1, \ldots, c_\ell; k_1, \ldots, k_\ell; \lambda_1, \ldots, \lambda_\ell)$-**partitioned external difference family** (or $(n, m; c_1, \ldots, c_\ell; k_1, \ldots, k_\ell; \lambda_1, \ldots, \lambda_\ell)$-**PEDF**) is a set of $m = \sum_i c_i$ disjoint subsets of $\mathcal{G}$, say $A_1, \ldots, A_m$, such that there are c_h subsets of size k_h, for $1 \leq h \leq \ell$, and the following multiset equation holds for every h, $1 \leq h \leq \ell$:*

$$\bigcup_{\{i:|A_i|=c_h\}} \bigcup_{\{j:j\neq i\}} \mathcal{D}(A_i, A_j) = \lambda_i(\mathcal{G} \setminus \{0\}).$$

We note the following:

- an $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-GSEDF is an $(n, m; 1, \ldots, 1; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-PEDF
- an (n, m, k, λ)-EDF is an $(n, m; m; k; \lambda)$-PEDF
- an $(n, m; c_1, \ldots, c_\ell; k_1, \ldots, k_\ell; \lambda_1, \ldots, \lambda_\ell)$-PEDF is an $(n, m; {k_1}^{c_1}, \ldots, {k_\ell}^{c_\ell}; \lambda)$-GEDF in which
$$\lambda = \sum_{i=1}^{\ell} \lambda_i,$$
where the notation ${k_i}^{c_i}$ denotes c_i occurrences of k_i, for $1 \leq h \leq \ell$.

Here is an example of a PEDF that is not an EDF or GSEDF.

Example 2.8. *Let $\mathcal{G} = (\mathbb{Z}_{13}, +)$, $A_1 = \{0, 1, 4\}$, $A_2 = \{3, 5, 10\}$, $A_3 = \{2, 6, 7, 9\}$, $A_4 = \{8\}$, $A_5 = \{11\}$, $A_6 = \{12\}$. It can be verified that $A_1, \ldots, A_6$ is a $(13, 6; 2, 1, 3; 3, 4, 1; 5, 3, 3)$-PEDF. To see that it is not a GSEDF, we first compute the occurrence of differences from A_1 to the union of the other A_i's:*

difference	1	2	3	4	5	6	7	8	9	10	11	12
frequency	2	3	2	2	3	3	3	3	2	2	3	2

Then we compute the occurrence of differences from A_2 to the union of the other A_i's:

difference	1	2	3	4	5	6	7	8	9	10	11	12
frequency	3	2	3	3	2	2	2	2	3	3	2	3

These two lists of occurrences of differences are not uniform, so we do not have a GSEDF. However, each difference occurs a total of five times in the two lists.

Theorem 2.6. *Suppose $A_1, \ldots, A_m$ is a partition of $\mathcal{G}$ (where $|\mathcal{G}| = n$) such that there are c_h subsets of size k_h for $1 \leq h \leq \ell$. Then $A_1, \ldots, A_m$ is a (maximal) $(n, m; c_1, \ldots, c_\ell; k_1, \ldots, k_\ell; \lambda_1, \ldots, \lambda_\ell)$-PEDF if and only if the subsets of cardinality k_h form an $(n, k_h, c_h k_h - \lambda_h)$-DF in $\mathcal{G}$, for $1 \leq h \leq \ell$.*

Proof. We omit the proof, which is similar to the proof of Theorem 2.4. $\square$

Example 2.9. *Let's look again at the PEDF in Example 2.8. Here the two sets of size 3 form a $(13, 2, 3, 1)$-DF; the set of size 4 is a $(13, 1, 4, 1)$-DF; and the three sets of size 1 form a $(13, 3, 1, 0)$-DF.*

In Figure 1, we indicate the relationship between the various types of difference families we have defined. If we designate $X \rightarrow Y$, this indicates that any example of "X" automatically satisfies the properties of "Y".

3 Weak AMD Codes

Our goal is to prove lower bounds on the adversary's optimal success probability, $\hat{\epsilon}$. Note that a *lower* bound on $\hat{\epsilon}$ states that there exists an adversary who wins the relevant game with *at least* some specified probability. Then we construct codes that meet these lower bounds, i.e., codes in which the adversary cannot succeed with higher probability. Whenever possible, we will prove bounds without assuming that the code is uniform or has equiprobable encoding (we do assume equiprobable sources, however).

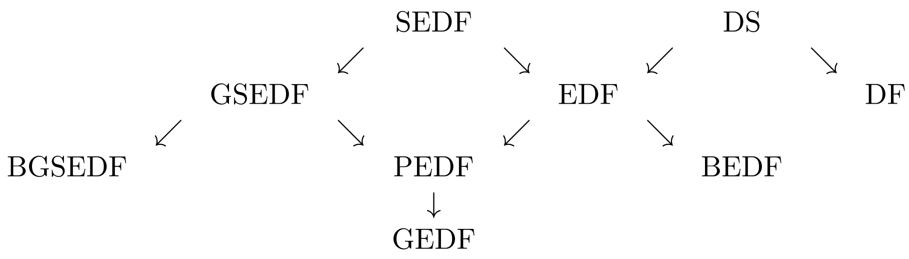

Figure 1: Relationships between various types of difference families

3.1 Bounds for Weak AMD Codes

Theorem 3.1. *In any weak $(m, n, \hat{\epsilon})$-AMD code, it holds that*

$$\hat{\epsilon} \geq \frac{a(m-1)}{m(n-1)}.$$

Proof. Suppose the adversary chooses the value $\Delta \in \mathcal{G} \setminus \{0\}$ uniformly at random. For any given $g \in A(s)$ and for a randomly chosen Δ, the probability that the adversary wins is $(a - a_s)/(n-1)$. The success probability ϵ_{rand} of this random strategy rand is

