Splitting authentication codes with perfect secrecy: new results, constructions and connections with algebraic manipulation detection codes

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

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

April 22, 2021

Abstract

A splitting BIBD is a type of combinatorial design that can be used to construct splitting authentication codes with good properties. In this paper we show that a design-theoretic approach is useful in the analysis of more general splitting authentication codes. Motivated by the study of algebraic manipulation detection (AMD) codes, we define the concept of a *group generated* splitting authentication code. We show that all group-generated authentication codes have perfect secrecy, which allows us to demonstrate that algebraic manipulation detection codes can be considered to be a special case of an authentication code with perfect secrecy.

We also investigate splitting BIBDs that can be "equitably ordered". These splitting BIBDs yield authentication codes with splitting that also have perfect secrecy. We show that, while group generated BIBDs are inherently equitably ordered, the concept is applicable to more general splitting BIBDs. For various pairs (k, c), we determine necessary and sufficient (or almost sufficient) conditions for the existence of $(v, k \times c, 1)$-splitting BIBDs that can be equitably ordered. The pairs for which we can solve this problem are $(k, c) = (3, 2), (4, 2), (3, 3)$ and $(3, 4)$, as well as all cases with $k = 2$.

1 Introduction

The use of authentication codes for providing authentication in an unconditionally secure setting has long been studied, following models developed by Simmons [12]. Authentication codes with perfect secrecy ensure confidentiality of sources as well as authenticity. There is a considerable literature on authentication and secrecy codes, including models that make different assumptions about the distribution of the sources [14, 15]. We observe that the majority of the focus has been on the case where, for a given key, there is a unique encoding for each source. On the other hand, *Splitting authentication codes* allow multiple different encodings of a source under a specific key. Allowing splitting can facilitate better performance for certain parameter settings, and can also yield constructions that work for any source distribution. There is also a wide literature on splitting authentication codes, including many constructions [1, 7, 8, 9, 13, 16, 17]. However,

*D.R. Stinson's research is supported by NSERC discovery grant RGPIN-03882

the case of splitting authentication codes with perfect secrecy has not been systematically considered.

Our investigation of splitting authentication codes with perfect secrecy is motivated by consideration of the properties and structure of *algebraic manipulation detection (AMD) codes* with a view to better characterising those application contexts in which they can be usefully applied. AMD codes were introduced by Cramer, Dodis, Fehr, Padró and Wichs in EUROCRYPT 2008 as a way of abstracting ideas used in the construction of robust secret sharing schemes into more general tools for providing robustness against active manipulation in cryptographic systems [5]. The definitions of these objects have certain similarities with authentication codes, in that both aim to detect whether an adversary has tampered with an encoded element. Connections noted in the literature include the use of AMD codes by Cramer et al. in the construction of a primitive they call a *KMS-MAC*, which could be viewed as a variant of an authentication code [5]. However, there are also clear differences in the two definitions. For example, authentication codes rely on the use of a shared key, whereas there are no keys involved in the definition of an AMD code. Also, the underlying context for their use and the corresponding security definitions are different. The definition of an authentication code is purely combinatorial, as are many of the known constructions, whereas an AMD code inherently requires the algebraic structure of an abelian group.

In Section 2 of this paper we connect the combinatorial and algebraic perspectives by taking a design-theoretic approach to studying splitting authentication codes, with a particular focus on their automorphism groups. We introduce the notion of a *group generated authentication code,* and show that the property of being group generated is sufficient to ensure the authentication code has perfect secrecy, and it also gives other desirable properties such as optimal protection against impersonation attacks. We clarify the relationship between authentication codes and AMD codes by demonstrating that, in terms of their mathematical structure, an AMD code is a special case of a group generated authentication code, with weak AMD codes corresponding to authentication codes that require a uniform distribution on the sources and strong AMD codes corresponding to authentication codes that work for any source distribution. We discuss the consequences of this connection for our understanding of AMD codes.

In Section 3.1 we consider perfect secrecy for certain optimal authentication codes that are not necessarily group generated. Splitting BIBDs are a type of combinatorial design that give rise to splitting authentication codes that are optimal with respect to certain bounds on the adversary's success probability in substitution attacks. In Section 3.1 we define the *equitable ordering* property for splitting BIBDs, which guarantees that the corresponding authentication codes offer perfect secrecy. We give techniques to provide equitable ordering for splitting BIBDs with a range of parameters, which permits the conversion of a wide class of splitting authentication codes into splitting authentication codes with perfect secrecy.

