• REVIEWS •

SCIENCE CHINA
Mathematics

COMBINATORICS
CODING THEORY
CRYPTOGRAPHY

July 2013 Vol. 56 No. 7: 1349–1358
doi: 10.1007/s11425-013-4654-5

Algebraic manipulation detection codes

CRAMER Ronald[1,2], FEHR Serge[1] & PADRÓ Carles[3,*]

[1]*Centrum voor Wiskunde en Informatica (CWI), P.O.Box* 94079*, NL-*1090 *GB, Amsterdam, The Netherlands;*
[2]*Mathematical Institute, Leiden University, PB* 9512, 2300 *RA Leiden, The Netherlands;*
[3]*Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University,* 637616*, Singapore*
Email: Ronald.Cramer@cwi.nl, Serge.Fehr@cwi.nl, carlespl@ntu.edu.sg

Received December 31, 2012; accepted March 17, 2013

Abstract Algebraic manipulation detection codes are a cryptographic primitive that was introduced by Cramer et al. (Eurocrypt 2008). It encompasses several methods that were previously used in cheater detection in secret sharing. Since its introduction, a number of additional applications have been found. This paper contains a detailed exposition of the known results about algebraic manipulation detection codes as well as some new results.

Keywords cryptography, keyless message authentication, algebraic manipulation

MSC(2010) 94A62, 94A60, 94B60

Citation: Cramer R, Fehr S, Padró C. Algebraic manipulation detection codes. Sci China Math, 2013, 56: 1349–1358, doi: 10.1007/s11425-013-4654-5

1 Introduction

Consider an abstract device $\Sigma(\mathcal{G})$ that can store a single element x from a fixed, publicly known finite abelian group $\mathcal{G}$. An adversary cannot obtain any information on the value stored in $\Sigma(\mathcal{G})$, but he can manipulate the stored data by adding some value $\delta \in \mathcal{G}$ of his choice. As a result, $\Sigma(\mathcal{G})$ stores the value $x + \delta$. This value depends only on the adversary's *a priori* knowledge of x. Such a tampering is called an *algebraic manipulation*.

One-time-pad encryption is an example of such a device. It hides the message perfectly but an adversary can add any string to it without being detected. Linear secret sharing provides another example. The reader is referred to [1] for a recent survey on secret sharing. In a linear secret sharing scheme, a secret value is distributed into shares among a set of participants. Only some qualified subsets of participants can recover the secret value form their shares. Both the secret and the shares are vectors over some finite field, and the secret is reconstructed by computing a linear map on the shares of a qualified set. A coalition of dishonest players may have no information about the secret value, but they can manipulate the output of the reconstruction process by providing forged shares. Because of the homomorphic properties of such schemes, they can control the difference between the shared secret s and the incorrect reconstructed value s'.

An *algebraic manipulation detection* (*AMD*) *code* encodes a source s into a value $x \in \mathcal{G}$ stored in $\Sigma(\mathcal{G})$ in such a way that any algebraic manipulation is detected with high probability, even if the adversary knows

*Corresponding author

the value of s. In contrast to standard message authentication codes, there is no secret key involved. Algebraic manipulation detection codes were introduced in [5] as an abstraction of several previously presented methods for cheating detection in linear secret sharing schemes [4, 15–18, 22]. Following the ideas of [7], [5] also showed the usefulness of AMD codes to the design of robust fuzzy extractors.

Several other applications of AMD codes were found subsequently. For instance, to unconditionally secure multiparty computation with dishonest majority [3], anonymous quantum communication [2], non-malleable codes [9], codes for computationally simple channels [11], and public key encryption against related key attacks [24].

This paper contains a detailed presentation of the known results about AMD codes, and also some new results: the bound in Theorem 5.4, and the construction strategies in Subsections 7.2 and 7.3.

2 Definitions and lower bounds

Definition 2.1. Let $\mathcal{S}$ be a set of size $m > 1$ and $\mathcal{G}$ a commutative group of order n. Consider a pair (E, D) formed by a probabilistic *encoding* map $E\colon \mathcal{S} \to \mathcal{G}$ and a deterministic *decoding* map $D : \mathcal{G} \to \mathcal{S} \cup \{\perp\}$ such that $D(E(s)) = s$ with probability 1 for every $s \in \mathcal{S}$.

• The pair (E, D) is an (m, n, ε)-*algebraic manipulation detection* (*AMD*) *code* if for every $s \in \mathcal{S}$ and for every $\delta \in \mathcal{G}$, the probability that $D(E(s) + \delta) \notin \{s, \perp\}$ is at most ε.

• The pair (E, D) is a *weak* (m, n, ε)-*algebraic manipulation detection code* if for every $\delta \in \mathcal{G}$ and for s sampled uniformly at random from $\mathcal{S}$ (independent of δ), the probability that $D(E(s) + \delta) \notin \{s, \perp\}$ is at most ε.

Trivially, since the bound ε on the probability that $D(E(s) + \delta) \notin \{s, \perp\}$ holds for every *fixed* choice of $\delta \in \mathcal{G}$, it also holds for δ chosen in a randomized way, as long as it is chosen *independently* of the randomness of the probabilistic encoding map E, and of the uniform choice of s in the case of a weak AMD code.

The following bounds on the size of codewords in AMD codes were proved by Ogata and Kurosawa [16] in the framework of cheating detection in secret sharing.