$$\begin{aligned}
\epsilon_{\mathsf{rand}} &= \sum_s \mathbf{Pr}[s] \sum_{g \in A(s)} \left(\mathbf{Pr}[E(s) = g] \times \frac{a - a_s}{n-1} \right) \\
&= \sum_s \left(\mathbf{Pr}[s] \times \frac{a - a_s}{n-1} \right) \\
&= \frac{a}{n-1} - \sum_s \frac{a_s}{m(n-1)} \quad \text{(because the sources are equiprobable)} \\
&= \frac{a}{n-1} - \frac{a}{m(n-1)} \\
&= \frac{a(m-1)}{m(n-1)}.
\end{aligned}$$

□

Corollary 3.2. *In any k-uniform weak $(m, n, \hat{\epsilon})$-AMD code, it holds that*

$$\hat{\epsilon} \geq \frac{k(m-1)}{n-1}.$$

Proof. Note that $a = km$ in a k-uniform code and apply Theorem 3.1. □

Definition 3.1. *We will define a weak AMD code that meets the bound of Theorem 3.1 (or Corollary 3.2, in the case that the code is k-uniform) with equality to be* R-optimal*. Here, "R" is used to indicate that* rand *is an optimal strategy.*

Corollary 3.3. *[5, Theorem 2.2] In any weak $(m, n, \hat{\epsilon})$-AMD code, it holds that*

$$\hat{\epsilon} \geq \frac{m-1}{n-1}.$$

Proof. Note that $a \geq m$ and apply Theorem 3.1. □

Remark: The bound of Corollary 3.3 is met with equality only if the code is deterministic.

Here is a new bound for weak AMD codes, that arises from a different adversarial strategy.

Theorem 3.4. *In any weak $(m, n, \hat{\epsilon})$-AMD code, it holds that*

$$\hat{\epsilon} \geq \frac{1}{a}.$$

Proof. We consider the following strategy guess for the adversary:

1. Find the encoding $\hat{g} \in \mathcal{A}$ that occurs with the highest probability. Observe that $\mathbf{Pr}[\hat{g}] \geq 1/a$.
2. Pick a Δ that will work for the particular encoding $\hat{g}$.

Clearly, the success probability $\epsilon_{\mathsf{guess}}$ of the strategy guess is equal to $\mathbf{Pr}[\hat{g}] \geq 1/a$. □

Definition 3.2. *We will define a weak AMD code that meets the bound of Theorem 3.4 with equality to be* G-optimal*. Here, "G" is used to indicate that* guess *is an optimal strategy.*

Theorem 3.5. *In any weak $(m, n, \hat{\epsilon})$-AMD code, it holds that*

$$\hat{\epsilon}^2 \geq \frac{m-1}{m(n-1)}.$$

Proof. Multiply the bounds proven in Theorems 3.1 and 3.4. □

A code that meets the bound of Theorem 3.5 with equality is simultaneously R-optimal and G-optimal.

3.2 Optimal Weak AMD Codes

In this section, we consider weak AMD codes that are R-optimal and/or G-optimal. Recall that a weak AMD code is R-optimal if $\hat{\epsilon} = a(m-1)/(m(n-1))$ and it is G-optimal if $\hat{\epsilon} = 1/a$.

3.2.1 R-Optimal Weak AMD Codes

First, we consider R-optimality. Consider the strategy $g \mapsto g + \Delta$, where $\Delta \neq 0$, and let ϵ_Δ denote the success probability of this strategy. Clearly, we have

$$\hat{\epsilon} = \max\{\epsilon_\Delta : \Delta \neq 0\}. \tag{6}$$

For any $\Delta \neq 0$, define

$$\mathsf{Good}(\Delta) = \{g \in \mathcal{G}_0 : g \in A(s) \text{ and } g + \Delta \in A(s'), \text{where } s' \neq s\}. \tag{7}$$

$\mathsf{Good}(\Delta)$ denotes the set of encodings g under which a substitution $g \mapsto g + \Delta$ will result in the adversary winning the game.