1.1 Definitions

1.2 Authentication Codes

An *authentication code* consists of a 4-tuple $(\mathcal{S}, \mathcal{T}, \mathcal{K}, \mathcal{E})$ where $\mathcal{S}$ is a finite set of *sources*, the set $\mathcal{T}$ is a finite set of *messages*, the set $\mathcal{K}$ is a finite set of *keys* and $\mathcal{E}$ is a set of *encoding rules*. The encoding rules are (possibly randomised) maps from $\mathcal{S}$ to $\mathcal{T}$ that are indexed by the keys in $\mathcal{K}$. We use the notation $e_k(s) \subset \mathcal{T}$ to denote the set of possible encodings of source s under the encoding rule e_k. Note that, for distinct sources s and s', we require $e_k(s) \cap e_k(s') = \emptyset$ for each $k \in \mathcal{K}$; in practical terms, this means that knowledge of k enables the unique identification of the source from the encoding. We assume that the keys are drawn uniformly at random from $\mathcal{K}$, independently of s. A sender who shares a key $k \in \mathcal{K}$ with a receiver authenticates a source

value $s \in \mathcal{S}$ by calculating a message $t \in e_k(s)$ and transmitting it to the receiver. The receiver accepts the message as authentic if $t \in e_k(s)$.

Example 1.1. Let $\mathcal{S} = \{0, 1\}$ and $\mathcal{K} = \{0, 1, 2, 3, 4\}$. We can define an authentication code by means of the following table. To generate an encoding for source s and key k, we choose one of the two entries in the corresponding row/column uniformly at random.

k	$e_k(0)$	$e_k(1)$
0	$\{1, 4\}$	$\{2, 3\}$
1	$\{2, 0\}$	$\{3, 4\}$
2	$\{3, 1\}$	$\{4, 0\}$
3	$\{4, 2\}$	$\{0, 1\}$
4	$\{0, 3\}$	$\{1, 2\}$

If a receiver possesses the key 3, for example, then they would accept the message 2 as being an authentic encoding of the source 0. However, if they received the message 3 they would reject this as being inauthentic.

If $|e_k(s)| = 1$ for all $k \in \mathcal{K}$, $s \in \mathcal{S}$, then the authentication code is *deterministic*, otherwise it is said to be a *splitting authentication code*. If $|e_k(s)| = c$ for all $k \in \mathcal{K}$, $s \in \mathcal{S}$, then we say the authentication code is *c-splitting*. In the case where the encoding of each source s under any encoding rule e_k is chosen uniformly from the messages in $e_k(s)$ the authentication code is said to have *equiprobable encoding*. For instance, the authentication code described in Example 1.1 is a 2-splitting authentication code with equiprobable encoding. For all authentication codes considered in this paper, we assume we have equiprobable encoding.

There are several relevant probability distributions associated with an authentication code. There is the distribution on the sources; in some circumstances we consider the case where this distribution is uniform, although we also consider authentication codes with arbitrary source distributions. There is the distribution on the keys, which is generally assumed to be uniform and independent of the source distribution. Additionally there is the distribution associated with encoding of a source s under a key k, which we assume is uniform. Finally, there is the resulting distribution induced on the space of messages. For a key k, source s and message t, the probability that the message t results from encoding source s under key k can be expressed as

$$\begin{aligned} \Pr(k, s, t) &= \Pr(s) \Pr(k) \Pr(t|k, s), \\ &= \frac{\Pr(s)}{|\mathcal{K}||e_k(s)|}. \end{aligned}$$

For a c-splitting authentication code this becomes

$$\Pr(k, s, t) = \frac{\Pr(s)}{|\mathcal{K}|c}.$$

An adversary who has seen a valid message $t \in e_k(s)$ can try and trick the receiver into accepting as valid a different message t'. This attack is known as *substitution*, and it succeeds if $t' \in e_k(s')$ for some source $s' \neq s$. It is desirable to construct authentication codes for which the probability of a successful substitution attack is as small as possible. We assume that the adversary is aware of the distribution from which the source is drawn, and in response they choose a *substitution strategy* σ that consists of a choice of replacement message $\sigma(t)$ for each possible message $t \in \mathcal{T}$.