Theorem 2.2. *In every* (m, n, ε)*-AMD code, respectively weak* (m, n, ε)*-AMD code,*

$$n \geqslant \frac{m-1}{\varepsilon^2} + 1\,, \quad \textit{respectively} \quad n \geqslant \frac{m-1}{\varepsilon} + 1\,.$$

Proof. For every $s \in \mathcal{S}$, consider

$$\rho(s) = \frac{|\bigcup_{s' \neq s} D^{-1}(s')|}{n-1} = \frac{\sum_{s' \neq s} |D^{-1}(s')|}{n-1},$$

where $D^{-1}(s')$ naturally denotes the set of all $g \in \mathcal{G}$ with $D(g) = s'$. First note that for arbitrary but fixed $s \in \mathcal{S}$ and $E(s) \in \mathcal{G}$, and for $\delta \in \mathcal{G} \setminus \{0\}$ chosen uniformly at random, $E(s) + \delta$ is uniformly distributed over $\mathcal{G} \setminus \{E(s)\}$, and thus the probability that $D(E(s) + \delta) \notin \{s, \perp\}$ is $\rho(s)$.

The bound for weak AMD codes now follows from observing that $|D^{-1}(s')| \geqslant 1$ for every $s' \in \mathcal{S}$ and thus $\rho(s) \geqslant (m-1)/(n-1)$, and from noting that if this bound $(m-1)/(n-1)$ on the probability that $D(E(s) + \delta) \notin \{s, \perp\}$ holds for every fixed choice of s and $E(s)$, it also hold for randomized choices, so that $\varepsilon \geqslant (m-1)/(n-1)$.

The bound for ordinary (i.e., non-weak) AMD codes follows from observing that here $|D^{-1}(s')| \geqslant 1/\varepsilon$ for every $s' \in \mathcal{S}$, as can easily be seen, and from noting similarly to above that the resulting bound on the probability that $D(E(s) + \delta) \notin \{s, \perp\}$ also holds for a randomized choice of $E(s)$. □

The *tag length* $\log n - \log m$ of an (m, n, ε)-AMD code measures the number of redundant bits that are needed to encode the source. A natural optimization problem is to minimize the tag length for any given values of the source length $\log m$ and the adversary's success probability ε. More specifically, minimizing the tag length for families of AMD codes encoding arbitrarily long messages for a fixed value of ε.

3 Weak algebraic manipulation detection codes

In this section, we present the known constructions of weak AMD codes. All of them are *deterministic*, that is, the encoding map $E\colon \mathcal{S} \to \mathcal{G}$ is deterministic. Then, the decoding map $D : \mathcal{G} \to \mathcal{S} \cup \{\perp\}$ is determined by $D(E(s)) = s$ and $D(g) = \perp$ if $g \notin E(\mathcal{S})$. Deterministic weak AMD codes have been implicitly used in [4, 16, 18] to construct secret sharing schemes with cheating detection. In addition, the *robust codes* introduced by Karpovsky and Taubin [13] are equivalent to a special class of deterministic weak AMD codes (as specified later).

Deterministic weak AMD codes are closely related to difference sets, specifically to a slightly more general object that is introduced here. Recall that in an (n, m, t)-*difference set*, every nonzero element of $\mathcal{G}$ appears *exactly* t times in the list of differences.

Definition 3.1. Let $\mathcal{G}$ be a group of finite order n. A subset $V \subseteq \mathcal{G}$ of size m is an (n, m, t)-*bounded difference set* if the list of differences $(v - w)_{v,w\in V}$, contains every nonzero element of $\mathcal{G}$ at most t times.

We prove in the following that there is an equivalence between bounded difference sets and deterministic weak AMD codes.

Theorem 3.2. *An injective map $E\colon \mathcal{S} \to \mathcal{G}$ determines a deterministic weak (m, n, ε)-AMD code if and only if $E(\mathcal{S}) \subseteq \mathcal{G}$ is an (n, m, t)-bounded difference set with $t \leqslant \varepsilon m$.*

Proof. Consider $V = E(\mathcal{S})$ and let t be the maximum number of times that a nonzero element $\delta \in \mathcal{G}$ occurs in the list of differences $(v - w)_{v,w\in V}$. Clearly, if s is sampled uniformly at random from $\mathcal{S}$, then $\max_{\delta\in\mathcal{G}} \Pr(D(E(s) + \delta) \notin \{s, \perp\}) = t/m$. □

Example 3.3. Consider $\mathcal{G} = \mathbb{Z}_7$ and $V = \{0, 1, 3\} \subseteq \mathcal{G}$. Clearly $V \subseteq \mathcal{G}$ is a $(7, 3, 1)$-difference set. Therefore, if $|\mathcal{S}| = 3$, any injective map $E\colon \mathcal{S} \to \mathcal{G}$ with $E(\mathcal{S}) = V$ determines a weak $(3, 7, 1/3)$-AMD code. This weak AMD code is optimal because $n = 1 + (m - 1)/\varepsilon$, that is, the bound in Theorem 2.2 is attained.

Example 3.4. More generally, for ever prime power q, there exists an $(n, m, 1)$-difference set in $\mathcal{G} = \mathbb{Z}_n$ with $n = q^2 + q + 1$ and $m = q + 1$. Each one of these difference sets provides a weak $(m, n, 1/m)$-AMD code attaining the bound in Theorem 2.2. This fact was used by Ogata and Kurosawa [16] to present a family of optimal secret sharing schemes with cheating detection.

Example 3.5. Let $\mathbb{F}_q$ be a finite field with characteristic different from 2 and $\mathcal{G} = \mathbb{F}_q^2$. Then $V = \{(x, x^2) \in \mathcal{G} : x \in \mathbb{F}_q\}$ is a $(q^2, q, 1)$-bounded difference set. This implies the existence of a weak $(q, q^2, 1/q)$-AMD code for every odd prime power q. They were used in [18], also for cheating detection in secret sharing.

In the weak AMD codes in the previous examples, $\varepsilon = 1/|\mathcal{S}|$. A more flexible family of weakly secure AMD codes is obtained from the bounded difference sets in the following proposition, which generalizes the ones in Example 3.5.