Lemma 3.6. *For any $\Delta \neq 0$, it holds that*

$$\epsilon_\Delta = \sum_{g \in \mathsf{Good}(\Delta)} \mathbf{Pr}[g]. \tag{8}$$

Proof. It is clear that

$$\begin{aligned} \epsilon_\Delta &= \mathbf{Pr}[g \in \mathsf{Good}(\Delta)] \\ &= \sum_{g \in \mathsf{Good}(\Delta)} \mathbf{Pr}[g]. \end{aligned}$$

□

Theorem 3.7. *A weak AMD code is R-optimal if and only if $\epsilon_\Delta = a(m-1)/(m(n-1))$ for all $\Delta \neq 0$.*

Proof. Suppose we have an R-optimal weak AMD code. It is not hard to compute

$$\begin{aligned} \sum_{\Delta \neq 0} \epsilon_\Delta &= \sum_{\Delta \neq 0} \sum_{g \in \mathsf{Good}(\Delta)} \mathbf{Pr}[g] \\ &= \sum_{g \in \mathcal{G}_0} \mathbf{Pr}[g] \times |\{\Delta : g \in \mathsf{Good}(\Delta)\}| \\ &= \sum_{s \in \mathcal{S}} \sum_{g \in A(s)} \mathbf{Pr}[s]\, \mathbf{Pr}[E(s) = g] \times |\{\Delta : g \in \mathsf{Good}(\Delta)\}| \\ &= \sum_{s \in \mathcal{S}} \mathbf{Pr}[s] \sum_{g \in A(s)} \mathbf{Pr}[E(s) = g](a - a_s) \\ &= \sum_{s \in \mathcal{S}} \mathbf{Pr}[s](a - a_s) \\ &= \sum_{s \in \mathcal{S}} \frac{1}{m}(a - a_s) \\ &= \frac{a(m-1)}{m}. \end{aligned}$$

Therefore the average of the quantities ϵ_Δ ($\Delta \neq 0$) is equal to $a(m-1)/(m(n-1))$. In order to have $\hat{\epsilon} = a(m-1)/(m(n-1))$, it must be the case that $\epsilon_\Delta = a(m-1)/(m(n-1))$ for all $\Delta \neq 0$. □

We next present a method of constructing R-optimal weak AMD codes.

Theorem 3.8. *Suppose there is an $(n, m; k_1, \dots, k_m; \lambda_1, \dots, \lambda_m)$-GSEDF. Then there is an (R-optimal) weak $(m, n, a(m-1)/(m(n-1)))$-AMD code, where $a = \sum_{i=1}^m k_i$.*

Proof. Suppose the GSEDF is given by $A_1, \dots, A_m$. Let $a = \sum_{i=1}^m k_i$. Observe that

$$k_i(a - k_i) = \lambda_i(n-1) \tag{9}$$

for $1 \leq i \leq m$. Let $\mathcal{S} = \{s_1, \dots, s_m\}$ be a set of m sources. For $1 \leq i \leq m$, define $A(s_i) = A_i$ and suppose the encoding function $E(s_i)$ is equiprobable. We show that $\epsilon_\Delta = a(m-1)/(m(n-1))$ for all $\Delta \neq 0$. We have

$$\begin{aligned}
\epsilon_\Delta &= \sum_{g \in \mathsf{Good}(\Delta)} \mathbf{Pr}[g] \\
&= \sum_{i=1}^{m} \frac{1}{m} \times \frac{\lambda_i}{k_i} \\
&= \frac{1}{m} \sum_{i=1}^{m} \frac{a-k_i}{n-1} \quad \text{from (9)} \\
&= \frac{1}{m(n-1)} \sum_{i=1}^{m} (a-k_i) \\
&= \frac{a(m-1)}{m(n-1)}.
\end{aligned}$$

□

In fact, we can obtain R-optimal weak AMD codes from a weaker type of difference family, namely, a PEDF.

Theorem 3.9. *Suppose there is an $(n, m; c_1, \ldots, c_\ell; k_1, \ldots, k_\ell; \lambda_1, \ldots, \lambda_\ell)$-PEDF. Then there is an (R-optimal) weak $(m, n, a(m-1)/(m(n-1)))$-AMD code, where $a = \sum_{h=1}^{\ell} c_h k_h$.*

Proof. We omit the proof, which is similar to the proof of Theorem 3.8. □

It is interesting to note that the we do not necessarily obtain an R-optimal AMD code if we start from an arbitrary generalized external difference family. As an example, suppose we construct an AMD code with equiprobable encoding for two sources using the GEDF presented in Example 2.4. Here it is easy to compute

$$\epsilon_1 = \frac{1}{4} > \frac{a(m-1)}{m(n-1)} = \frac{5 \times 1}{2 \times 12} = \frac{5}{24},$$

so this code is not R-optimal

It is an open problem to characterize R-optimal (weak) AMD codes. The following example illustrates that the converse of Theorem 3.9 is not true in general. That, is we can construct R-optimal codes that do not come from PEDFs.

Example 3.1. *Let $\mathcal{S} = \{1, 2, 3, 4\}$ and let $\mathcal{G} = \mathbb{Z}_{10}$. The encoding function E is defined by $E(1) = 0$, $E(2) = 5$, $E(3) \in_R \{1, 9\}$ and $E(4) \in_R \{2, 3\}$.*

Suppose the adversary chooses $\Delta = 5$; then the adversary wins if $s \in \{1, 2\}$, which occurs with probability $1/2$. Suppose the adversary chooses $\Delta = 1$; then the adversary succeeds if $s \in \{1, 3\}$, which occurs with probability $1/2$. Suppose the the adversary chooses $\Delta = 2$; then the adversary succeeds if $s = 1$, if $s = 3$ and $E(s) = 1$, or if $s = 4$ and $E(s) = 3$. The success probability here is

$$\frac{1}{4} + \frac{1}{4} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2} = \frac{1}{2}.$$

The remaining choices for Δ can be checked in a similar way. We obtain a code with success probability $1/2$. Since $m = 4$, $n = 10$ and $a = 6$, we have $a(m-1)/(m(n-1)) = 18/36 = 1/2$, so the code is R-optimal. However, the sets $\{0\}, \{5\}, \{1, 9\}, \{2, 8\}$ do not form a PEDF.