Let σ be a substitution strategy for attacking an authentication code $(\mathcal{S}, \mathcal{T}, \mathcal{K}, \mathcal{E})$. If the key is $k \in \mathcal{K}$ and the source is $s \in \mathcal{S}$, then the adversary's strategy succeeds whenever the message is a value t from the set

$$X_{k,s}^{\sigma} = \{t \in e_k(s) : \sigma(t) \in e_k(s') \text{ for some } s' \neq s\}.$$

The overall success probability ϵ_σ of the strategy σ is given by

$$\epsilon_\sigma = \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \sum_{t \in X_{k,s}^{\sigma}} \Pr(k, s, t). \tag{1}$$

The authentication code is said to have *substitution probability* at most ϵ if $e_\sigma \leq \epsilon$ for every strategy σ. We observe that the expression in (1) can be written as follows:

$$\epsilon_\sigma = \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \frac{|X_{k,s}^{\sigma}| \Pr(s)}{|\mathcal{K}||e_k(s)|}. \tag{2}$$

In [1] it was shown that the substitution probability ϵ is at least

$$\min_{k \in \mathcal{K}} \frac{\left|\bigcup_{s \in \mathcal{S}} e_k(s)\right| - \max_{s \in \mathcal{S}} |e_k(s)|}{|\mathcal{T}| - 1}. \tag{3}$$

(This was a correction of a result from [13].) An authentication code for which this bound is satisfied is said to have *optimal substitution probability.*

Example 1.2. Consider the authentication code of Example 1.1. As this is a 2-splitting authentication code with 5 keys and equiprobable encoding, the success probability of a substitution strategy σ is given by

$$\epsilon_\sigma = \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} \frac{|X_{k,s}^{\sigma}| \Pr(s)}{10}.$$

We first observe that for any $t \in \mathcal{T}$, if $\sigma(t) = t$ then it is the case that $t \notin X_{k,s}^{\sigma}$ for any choice of k or s. Consider now the element $0 \in \mathcal{T}$. If $\sigma(0) = 1$, then $0 \in X_{4,0}^{\sigma}$ and $0 \in X_{2,1}^{\sigma}$ but $0 \notin X_{k,s}^{\sigma}$ for any other choice of k and s. Similarly, for any other nonzero choice of $\sigma(0)$, we can check that there is one value of k with $0 \in X_{k,0}^{\sigma}$ and one value of k with $0 \in X_{k,1}^{\sigma}$. The same holds true for every other element t of $\mathcal{T}$: if $\sigma(t) \neq t$ then $t \in X_{k,0}^{\sigma}$ for precisely one value of k, and $t \in X_{k,1}^{\sigma}$ for precisely one value of k. Thus for any strategy σ it is the case that $\sum_{k \in \mathcal{K}} |X_{k,0}^{\sigma}| \leq 5$, and also $\sum_{k \in \mathcal{K}} |X_{k,1}^{\sigma}| \leq 5$. Hence we have

$$\begin{aligned}
\epsilon_\sigma &= \frac{1}{10} \sum_{s \in \mathcal{S}} \sum_{k \in \mathcal{K}} \Pr(s) |X_{k,s}^{\sigma}|, \\
&= \frac{1}{10} \sum_{k \in \mathcal{K}} \left(\Pr(0) |X_{k,0}^{\sigma}| + \Pr(1) |X_{k,1}^{\sigma}| \right), \\
&\leq \frac{1}{10} \left(5 \Pr(0) + 5 \Pr(1) \right), \\
&= \frac{1}{2}.
\end{aligned}$$

Hence $\epsilon_\sigma \leq 1/2$ for any σ, and we note further that $\epsilon_\sigma = 1/2$ for any strategy σ that satisfies $\sigma(t) \neq t$ for all $t \in \mathcal{T}$. This holds true for any source distribution.

If we consider (3) for this authentication code we have

$$\epsilon \geq \frac{4-2}{5-1} = \frac{1}{2},$$

and so this authentication code has optimal substitution probability.