Proposition 3.6. *Let q be a prime and consider integers $1 \leqslant r \leqslant k$ and a surjective $\mathbb{F}_q$-linear map $\phi\colon \mathbb{F}_{q^k} \to \mathbb{F}_{q^r}$. If q is odd, then*

$$V = \{(x, \phi(x^2)) : x \in \mathbb{F}_{q^k}\} \subseteq \mathbb{F}_{q^k} \times \mathbb{F}_{q^r}$$

is a (q^{k+r}, q^k, q^{k-r})-bounded difference set. If $q = 2$, then

$$V = \{(x, \phi(x^3)) : x \in \mathbb{F}_{2^k}\} \subseteq \mathbb{F}_{2^k} \times \mathbb{F}_{2^r}$$

is a $(2^{k+r}, 2^k, 2^{k-r+1})$-bounded difference set.

Proof. We have to check how many times a nonzero element $(\delta_1, \delta_2) \in \mathbb{F}_{q^k} \times \mathbb{F}_{q^r}$ appears in the list of differences of pairs of elements in V. If q is odd, $(x, \phi(x^2)) + (\delta_1, \delta_2) \in V$ if and only if $\phi(2x\delta_1 + \delta_1^2) = \delta_2$. If $\delta_1 = 0$, then this condition is only satisfied if $\delta_2 = 0$. If $\delta_1 \neq 0$, this is equivalent to $2x\delta_1 + \delta_1^2 \in \phi^{-1}(\delta_2)$. Since $|\phi^{-1}(\delta_2)| = q^{k-r}$, there are exactly q^{k-r} elements $x \in \mathbb{F}_{q^k}$ in this situation. If $q = 2$, then

$(x, \phi(x^3)) + (\delta_1, \delta_2) \in V$ if and only if $\phi(3x^2\delta_1 + 3x\delta_1^2 + \delta_1^3) = \delta_2$. For every $y \in \phi^{-1}(\delta_2)$, there are at most two elements $x \in \mathbb{F}_{2^k}$ such that $3x^2\delta_1 + 3x\delta_1^2 + \delta_1^3 = y$. □

The weak $(q^k, q^{k+r}, 1/q^r)$-AMD codes that are determined by these bounded difference sets when q is odd were implicitly used in [4] to construct almost optimal secret sharing schemes with cheater detection. The construction for $q = 2$ provides weak $(2^k, 2^{k+r}, 1/2^{r-1})$-AMD codes. Observe that arbitrarily long messages can be encoded for fixed values of ε. These deterministic weak AMD codes are *systematic*, that is, $\mathcal{S}$ is a group and the encoding map is of the form $E : \mathcal{S} \to \mathcal{S} \times \mathcal{G}'$ with $E(s) = (s, f(s))$ for some group $\mathcal{G}'$ and some deterministic map $f : \mathcal{S} \to \mathcal{G}'$. The bound in Theorem 2.2 cannot be attained by any systematic weak AMD code because $\varepsilon \geqslant 1/|\mathcal{G}'| = n/m$. Therefore, the above AMD codes for odd q are optimal in that they attain this bound for weak systematic AMD codes. Systematic weak AMD codes are equivalent to the *robust codes* introduced in [13]. In fact, the constructions of robust codes presented in [13] are also based on the bounded difference sets described in Proposition 3.6.

4 Algebraic manipulation detection codes and differential structures

We present in this section a characterization of AMD codes in terms of *differential structures*, a combinatorial object that was introduced in [5].

For a group $\mathcal{G}$ of order n and a set $\mathcal{S}$ of size m, let $(V_s)_{s\in\mathcal{S}}$ be a family of disjoint non-empty subsets of $\mathcal{G}$. We call $(\mathcal{G}, (V_s)_{s\in\mathcal{S}})$ a *differential structure*. The parameters of interest related to a differential structure are as follows. For any $s \in \mathcal{S}$, we write t_s for the maximal overlap between any translation of V_s and the union of the other sets. Namely,

$$t_s = \max_{\delta\in\mathcal{G}} \Big|(V_s + \delta) \cap \bigcup_{s'\neq s} V_{s'}\Big|.$$

Differential structures are closely related to AMD codes. An AMD code (E, D) is said to be with *uniform selection* if the encoding $E(s)$ is uniformly distributed over $D^{-1}(s) = \{g \in \mathcal{G} : D(g) = s\}$ for every $s \in \mathcal{S}$. Only AMD codes with uniform selection appear in the natural constructions we are aware of.

Let (E, D) be an (m, n, ε)-AMD code with uniform selection with $E : \mathcal{S} \to \mathcal{G}$ and take $V_s = D^{-1}(s)$ for every $s \in \mathcal{S}$. Then $(\mathcal{G}, (V_s)_{s\in\mathcal{S}})$ is a differential structure with $t_s \leqslant \varepsilon\,|V_s|$ for every $s \in \mathcal{S}$. Conversely, given a differential structure $(\mathcal{G}, (V_s)_{s\in\mathcal{S}})$, consider $E : \mathcal{S} \to \mathcal{G}$ in which $E(s)$ is taken uniformly at random from V_s and $D : \mathcal{G} \to \mathcal{S}$ is given by $D(g) = s$ if $g \in V_s$ and $D(g) = \perp$ if $g \notin \bigcup_{s\in\mathcal{S}} V_s$. Then (E, D) is an (m, n, ε)-AMD code with uniform selection, where $m = |\mathcal{S}|$, $n = |\mathcal{G}|$ and $\varepsilon = \max_{s\in\mathcal{S}} t_s/|V_s|$.