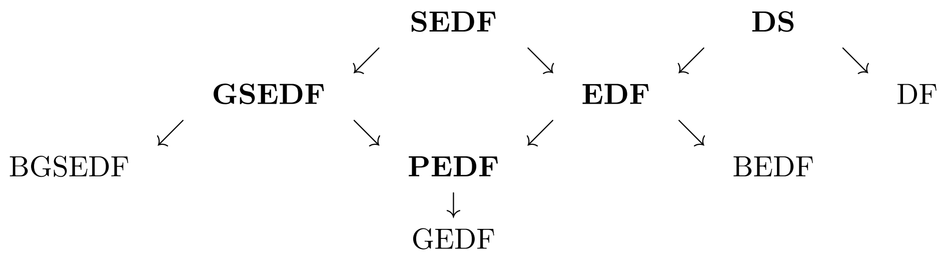

Figure 2: Difference families that yield R-optimal weak AMD codes (indicated in boldface type)

We can give a tight characterization of *k-regular* R-optimal weak AMD codes, however, as follows.

Theorem 3.10. *An (R-optimal) k-regular weak $(m, n, k(m-1)/(n-1))$-AMD code is equivalent to an (n, m, k, λ)-EDF.*

Proof. Suppose $A_1, \dots, A_m$ is an (n, m, k, λ)-EDF. Let $\mathcal{S} = \{s_1, \dots, s_m\}$ be a set of m sources. For $1 \leq i \leq m$, suppose the encoding function $E(s_i)$ is equiprobable. The resulting weak AMD code is k-regular. Choose any $\Delta \in \mathcal{G}$, $\Delta \neq 0$. The strategy $g \mapsto g + \Delta$ succeeds with probability $\epsilon_\Delta = \lambda/(km) = k(m-1)/(n-1)$. (In fact, this follows from Theorem 3.9.)

Conversely, suppose we have an R-optimal k-regular weak AMD code. Then it must be the case that $\epsilon_\Delta = k(m-1)/(n-1)$ for all $\Delta \neq 0$. Using the fact that the code is a k-regular AMD, we have

$$\frac{k(m-1)}{n-1} = \epsilon_\Delta = \mathbf{Pr}[E(s) \in \mathsf{Good}(\Delta)] = \frac{|\mathsf{Good}(\Delta)|}{km}.$$

Therefore,

$$|\mathsf{Good}(\Delta)| = \frac{k^2 m(m-1)}{n-1}.$$

It then follows that $\{A(s) : s \in \mathcal{S}\}$ is an (n, m, k, λ)-EDF, where $\lambda = k^2 m(m-1)/(n-1)$. □

In Figure 2 we indicate the types of difference families that yield R-optimal weak AMD codes. This summarizes the results proven in this section.

3.2.2 G-Optimal Weak AMD Codes

Now we turn to G-optimality. We have the following characterization of G-optimal weak AMD codes.

Theorem 3.11. *A (G-optimal) weak $\left(m, n, \frac{1}{a}\right)$-AMD code is equivalent to an $(n, m, k, 1)$-BEDF, where $a = km$.*

Proof. Suppose $A_1, \dots, A_m$ is an $(n, m, k, 1)$-BEDF. Let $\mathcal{S} = \{s_1, \dots, s_m\}$ be a set of m sources. For $1 \leq i \leq m$, define an encoding function $E(s_i)$ which chooses an element of A_i uniformly at random. Choose any $\Delta \in \mathcal{G}$, $\Delta \neq 0$. The strategy $g \mapsto g + \Delta$ succeeds with probability $\epsilon_\Delta \leq 1/(km) = 1/a$, since there is at most one occurrence of the difference Δ in the BEDF. Further, if $\Delta \in \mathcal{D}(A_j, A_i)$ where $i \neq j$, then the strategy $g \mapsto g + \Delta$ succeeds with probability $1/a$.

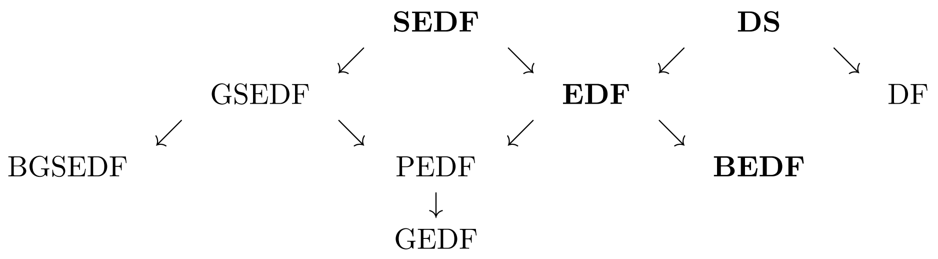

Figure 3: Difference families with $\lambda = 1$ that yield G-optimal weak AMD codes (indicated in boldface type)

Conversely, suppose we have a G-optimal weak AMD code. From the proof of Theorem 3.4, we see that all encodings must occur with the same probability, $1/a$. Since the sources are equiprobable, this happens only if the code is k-regular with $k = a/m$. Now we claim that $\{A(s) : s \in \mathcal{S}\}$ is an $(n, m, k, 1)$-BEDF. This is easy to see, because if some difference occurred more than once, it would immediately follow that $\epsilon \geq 2/a$. □

Now we characterize k-regular weak AMD codes that are simultaneously R-optimal and G-optimal.