Another attack considered in the literature is that of *impersonation*, in which an adversary who has not seen any transmitted messages sends a message to the receiver in the hopes that it will be accepted as valid. The probability that an adversary sending message t succeeds is given by

$$\frac{\left|\{k \in \mathcal{K} : t \in \bigcup_{s\in\mathcal{S}} e_k(s)\}\right|}{|\mathcal{K}|}, \tag{4}$$

and the *impersonation probability* of the authentication code is the maximum over all t of these success probabilities. Simmons observed in [12] that the impersonation probability of an authentication code is at least

$$\min_{k\in\mathcal{K}} \frac{\left|\bigcup_{s\in\mathcal{S}} e_k(s)\right|}{|\mathcal{T}|}. \tag{5}$$

An authentication code that meets this bound is said to have *optimal impersonation probability.*

Example 1.3. For the authentication code of Example 1.1 we observe that $|\{k \in \mathcal{K} : t \in \bigcup_{s\in\mathcal{S}} e_k(s)\}| = 4$ for any choice of t, hence the impersonation probability is $4/5$. This is in fact optimal, as $|\bigcup_{s\in\mathcal{S}} e_k(s)| = 4$, so the expression in (5) also evaluates to $4/5$.

Definition 1.1. An authentication code $(\mathcal{S}, \mathcal{T}, \mathcal{K}, \mathcal{E})$ has *perfect secrecy* if the message t reveals no information about the source s, that is if

$$\Pr(s \mid t) = \Pr(s),$$

for all $t \in \mathcal{T}$ and $s \in \mathcal{S}$.

While generalisations of these notions where the adversary sees more than one message have been considered in the literature, e.g. [15], in this paper we restrict our attention to the case where the adversary sees a single message.

1.3 AMD codes

An *algebraic manipulation detection code* (AMD code) is a 4-tuple $(\mathcal{S}, \mathcal{G}, A, E)$, where $\mathcal{S}$ is a finite set of *sources*, $\mathcal{G}$ is a finite additive group, $A \subset \mathcal{G}$ is a set of *valid encodings* and $E\colon \mathcal{S} \to A$ is a (possibly randomised) encoding rule [10]. We use the notation $A(s) \subset \mathcal{G}$ to denote the set of valid encodings of source $s \in S$, and we require $A(s) \cap A(s') = \emptyset$ whenever $s \neq s'$. We have $A = \cup_{s\in\mathcal{S}} A(s)$, and we will often use the notation $\mathcal{A}$ to denote the collection of disjoint subsets of $\mathcal{G}$ given by $\{A(s) : s \in \mathcal{S}\}$. We set $a_s = |A(s)|$ and $a = |A| = \sum_{s\in\mathcal{S}} a_s$.

A user selects a source $s \in \mathcal{S}$ randomly according to a distribution that is known to the adversary then the encoding rule E is used to encode s as an element $g \in A(s)$. If g is chosen uniformly at random from $A(s)$, then the AMD code is said to have *equiprobable encoding.* Throughout this paper we assume all AMD codes we consider have equiprobable encoding.

Example 1.4. Let $\mathcal{S} = \{0, 1\}$, let $\mathcal{G} = \mathbb{Z}_9$, and let $\mathcal{A} = \{\{0, 1\}, \{2, 4\}\}$, so $A = \{0, 1, 2, 4\}$. We can construct an AMD code $(\mathcal{S}, \mathcal{G}, A, E)$ by defining an encoding rule E that encodes the source 0 as either 0 or 1, each with probability $1/2$, and encodes the source 1 as either 2 or 4, each with probability $1/2$. We typically refer to $\mathcal{A}$ as an AMD code, since E is implied once we assume equiprobable encodings.

An adversary selects an element $\Delta \in \mathcal{G}$ to be added to g. The user accepts $g + \Delta$ if it is a valid encoding of some source, that is, if $g + \Delta \in A(s')$ for some $s' \in \mathcal{S}$, in which case it is decoded to s'. The adversary wins if $s' \neq s$, that is if their algebraic manipulation has succeeded in causing the user to decode the stored value incorrectly. Given a source $s \in \mathcal{S}$ and an element $\Delta \in \mathcal{G}$, define the set X_s^Δ to be

$$X_s^\Delta = \{g \in A(s) : g + \Delta \in A(s') \text{ for some } s' \neq s\}.$$

Then the probability that an adversary who chooses Δ succeeds is

$$\epsilon_\Delta = \sum_{s \in \mathcal{S}} \sum_{g \in X_s^\Delta} \Pr(s, g).$$