Example 4.1. Consider a finite field $\mathbb{F}_q$ and, for every $s \in \mathcal{S}$, the vector $v_s = (0, 1, s) \in \mathbb{F}_q^3$ and the line $V_s = \{(s, 0, 0) + \lambda v_s : \lambda \in \mathbb{F}_q\} \subseteq \mathbb{F}_q^3$. Then $(\mathbb{F}_q^3, (V_s)_{s\in\mathbb{F}_q})$ is a differential structure with $t_s = 1$ for every $s \in \mathbb{F}_q$. Indeed, for every $\delta = (\delta_1, \delta_2, \delta_3) \in \mathbb{F}_q^3$,

$$V_s + \delta = \{(s, 0, 0) + \delta + \lambda v_s : \lambda \in \mathbb{F}_q\}$$

is a line in $\mathbb{F}_q^3$ that lies on the plane $x = s + \delta_1$. Therefore, $(V_s + \delta) \cap V_{s'} = \emptyset$ if $s' \neq s + \delta_1$ and $|(V_s + \delta) \cap V_{s+\delta_1}| = 1$ if $\delta_1 \neq 0$ because the vectors v_s and $v_{s+\delta_1}$ are linearly independent. A $(q, q^3, 1/q)$-AMD code is obtained from this differential structure.

Another differential structure that may be useful for the construction of AMD codes is obtained from caps in projective spaces. A *cap* is a set of points in a projective space over a finite field such that no three of them are collinear. A survey on this topic can be found in [12, Section 4]. Equivalently, a cap is a set of nonzero vectors in a vector space over a finite field such that no three of them are linearly dependent.

Let $\mathcal{S} \subseteq \mathbb{F}_q^r$ be a cap and, for every $s \in \mathcal{S}$, consider

$$V_s = \{\lambda s : \lambda \in \mathbb{F}_q \setminus \{0\}\} \subseteq \mathbb{F}_q^r.$$

Then $(\mathbb{F}_q^r, (V_s)_{s\in\mathcal{S}})$ is a differential structure with $t_s = 1$ for every $s \in \mathcal{S}$. Indeed, if there exists $\delta \in \mathbb{F}_q^r$ such that

$$\left|(V_s + \delta) \cap \bigcup_{s'\neq s} V_{s'}\right| > 1,$$

then $\lambda s + \delta = \lambda_1 s_1$ and $\lambda' s + \delta = \lambda_2 s_2$, for some $\lambda, \lambda', \lambda_1, \lambda_2 \in \mathbb{F}_q \setminus \{0\}$ with $\lambda \neq \lambda'$ and $s_1, s_2 \in \mathcal{S} \setminus \{s\}$. Therefore, $(\lambda - \lambda')s = \lambda_1 s_1 - \lambda_2 s_2$. If $s_1 = s_2$, the set $\{s, s_1\}$ is linearly dependent, otherwise the set $\{s, s_1, s_2\}$ is linearly dependent, a contradiction in both cases.

5 Systematic algebraic manipulation detection codes

An AMD code is *systematic* if the source set $\mathcal{S}$ is a group and the encoding is of the form

$$E : \mathcal{S} \to \mathcal{G} = \mathcal{S} \times \mathcal{G}_1 \times \mathcal{G}_2, \quad s \mapsto (s, x, f(x, s))$$

for some function $f : \mathcal{G}_1 \times \mathcal{S} \to \mathcal{G}_2$, where $\mathcal{G}_1$ and $\mathcal{G}_2$ are groups and x is taken uniformly at random from $\mathcal{G}_1$. The decoding function of a systematic AMD code is naturally given by

$$D(s, x, e) = \begin{cases} s, & \text{if } e = f(x, s), \\ \perp, & \text{otherwise.} \end{cases}$$

Clearly, a systematic AMD code is with uniform selection. The underlying differential structure is given by the sets $V_s = \{(s, x, f(x, s)) : x \in \mathcal{G}_1\}$. In such differential structures,

$$t_s = \max_{\substack{\delta\in\mathcal{G}\\ s'\neq s}} |(V_s + \delta) \cap V_{s'}| = \max_{\substack{(\delta_1,\delta_2)\in\mathcal{G}_1\times\mathcal{G}_2\\ s'\neq s}} |\{x \in \mathcal{G}_1 : f(x, s) + \delta_2 = f(x + \delta_1, s')\}|. \tag{5.1}$$

For systematic AMD codes, we notate $n_1 = |\mathcal{G}_1|$ and $n_2 = |\mathcal{G}_2|$. That is, in a systematic (m, n, ε)-AMD code, $n = mn_1n_2$. We give first a quite obvious bound on the values of n_1, n_2 that excludes the possibility that the bound in Theorem 2.2 is attained by a systematic AMD code. The intuition behind this result is derived from the fact that an adversary can always try to guess the value of $x \in \mathcal{G}_1$ or the value of the difference $\delta_2 = f(x, s') - f(x, s) \in \mathcal{G}_2$.

Proposition 5.1. *In every systematic AMD code, $n_i \geqslant 1/\varepsilon$ for $i = 1, 2$. As a consequence, $n \geqslant m/\varepsilon^2$.*

Proof. Let $(G, (V_s)_{s\in\mathcal{S}})$ be the underlying differential structure. Since $|V_s| = n_1$ for every $s \in \mathcal{S}$, it is clear that $\varepsilon \geqslant 1/n_1$. Given $s, s' \in \mathcal{S}$ with $s \neq s'$ and $\delta_1 \in \mathcal{G}_1$, consider the mapping $\phi : \mathcal{G}_1 \to \mathcal{G}_2$ defined by $\phi(x) = f(x + \delta_1, s') - f(x, s)$. Since $|\phi^{-1}(\delta_2)| \geqslant n_1/n_2$ for some $\delta_2 \in \mathcal{G}_2$, we have that $t_s \geqslant n_1/n_2$ by Equation (5.1), and hence $\varepsilon \geqslant t_s/|V_s| \geqslant 1/n_2$. □

Example 5.2. A systematic $(q, q^3, 1/q)$-AMD code is obtained by taking $\mathcal{S} = \mathcal{G}_1 = \mathcal{G}_2 = \mathbb{F}_q$ and $f : \mathcal{G}_1 \times \mathcal{S} \to \mathcal{G}_2$ with $f(x, s) = xs$. This AMD code is equivalent to the one that is obtained from the differential structure in Example 4.1.