Theorem 3.12. *A k-regular weak AMD code that is simultaneously R-optimal and G-optimal is equivalent to an $(n, m, k, 1)$-EDF.*

Proof. From Theorem 3.5, the code has success probability $\sqrt{\frac{m-1}{m(n-1)}}$. In order for this to occur, the bounds of Corollary 3.2 and 3.4 both must hold with equality. Therefore the AMD code is simultaneously an (n, m, k, λ)-EDF (from Theorem 3.10) and an $(n, m, k, 1)$-BEDF (from Theorem 3.11). Hence, it is an $(n, m, k, 1)$-EDF. □

In Figure 3 we indicate the types of difference families that yield G-optimal weak AMD codes. Note that the relevant difference families are assumed to have $\lambda = 1$ in this figure.

4 Strong AMD Codes

We begin by focussing on the success probability of the adversary when the source is fixed to be s. Let $\hat{\epsilon}_s$ be the success probability of the optimal strategy for the given source s.

Theorem 4.1. *In any strong AMD code, it holds that*

$$\hat{\epsilon}_s \geq \frac{a - a_s}{n - 1}$$

for any source $s \in \mathcal{S}$.

Proof. As in the proof of Theorem 3.1, we consider a random strategy, i.e., $\Delta \neq 0$ is chosen uniformly at random. Given that the source is s, it is easy to see that the success probability of this strategy will be

$$\frac{a - a_s}{n - 1}.$$

□

Definition 4.1. *We will define a strong AMD code that meets the bound of Theorem 4.1 with equality for every possible source s to be* R-optimal. *Again, "R" is used to indicate that choosing $\Delta \neq 0$ uniformly at random is an optimal strategy.*

Corollary 4.2. *In any strong $(m, n, \hat{\epsilon})$-AMD code, it holds that*

$$\hat{\epsilon} \geq \frac{a - a_{s'}}{n - 1},$$

where $a_{s'} = \min\{a_s : s \in \mathcal{S}\}$.

Proof. The quantity $(a - a_s)/(n - 1)$ is maximized when a_s is minimized. □

If the code is k-uniform, then the previous bound takes a simpler form.

Corollary 4.3. *In any k-uniform strong $(m, n, \hat{\epsilon})$-AMD code, it holds that*

$$\hat{\epsilon} \geq \frac{k(m - 1)}{n - 1}.$$

Proof. Here $a_s = k$ for all s and $a = km$. Apply Corollary 4.2. □

Theorem 4.4. *In any strong AMD code, it holds that $\hat{\epsilon}_s \geq 1/a_s$, for any source $s \in \mathcal{S}$.*

Proof. Given any source s, the adversary can try to guess the encoded message $E(s)$ that is output. The adversary will maximize his probability of success by choosing g such that $\mathbf{Pr}[g \mid s]$ is maximized. Note that there exists a g such that $\mathbf{Pr}[g \mid s] \geq 1/a_s$. Then the adversary can choose Δ such that $g + \Delta \in \mathcal{G}_0 \setminus A(s)$. The success probability of this strategy is clearly at least $1/a_s$. □

Definition 4.2. *We will define a strong AMD code that meets the bound of Theorem 4.4 with equality for every possible source s to be* G-optimal. *Again, "G" is used to indicate that guessing the most likely encoding is an optimal strategy.*

Corollary 4.5. *In any strong $(m, n, \hat{\epsilon})$-AMD code, it holds that $\hat{\epsilon} \geq 1/a_{s'}$, where $a_{s'} = \min\{a_s : s \in \mathcal{S}\}$.*

Proof. The quantity $1/a_s$ is maximized when a_s is minimized. □

In the case of a k-regular code, we have the following corollary.

Corollary 4.6. *In any k-regular strong $(m, n, \hat{\epsilon})$-AMD code, it holds that $\hat{\epsilon} \geq 1/k$.*

We now have an easy proof of the following previously known bound.

Theorem 4.7. *[5, Theorem 2.2] In any k-uniform, strong $(m, n, \hat{\epsilon})$-AMD code, it holds that*

$$\hat{\epsilon}^2 \geq \frac{m - 1}{n - 1}.$$

Proof. From Corollary 4.3, we have

$$\hat{\epsilon} \geq \frac{k(m - 1)}{n - 1}.$$

Furthermore, from Corollary 4.6, we have $\hat{\epsilon} \geq 1/k$. Multiplying these two inequalities, we get

$$\hat{\epsilon}^2 \geq \frac{m - 1}{n - 1}.$$

□

Remark: We will determine in Theorem 4.14 necessary and sufficient conditions for the bound of Theorem 4.7 to be met with equality in all nontrivial cases, i.e., when $\hat{\epsilon} < 1$.

4.1 Optimal Strong AMD Codes

4.1.1 R-Optimal Strong AMD Codes

Suppose the source s is fixed. Consider the strategy $g \mapsto g + \Delta$, where $\Delta \neq 0$. Let $\epsilon_{\Delta,s}$ denote the success probability of this strategy. Then it is clear that

$$\hat{\epsilon}_s = \max\{\epsilon_{\Delta,s} : \Delta \neq 0\}. \tag{10}$$

For any $\Delta \neq 0$, define

$$\mathsf{Good}(\Delta, s) = \{g : g \in A(s) \text{ and } g + \Delta \in A(s'), \text{where } s' \neq s\}. \tag{11}$$

This is the same definition as (7), except that s is now fixed.

Lemma 4.8. *For any $\Delta \neq 0$, it holds that*

$$\epsilon_{\Delta,s} = \sum_{g \in \mathsf{Good}(\Delta,s)} \mathbf{Pr}[E(s) = g]. \tag{12}$$

Proof. It is clear that

$$\begin{aligned} \epsilon_{\Delta,s} &= \mathbf{Pr}[E(s) \in \mathsf{Good}(\Delta, s)] \\ &= \sum_{g \in \mathsf{Good}(\Delta,s)} \mathbf{Pr}[E(s) = g]. \end{aligned}$$

□

Theorem 4.9. *In any strong AMD code, $\hat{\epsilon}_s = (a - a_s)/(n-1)$ if and only if $\epsilon_{\Delta,s} = (a - a_s)/(n-1)$ for all $\Delta \neq 0$.*

Proof. Suppose we have an AMD code where $\hat{\epsilon}_s = (a - a_s)/(n - 1)$. It is not hard to compute

$$\begin{aligned} \sum_{\Delta \neq 0} \epsilon_{\Delta,s} &= \sum_{\Delta \neq 0} \sum_{g \in \mathsf{Good}(\Delta,s)} \mathbf{Pr}[E(s) = g] \\ &= \sum_{g \in A(s)} \mathbf{Pr}[E(s) = g] \times |\{\Delta : g \in \mathsf{Good}(\Delta, s)\}| \\ &= \sum_{g \in A(s)} \mathbf{Pr}[E(s) = g] \times (a - a_s) \\ &= a - a_s. \end{aligned}$$

Therefore the average of the quantities $\epsilon_{\Delta,s}$ ($\Delta \neq 0$) is equal to $(a - a_s)/(n - 1)$. In order to have $\hat{\epsilon}_s = (a - a_s)/(n - 1)$, it must be the case that $\epsilon_{\Delta,s} = (a - a_s)/(n - 1)$ for all $\Delta \neq 0$. □

Theorem 4.10. *Suppose there is an $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-GSEDF. Then there is an R-optimal strong AMD code where $a = \sum_{i=1}^{m} k_i$.*

Proof. Suppose the GSEDF is given by $A_1, \ldots, A_m$. Let $\mathcal{S} = \{s_1, \ldots, s_m\}$ be a set of m sources. For $1 \leq i \leq m$, define $A(s_i) = A_i$, so $a_{s_i} = k_i$, and suppose the encoding function $E(s_i)$ is equiprobable. We show that $\epsilon_{\Delta,s_i} = (a - a_{s_i})/(n-1)$ for $1 \leq i \leq m$ and all $\Delta \neq 0$. We have

$$\begin{aligned}\epsilon_{\Delta,s_i} &= \sum_{g \in \mathsf{Good}(\Delta,s_i)} \mathbf{Pr}[g] \\ &= \frac{\lambda_i}{k_i} \\ &= \frac{a - k_i}{n-1} \quad \text{from (9)} \\ &= \frac{a - a_{s_i}}{n-1}.\end{aligned}$$

□

It is possible to prove a converse to Theorem 4.10 in the case where the AMD code has equiprobable encoding.

Theorem 4.11. *Suppose there is an R-optimal strong AMD code with equiprobable encoding. Then the sets $A(s)$ $(s \in \mathcal{S})$ form an $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-GSEDF.*

Proof. Suppose the sources are denoted $\mathcal{S} = \{s_1, \ldots, s_m\}$. Fix a value i, $1 \leq i \leq m$ and let $\Delta \neq 0$. We have

$$\begin{aligned}\epsilon_{\Delta,s_i} &= \frac{a - a_{s_i}}{n-1} \\ &= \sum_{g \in \mathsf{Good}(\Delta,s_i)} \mathbf{Pr}[g] \\ &= \frac{|\mathsf{Good}(\Delta, s_i)|}{a_{s_i}}.\end{aligned}$$

Therefore, for a fixed value i, we have

$$|\mathsf{Good}(\Delta, s_i)| = \frac{a_{s_i}(a - a_{s_i})}{n-1}$$

for all $\Delta \neq 0$. This says that

$$\mathcal{D}(A(s_i), \mathcal{G}_0 \setminus A(s_i)) = \lambda_i(\mathcal{G} \setminus \{0\}),$$

where

$$\lambda_i = \frac{a_{s_i}(a - a_{s_i})}{n-1}.$$

□

Remark: The results we have proven in Theorems 4.10 and 4.11 establish a close connection between R-optimal strong AMD codes and GSEDF. In [5], similar results were proven, using the language of differential structures, that showed the link between (not necessarily optimal) strong AMD codes and BGSEDF.