Example 5.3. Consider $\mathcal{S} = \mathbb{F}_q$ and $\mathcal{G}_1 = \mathcal{G}_2 = \mathbb{F}_q^r$. The function $f : \mathcal{G}_1 \times \mathcal{S} \to \mathcal{G}_2$ defined by $f((x_1, \ldots, x_r), s) = (sx_1, \ldots, sx_r)$ determines a systematic $(q, q^{2r+1}, 1/q^r)$-AMD code.

The AMD codes in the two previous examples were implicitly used in [4] in the framework of cheater detection in secret sharing. They satisfy $\log n - \log m = -2 \log \varepsilon$, and hence their tag length is optimal because the bound in Proposition 5.1 is attained. Nevertheless, $\varepsilon \leqslant 1/m$ is too small for the applications, in which we need to encode arbitrarily long source messages with a fixed adversary's success probability.

By the bound we present in Theorem 5.4, it is not possible to attain the bound in Proposition 5.1 when m is much larger than $1/\varepsilon$. This result is an adaptation of [23, Theorem 3.1], which corresponds to a different definition of AMD codes (see Subsection 7.4).

Theorem 5.4. *In every systematic $(m, mn_1n_2, \varepsilon)$-AMD code with $\varepsilon < 1$,*

$$\varepsilon \geqslant \frac{\log m}{n_1 \log n_2}.$$

Proof. Given a systematic $(m, mn_1n_2, \varepsilon)$-AMD code with $\varepsilon < 1$ defined by a function $f : \mathcal{G}_1 \times \mathcal{S} \to \mathcal{G}_2$, consider, for every pair $(\delta_2, s) \in \mathcal{G}_2 \times \mathcal{S}$, the function $G_{\delta_2,s} : \mathcal{G}_1 \to \mathcal{G}_2$ defined by $G_{\delta_2,s}(x) = f(x, s) + \delta_2$. We affirm that all these functions are different. Indeed, suppose that $G_{\delta_2,s} = G_{\delta_2',s'}$, that is, $f(x, s)+\delta_2 = f(x, s') + \delta_2'$ for every $x \in \mathcal{G}_1$. Obviously, $\delta_2 = \delta_2'$ if $s = s'$ and, by (5.1), $t_s = n_1$ if $s \neq s'$, a contradiction with $\varepsilon < 1$. Therefore,

$$C = \{(G_{\delta_2,s}(x))_{x\in\mathcal{G}_1} \,:\, (\delta_2, s) \in \mathcal{G}_2 \times \mathcal{S}\} \subseteq \mathcal{G}_2^{n_1}$$

is a code with mn_2 codewords of length n_1 over an alphabet of size n_2. The Hamming distance between the codewords corresponding to $G_{\delta_2,s}$ and $G_{\delta_2',s'}$ is obviously equal to n_1 if $s = s'$ and $\delta_2 \neq \delta_2'$, and it is at least $n_1 - t_s$ if $s \neq s'$. Therefore, $\varepsilon \geqslant (n_1 - d)/n_1$ if d is the minimum distance of C. Finally, $mn_2 \leqslant n_2^{n_1-d+1}$ by the Singleton bound. □

As a consequence of Proposition 5.1 and Theorem 5.4, a lower bound on the tag length of systematic AMD codes is obtained, and also an impossibility result.

Corollary 5.5. *In every systematic $(m, mn_1n_2, \varepsilon)$-AMD code with $\varepsilon < 1$,*

$$\log n_1 + \log n_2 \geqslant -2\log\varepsilon + \log\log m - \log\log n_2.$$

Proof. Take into account that $\log n_1 \geqslant -\log\varepsilon+\log\log m-\log\log n_2$ by Theorem 5.4 and that $\log n_2 \geqslant -\log\varepsilon$ by Proposition 5.1. □

Corollary 5.6. *There is no systematic $(m, mn_1n_2, \varepsilon)$-AMD code with $\varepsilon < 1$ if $\log m \geqslant n_1 \log n_2$.*

6 An almost optimal flexible construction

The following construction of a systematic AMD code makes it possible to encode arbitrarily long source messages for fixed values of ε. In addition, the bound in Theorem 5.4 is almost attained. This AMD code construction first appeared in an early draft of [5]. Concurrently and independently, the same construction was used as a tool in the context of robust fuzzy extractors in [7], though without making its abstract properties explicit.

For a finite field $\mathbb{F}_q$ with characteristic p and a positive integer r such that $r + 2$ is not divisible by p, consider $\mathcal{S} = \mathbb{F}_q^r$ and $\mathcal{G}_1 = \mathcal{G}_2 = \mathbb{F}_q$. We prove next that the function

$$f(x, s) = f(x, (s_1, \ldots, s_r)) = x^{r+2} + \sum_{i=1}^{r} s_i x^i \tag{6.1}$$

determines a systematic $(q^r, q^{r+2}, (r + 1)/q)$-AMD code. By (5.1), it is enough to prove that, for every $s, s' \in \mathbb{F}_q^r$ with $s \neq s'$ and for every $\delta_1, \delta_2 \in \mathbb{F}_q$, there exist at most $r + 1$ values $x \in \mathbb{F}_q$ such that $f(x, s) + \delta_2 = f(x + \delta_1, s')$. Observe that

$$f(x + \delta_1, s') = x^{r+2} + (r + 2)\delta_1 x^{r+1} + \sum_{i=1}^{r} s_i' x^i + \delta_1 h(x),$$

where h is a polynomial of degree at most r. Therefore,

$$f(x + \delta_1, s') - (f(x, s) + \delta_2) = (r + 2)\delta_1 x^{r+1} + \sum_{i=1}^{r}(s_i' - s_i)x^i + \delta_1 h(x) - \delta_2. \tag{6.2}$$

The left side of Equation (6.2) is a nonzero polynomial on x of degree at most $r + 1$. Indeed, this is obvious if $\delta_1 \neq 0$ and, if $\delta_1 = 0$, the left side simplifies to $\sum_{i=1}^{r}(s_i' - s_i)x^i - \delta_2$, which is nonzero because $s \neq s'$. Therefore, this polynomial has at most $r + 1$ roots and the proof is concluded.