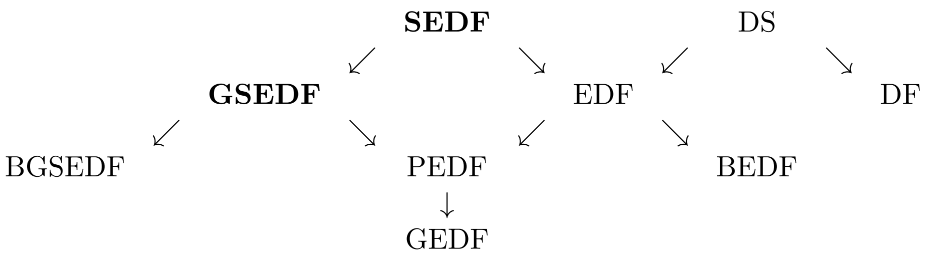

Figure 4: Difference families that yield R-optimal strong AMD codes (indicated in boldface type)

4.1.2 G-Optimal Strong AMD Codes

Now we turn to G-optimality. We have the following characterization of G-optimal strong AMD codes.

Theorem 4.12. *A G-optimal strong AMD code is equivalent to an $(n, m; k_1, \ldots, k_m; 1, \ldots, 1)$-BGSEDF.*

Proof. Suppose $A_1, \ldots, A_m$ is an $(n, m; k_1, \ldots, k_m; 1, \ldots, 1)$-BGSEDF. Let $\mathcal{S} = \{s_1, \ldots, s_m\}$ be a set of m sources. For $1 \leq i \leq m$, define an encoding function $E(s_i)$ which chooses an element of A_i uniformly at random. Let $1 \leq i \leq m$ and choose any $\Delta \in \mathcal{G} \setminus \{0\}$. Given that the source is s_i, the strategy $g \mapsto g + \Delta$ succeeds with probability at most $1/a_{s_i}$, since there is at most one $g \in A_i$ such that $g + \Delta \in A_j$ and $j \neq i$. Further, there exists a $\Delta \neq 0$ such that this strategy succeeds with probability $1/a_{s_i}$.

Conversely, suppose we have a G-optimal strong AMD code. Let $s \in \mathcal{S}$. From the proof of Theorem 4.4, it is easy to see that all encodings of s occur with the same probability $1/|A(s)|$. Now we claim that $\{A(s) : s \in \mathcal{S}\}$ is an $(n, m; k_1, \ldots, k_m; 1, \ldots, 1)$-BGSEDF. Suppose that there existed two different values $g, g' \in A_i$ such that $g + \Delta \in A_j$, $g' + \Delta \in A_{j'}$ and $j, j' \neq i$. It would then follow that $\hat{\epsilon}_s \geq 2/|A(s)|$, which is a contradiction. □

Now we show that k-regular strong AMD codes with $m \geq 3$ cannot be simultaneously R-optimal and G-optimal.

Theorem 4.13. *There does not exist a strong AMD code with $m \geq 3$ and $\hat{\epsilon} < 1$ that is simultaneously R-optimal and G-optimal.*

Proof. Since the code is G-optimal, it follows from Theorem 4.12 and its proof that the code has equiprobable encoding and is derived from $(n, m; k_1, \ldots, k_m; 1, \ldots, 1)$-BGSEDF. Now, since the code is R-optimal and it has equiprobable encoding, Theorem 4.11 shows that the code is derived from $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-GSEDF. Thus we have an $(n, m; k_1, \ldots, k_m; 1, \ldots, 1)$-BGSEDF that is also an $(n, m; k_1, \ldots, k_m; \lambda_1, \ldots, \lambda_m)$-GSEDF, so it must in fact be an $(n, m; k_1, \ldots, k_m; 1, \ldots, 1)$-GSEDF. This implies that $k_i(a - k_i) = n - 1$ for all i. Given a and n, the equation $x(a - x) = n - 1$ has at most two distinct roots, and these roots sum to a. Suppose that $k_i \neq k_j$ for some i, j. Then $k_i + k_j = a$, which implies that $m = 2$, a contradiction. Hence the code is k-uniform and the GSEDF is in fact an $(n, m; k; 1)$-SEDF. Now Theorem 2.3 implies that $k = 1$ and $n = m$. This code has $\hat{\epsilon} = 1$, so we are done. □

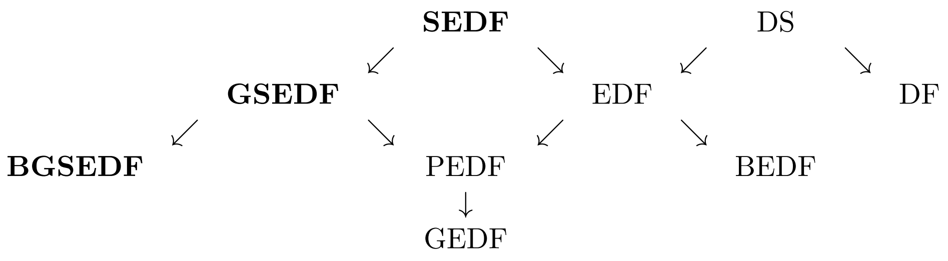

Figure 5: Difference families with $\lambda = 1$ that yield G-optimal strong AMD codes (indicated in boldface type)

Theorem 4.14. *There exists a k-uniform, strong $(m, n, \hat{\epsilon})$-AMD code with $\hat{\epsilon}^2 = \frac{m-1}{n-1} < 1$ if and only if $m = 2$ and $n = k^2 + 1$.*

Proof. Here we are considering k-uniform strong AMD codes that are simultaneously R-optimal and G-optimal. From the proof of Theorem 4.13, we see that $m = 2$ and $k(a - k) = n - 1$. Since $a = 2k$, we have $n = k^2 + 1$. Conversely, if $m = 2$ and $n = k^2 + 1$, then Example 2.2 shows the existence of a $(k^2 + 1, 2; k; 1)$-SEDF. This yields a strong AMD code with $\hat{\epsilon} = 1/k$, as desired. □

Figure 5 shows the types of difference families that yield G-optimal strong AMD codes. The relevant difference families are assumed to have $\lambda = 1$ in this figure.

5 Conclusion

We have studied weak and strong AMD codes that provide optimal protection against two specific adversarial substitution strategies. These codes are termed "R-optimal" and "G-optimal". We have considered various types of generalized difference families and determined when they yield R-optimal and/or G-optimal AMD codes. As well, we have proven in certain situations that R-optimal and/or G-optimal AMD codes imply the existence of the relevant difference families, thus providing a combinatorial characterization of the AMD codes under consideration.

It is an interesting open problem to construct additional examples of these generalized difference families. In particular, we ask if there are any examples of strong external difference families with $k > 1$ and $m > 2$. We are unaware of any such examples at the present time.

We should mention that there have been other bounds proven for AMD codes, in particular, for special classes of AMD codes, including *systematic* AMD codes. For example, [6, Proposition 6] uses a coding-theoretic approach to prove a necessary condition for the existence of a systematic AMD code. We focussed on bounds which can be met with equality if appropriate difference sets and difference families exist. We do not know if there are any natural connections between difference families and "optimal" systematic AMD codes.

References

[1] H. Ahmadi and R. Safavi-Naini. Detection of algebraic manipulation in the presence of leakage. *Lecture Notes in Computer Science* **8317** (2013), 238–258 (ICITS 2013).

[2] Y. Chang and C. Ding. Constructions of external difference families and disjoint difference families. *Designs, Codes and Cryptography* **40** (2006) 67–185.

[3] R. Cramer, Y. Dodis, S. Fehr, C. Padró and D. Wichs. Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. *Lecture Notes in Computer Science* **4965** (2008), 471–488 (Eurocrypt 2008).

[4] R. Cramer, Y. Dodis, S. Fehr, C. Padró and D. Wichs. Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. *Cryptology ePrint Archive: Report 2008/030.*

[5] R. Cramer, S. Fehr and C. Padró. Algebraic manipulation detection codes. *Science China Mathematics* **56** (2013), 1349–1358.

[6] R. Cramer. C. Padró and C. Xing. Optimal algebraic manipulation detection codes in the constant-error model. *Lecture Notes in Computer Science* **9014** (2015), 481–501 (TCC 2015).

[7] F. CuiLing, L. JianGuo and S. XiuLing. Constructions of optimal difference systems of sets. *Science China Mathematics* **54** (2011), 173–184.

[8] Y. Fujiwara K. Momihara and M. Yamada. Perfect difference systems of sets and Jacobi sums. *Discrete Mathematics* **309** (2009), 3954–3961.

[9] Y. Fujiwara and V.D. Tonchev. High-rate self-synchronizing codes. *IEEE Transactions on Information Theory* **59** (2013), 2328–2335.

[10] B. Huang and D. Wu. Cyclotomic constructions of external difference families and disjoint difference families. *Journal of Combinatorial Designs* **17** (2009), 333–341.

[11] J. Lei and C. Fan. Optimal difference systems of sets and partition-type cyclic difference packings. *Designs, Codes and Cryptography* **58** (2011), 135–153.

[12] V.I. Levenshtein. One method of constructing quasilinear codes providing synchronization in the presence of errors. *Problems of Information Transmission* **7** (1971), 215–222.

[13] W. Ogata and K. Kurosawa. Optimum secret sharing scheme against cheating. *Lecture Notes in Computer Science* **1070** (1996), 200–211 (EUROCRYPT '96).

[14] W. Ogata, K. Kurosawa, D.R. Stinson and H. Saido. New combinatorial designs and their applications to authentication codes and secret sharing schemes. *Discrete Mathematics* **279** (2004), 383–405.

[15] W. Ogata, K. Kurosawa and D.R. Stinson. Optimum secret sharing scheme secure against cheating. *SIAM Journal on Discrete Mathematics* **20** (2006), 79–95.

[16] G.J. Simmons. Authentication theory/coding theory. *Lecture Notes in Computer Science* **196** (1985), 411–431 (Proceedings of Crypto '84).

[17] M. Tompa and H. Woll. How to share a secret with cheaters. *Journal of Cryptology* **1** (1988), 133–138.

[18] V.D. Tonchev. Difference systems of sets and code synchronization. *Rendiconti del Seminario Matematico di Messina Series II* **9** (2003), 217–226.