The tag length of these AMD codes is $\log n_1 + \log n_2 = 2\log q = -2\log\varepsilon + 2\log(r+1)$, which is almost optimal according to Theorem 5.4. Moreover, this family is flexible in the sense that, given m_0 and ε_0, and a prime p, we can find an (m, n, ε)-AMD in that family with $m = p^{kr} \geqslant m_0$ and $\varepsilon = (r + 1)/p^k \leqslant \varepsilon_0$

such that the tag length $\log m - \log n$ is close to $-2\log\varepsilon_0$. Indeed, this is achieved by taking r equal to the smallest positive integer such that $r+2$ is not divisible by p and $\log m_0 \leqslant r(\log(r+1) - \log\varepsilon_0)$ and taking

$$k = \left\lceil \frac{\log(r+1) - \log\varepsilon_0}{\log p} \right\rceil.$$

Therefore,

$$\begin{aligned}\log n - \log m = 2k\log p &\leqslant -2\log\varepsilon_0 + 2\log(r+1) + 2\log p \\ &\leqslant -2\log\varepsilon_0 + 2\log\left(-\frac{\log m_0}{\log\varepsilon_0} + 3\right) + 2\log p,\end{aligned}$$

because $(k-1)\log p \leqslant \log(r-1) - \log\varepsilon_0$ and $(r-2)(\log(r-1) - \log\varepsilon_0) \leqslant \log m_0$.

7 Other constructions

We discuss here some strategies to find alternatives to the construction of AMD codes in Section 6.

7.1 From authentication codes

The first one, which was proposed in [5], consists of combining weak AMD codes and authentication codes [19–21]. Consider a systematic authentication code $A : \mathcal{S} \times \mathcal{K} \to \mathcal{T}$ such that $\mathcal{S}$ and $\mathcal{T}$ are groups. Such codes are used to authenticate a message $s \in S$ by adding to it the tag $\sigma = A(k, s)$, where k is taken uniformly at random from $\mathcal{K}$ and is known by the sender and the receiver. Let p_S be the success probability of the substitution attack, that is, the maximum over all $s \neq s'$ of the probability of successfully substituting the authenticated s by s'. We consider as well a systematic weak $(m, |\mathcal{K}|, \varepsilon_1)$-AMD code whose encoding map is of the form $E' : \mathcal{K} \to \mathcal{K} \times \mathcal{G}_3$ with $E'(k) = (k, g(k))$ for some group $\mathcal{G}_3$. Finally, take $\mathcal{G}_1 = \mathcal{K}$ and $\mathcal{G}_2 = \mathcal{G}_3 \times \mathcal{T}$. Then the map $E : \mathcal{S} \to \mathcal{S} \times \mathcal{G}_1 \times \mathcal{G}_2$ defined by $E(s) = (s, k, (g(k), A(k, s))$, where k is chosen uniformly at random from $\mathcal{K}$, determines a systematic $(m, mn_1n_2, \varepsilon)$-AMD code with $\varepsilon = \max\{\varepsilon_1, p_S\}$. We proceed now to estimate the tag length of these systematic AMD codes. Clearly, $p_S \geqslant 1/|\mathcal{T}|$ and, by [21, Theorem 5.2], $|\mathcal{K}| \geqslant |\mathcal{T}|^2$ if $p_S = 1/|\mathcal{T}|$. In addition, $\varepsilon \geqslant 1/|\mathcal{G}_3|$. Therefore, the tag length of such an AMD code is around $-4\log\varepsilon$.

In order to construct AMD codes with similar properties as the ones in Section 6, we should use authentication codes such that $\mathcal{S}$ is much larger that $\mathcal{K}$ and $\mathcal{T}$. A well-known family of such authentication codes is obtained by taking $\mathcal{S} = \mathbb{F}_q^r$, $\mathcal{K} = \mathbb{F}_q^2$ and $\mathcal{T} = \mathbb{F}_q$ together with

$$A[(s_1, \ldots, s_r), (x, b)] = b + \sum_{i=1}^{r} s_i x^i.$$

The substitution probability is $p_S = r/q$. By identifying $\mathbb{F}_q^2$ with $\mathbb{F}_{q^2}$, a systematic weak $(q^2, q^3, \varepsilon_1)$-AMD code $E' : \mathcal{K} \to \mathcal{K} \times \mathbb{F}_q$ with $\varepsilon_1 = 1/q$ if q is odd and $\varepsilon_1 = 2/q$ if q is even can be found in the family that is derived from Proposition 3.6. In conclusion, a systematic AMD code with $m = q^r$, $n_1 = n_2 = q^2$ and $\varepsilon = r/q$ is obtained. Clearly, this construction can not compete with the one in Section 6.

7.2 From error correcting codes

The bound for systematic AMD codes in Theorem 5.4 is proved from properties of error correcting codes that are derived from AMD codes. This connection also works in the other direction, that is, error correcting codes can be used to construct systematic AMD codes. This idea has been used by Karpovsky and Wang [14, 23], who introduced constructions of AMD codes from generalized Reed-Muller codes. Nevertheless, their constructions are difficult to analyze here because they consider a different definition of AMD code (see Subsection 7.4) and different optimization problems.

Consider groups $\mathcal{S}$, $\mathcal{G}_1$ and $\mathcal{G}_2$, a code $C \subseteq \mathcal{G}_2^{\mathcal{G}_1}$, and a surjective encoding map $\mathbf{c} : \mathcal{S} \to C$. The codewords in C are of the form $\mathbf{c}(s) = (\mathbf{c}_x(s))_{x \in \mathcal{G}_1}$, where $\mathbf{c}_x(s) \in \mathcal{G}_2$. A systematic AMD code is

obtained by considering the function $f : \mathcal{G}_1 \times \mathcal{S} \to \mathcal{G}_2$ with $f(x, s) = \mathbf{c}_x(s)$. In fact, this description applies to every systematic AMD code.

We next analyze the systematic AMD codes in Section 6 under this point of view. Clearly, those AMD codes are obtained from Reed-Solomon codes that have been modified in some way. Observe that the systematic AMD code associated with the Reed-Solomon code

$$C = \{(h(x))_{x\in\mathbb{F}_q} : h \in \mathbb{F}_q[X], \deg h \leqslant r\}$$

has $\varepsilon = 1$ because C contains constant codewords. We can remove them by considering the code

$$C' = \{(h(x))_{x\in\mathbb{F}_q} : h \in \mathbb{F}_q[X], h \text{ is monic}, \deg h = r, h(0) = 0\},$$

but the corresponding AMD code has $\varepsilon = 1$ as well. Indeed, given $h = X^r + \sum_{i=1}^r h_i X^i$ and $\delta_1 \in \mathbb{F}_q$, then $h(X + \delta_1) = X^r + \sum_{i=1}^r \widehat{h}_i X^i + b_0 = \widehat{h} + b_0$. This implies that $V_s + \delta = V'_s$ for some $\delta \in \mathcal{G}$ and $s, s' \in \mathcal{S}$ with $s \neq s'$. This problem appears because the Reed-Solomon code C is invariant under permutations $x \mapsto x + \delta_1$ of the indices. In particular, it is a cyclic code if q is prime. In order to avoid this problem, we can take the code

$$C'' = \{(h(x))_{x\in\mathbb{F}_q} : h \in \mathbb{F}_q[X], h \text{ is monic}, \deg h = r, h(0) = 0, h_{r-1} = 0\},$$

which determines an AMD code in the family described in Section 6. Because of the requirement $h_{r-1} = 0$, for every $h \in \mathbb{F}_q[X]$, at most one codeword in the orbit $\{(h(x + \delta_1))_{x\in\mathbb{F}_q} : \delta_1 \in \mathbb{F}_q\} \subseteq C$ is in C''.

A more general construction of systematic AMD codes is suggested by the previous discussion. Let $C \subseteq \mathcal{G}_1^{\mathcal{G}_2}$ be a code such that, for every $\mathbf{c} \in C$, the orbit $C_{\mathbf{c}} = \{(\mathbf{c}_{x+\delta_1})_{x\in G_1} : \delta_1 \in \mathcal{G}_1\}$ is contained in C. Observe that, if $\mathcal{G}_1$ is a cyclic group, this property is equivalent to C being a cyclic code. We introduce a special distance in C. Namely, for every two words $\mathbf{c}, \mathbf{c}' \in C$, consider $\Delta(\mathbf{c}, \mathbf{c}') = \min_{\delta_2\in\mathcal{G}_2} d(\mathbf{c} + (\delta_2, \ldots, \delta_2), \mathbf{c}')$, where d is the Hamming distance. We consider as well the distance between orbits $\Delta(C_{\mathbf{c}}, C_{\mathbf{c}'})$ defined in the obvious way. Finally, consider $C' \subseteq C$ such that $|C' \cap C_{\mathbf{c}}| \leqslant 1$ for every $\mathbf{c} \in C$ and take $\Delta(C') = \min\{\Delta(C_{\mathbf{c}}, C_{\mathbf{c}'}) : \mathbf{c}, \mathbf{c}' \in C', C_{\mathbf{c}} \neq C_{\mathbf{c}'}\}$. Then it is clear that a systematic AMD code with $m = |C'|$, $n_i = |\mathcal{G}_i|$ for $i = 1, 2$, and $\varepsilon = (n_1 - \Delta(C'))/n_1$ is obtained from C'. Intuitively, constructions of optimal or almost optimal systematic AMD codes should be obtained from suitable families of linear cyclic codes.

7.3 From caps in projective spaces

The bound in Theorem 5.4 applies only to systematic AMD codes. Therefore, the existence of non-systematic AMD codes with optimal tag length $\log n - \log m = -2\log\varepsilon$ when ε is fixed and m is arbitrarily large is still an open problem.

A possible strategy to find such AMD codes is to find suitable differential structures. We analyze here the possibilities of the differential structures from caps in projective spaces that are described in Section 4. From the discussion there, a cap $\mathcal{S} \subseteq \mathbb{F}_q^r$ with $|\mathcal{S}| = m$ determines an $(m, q^r, 1/(q-1))$-AMD code. For our problem, we need to find large caps in $\mathbb{F}_q^r$ for a fixed q and arbitrarily large values of r. Let $m(r, q)$ be the maximum size of a cap in $\mathbb{F}_q^r$. The best upper and lower bounds on $m(r, q)$ that were known in 2001 are surveyed in [12, Section 4]. For instance, $m(r, q) \leqslant q^{r-2} - q^{r-3} + 8q^{r-4} + 15q^{r-5} + 1$ if $r \geqslant 6$, and $q > 7$ and q odd [12, Table 4.4(i)]. A similar upper bound is given for q even in the same table. These upper bounds do not exclude the existence of AMD codes with optimal tag length. Nevertheless, it is not known if these upper bounds are attained, and the known lower bounds [12, Table 4.6(iii)] do not solve our problem.

7.4 Stronger algebraic manipulation detection codes

An alternative definition of AMD code, with a stronger requirement, was proposed in [14, 23]. Namely, it is required that, for every $s \in \mathcal{S}$ and for every $\delta \in \mathcal{G} \setminus \{0\}$, the probability that $D(E(s) + \delta) \neq \perp$ is at most ε. That is, every undetected algebraic manipulation of the encoding $E(s)$ is considered an

adversarial success, even if the source message s is not altered. These stronger AMD codes are useful in the construction of *non-malleable secret sharing schemes* [10, Definition 1], and other applications are mentioned in [14].

Observe that the AMD code in Example 5.2 does not satisfy that stronger requirement. Indeed, the algebraic manipulation $E(s) + \delta = (s, x, xs) + (0, \delta_1, s\delta_1)$ is not detected with probability 1. The same applies to the $(m, q^r, 1/(q-1))$-AMD codes that are obtained from caps in projective spaces (see Subsection 7.3), because the algebraic manipulation $E(s)+\delta = \lambda s+\mu s$ is detected if and only if $\lambda+\mu = 0$. On the other hand, it is not difficult to check that the AMD codes described in Section 6 satisfy the stronger requirement for the same value of ε. Other constructions of stronger AMD codes are given in [14, 23].

Acknowledgements The third author's work was supported by the Singapore National Research Foundation (Grant No. NRF-CRP2-2007-03).

References

1 Beimel A. Secret-sharing schemes: A survey on coding and cryptology. In: Third International Workshop, IWCC 2011, Lecture Notes in Computer Science, vol. 6639. Berlin: Springer, 2011, 11–46
2 Brassard G, Broadbent A, Fitzsimons J, et al. Anonymous quantum communication. In: Advances in Cryptology, Asiacrypt 2007, Lecture Notes in Computer Science, vol. 4833. Berlin: Springer, 2007, 460–473
3 Broadbent A, Tapp A. Information-theoretic security without an honest majority. In: Advances in Cryptology, Asiacrypt 2007, Lecture Notes in Computer Science, vol. 4833. Berlin: Springer, 2007, 410–426
4 Cabello S, Padró C, Sáez G. Secret sharing schemes with detection of cheaters for a general access structure. Des Codes Cryptogr, 2002, 25: 175–188
5 Cramer R, Dodis Y, Fehr S, et al. Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. In: Advances in Cryptology, Eurocrypt 2008, Lecture Notes in Computer Science, vol. 4965. Berlin: Springer, 2008, 471–488
6 Dodis Y, Kanukurthi B, Katz J, et al. Robust fuzzy extractors and authenticated key agreement from close secrets. IEEE Trans Inform Theory, 2012, 58: 6207–6222
7 Dodis Y, Katz J, Reyzin L, et al. Robust fuzzy extractors and authenticated key agreement from close secrets. In: Advances in Cryptology, Crypto 2006, Lecture Notes in Computer Science, vol. 4117. Berlin: Springer, 2006, 232–250
8 Dodis Y, Ostrovsky R, Reyzin L, et al. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. SIAM J Comput, 2008 38: 97–139
9 Dziembowski S, Pietrzak K, Wichs D. Non-malleable codes. In: Innovations in Computer Science, ICS 2010. Beijing: Tsinghua University Press, 2010, 434–452
10 Gordon S D, Ishai Y, Moran T, et al. On complete primitives for fairness. In: Theory of Cryptography, TCC 2010, Lecture Notes in Computer Science, vol. 5978. Berlin: Springer, 2010, 91–108
11 Guruswami V, Smith A. Codes for computationally simple channels: Explicit constructions with optimal rate. In: 51st Annal IEEE Symposium. Philadelphia: IEEE, 2010, 723–732
12 Hirschfeld J W P, Storme L. The packing problem in statistics, coding theory, and finite projective spaces: update 2001. In: Finite Geometries. Proceedings of the Fourth Isle of Thorns Conference, Developments in Mathematics 3. Boston: Kluwer Academic Publishers, 2000, 201–246
13 Karpovsky M, Taubin A. New class of nonlinear systematic error detecting codes. IEEE Trans Inform Theory, 2004, 50: 1818–1820
14 Karpovsky M, Wang Z. Algebraic manipulation detection codes and their applications for design of secure communication or computation channels. Http://mark.bu.edu/papers/226.pdf
15 Obana S, Araki T. Almost optimum secret sharing schemes secure against cheating for arbitrary secret distribution. In: Advances in Cryptology, Asiacrypt 2006. Lecture Notes in Computer Science, vol. 4284. Berlin: Springer, 2006, 364–379
16 Ogata W, Kurosawa K. Optimum secret sharing scheme secure against cheating. In: Advances in Cryptology, Eurocrypt'96, Lecture Notes in Computer Science, vol. 1070. Berlin: Springer, 1996, 200–211
17 Ogata W, Kurosawa K, Stinson D R, et al. New combinatorial designs and their applications to authentication codes and secret sharing schemes. Discrete Math, 2004, 279: 383–405
18 Padró C, Sáez G, Villar J L. Detection of cheaters in vector space secret sharing schemes. Des Codes Cryptogr, 1999, 16: 75–85
19 Simmons G J. Authentication theory/Coding theory. In: Advances in Cryptology, Crypto'84, Lecture Notes in Com-

puter Science, vol. 196. Berlin: Springer, 1985, 411–431
20 Stinson D R. Some constructions and bounds for authentication codes. J Cryptology, 1988, 1: 37–51
21 Stinson D R. The combinatorics of authentication and secrecy codes. J Cryptology, 1990, 2: 23–49
22 Tompa M, Woll H. How to share a secret with cheaters. J Cryptology, 1988, 1: 133–138
23 Wang Z, Karpovsky M. Algebraic manipulation detection codes and their applications for design of secure cryptographic devices. In: IEEE 17th International On-Line Testing Symposium. Philadelphia: IEEE, 2011, 234–239
24 Wee H. Public key encryption against related key attacks. In: Public Key Cryptography, PKC 2012, Lecture Notes in Computer Science, vol. 7293. Berlin: Springer, 2012, 262–279