We observe that $\Pr(s, g) = \Pr(s) \Pr(g \mid s)$, and that this is equal to $\Pr(s)|A(s)|^{-1}$ as we have equiprobable encodings. This allows us to express ϵ_Δ as

$$\epsilon_\Delta = \sum_{s \in \mathcal{S}} \frac{|X_s^\Delta| \Pr(s)}{|A(s)|}.$$

Definition 1.2. An AMD code with $|\mathcal{S}| = m$ and $|\mathcal{G}| = n$ is referred to as a *weak* (m, n, ϵ)*-AMD code* if an adversary who does not know the source has success probability at most ϵ in the case where the sources are uniformly distributed. Here we have

$$\epsilon_\Delta = \sum_{s \in \mathcal{S}} \frac{|X_s^\Delta|}{m|A(s)|},$$

and we require $\epsilon_\Delta \leq \epsilon$ for all $\Delta \in \mathcal{G}^*$.

Example 1.5. Consider the AMD code of Example 1.4, and suppose an adversary chooses $\Delta = 1$. Then $X_0^1 = \{1\}$ and $X_1^1 = \emptyset$, so $\epsilon_1 = \frac{1}{4}$. Similar calculations show that in fact $\epsilon_\Delta = 1/4$ for each $\Delta \in \mathbb{Z}_9^*$, and so this is a weak $(2, 9, 1/4)$-AMD code.

Definition 1.3. An AMD code with $|\mathcal{S}| = m$ and $|\mathcal{G}| = n$ is a *strong* (m, n, ϵ)*-AMD code* if the success probability of an adversary who knows the source is at most ϵ. Let $\epsilon_{s,\Delta}$ be the success probability of an adversary who selects the element Δ, conditioned on the event that the source is s. Then

$$\epsilon_{s,\Delta} = \frac{|X_S^\Delta|}{|A(s)|},$$

and we require $\epsilon_{s,\Delta} \leq \epsilon$ for all $s \in \mathcal{S}$ and $\Delta \in \mathcal{G}^*$.

Example 1.6. The AMD code of Example 1.4 has $\left|X_s^\Delta\right| \leq 1$ for each $s \in \mathcal{S}$ and $\Delta \in \mathcal{G}^*$, which implies $\epsilon_{s,\Delta} \leq 1/2$. Thus it is a strong $(2, 9, 1/2)$-AMD code.

An (m, n, ϵ)-AMD code is said to be *c-regular* if it has equiprobable encoding and $a_s = c$ for all $s \in \mathcal{S}$. A 1-regular AMD code is *deterministic*. We observe that a deterministic AMD cannot be a strong (m, n, ϵ)-AMD code for any $\epsilon < 1$, since an adversary who knows the source and knows the encoding of the source (due to the fact the encoding is deterministic) has enough information to pick a value of Δ that will succeed.

We observe that, while it would also be possible to study AMD codes with a specified distribution on the sources that is not the uniform distribution, this has not been considered in the literature. However, this notion does lead naturally to an alternative interpretation of strong AMD codes: rather than assume a model where the adversary knows the value of the source, we could instead view strong AMD codes as being ones that work for *any* source distribution, as demonstrated in the following theorem:

Theorem 1.1. *An AMD code $(\mathcal{S}, \mathcal{G}, \mathcal{A}, E)$ is a strong (m, n, ϵ)-AMD code if and only if the success probability of an adversary is at most ϵ for any choice of source distribution.*

Proof. Suppose an adversary's success probability against an AMD code $(\mathcal{S}, \mathcal{G}, \mathcal{A}, E)$ is at most ϵ for any source distribution. Then the adversary's success probability is at most ϵ for the distribution in which source s is chosen with probability 1, for any $s \in S$. Hence it is a strong (m, n, ϵ)-AMD code.

Conversely, suppose $(\mathcal{S}, \mathcal{G}, \mathcal{A}, E)$ is a strong (m, n, ϵ)-AMD code. Then for all $s \in \mathcal{S}$ and $\Delta \in \mathcal{G}$ we have

$$\frac{|X_S^\Delta|}{|A(s)|} \leq \epsilon.$$

Let the sources be chosen according to a distribution that chooses source $s \in \mathcal{S}$ with probability $\Pr(s)$. Then for any $\Delta \in \mathcal{G}$ we have

$$\begin{aligned} \epsilon_\Delta &= \sum_{s \in \mathcal{S}} \frac{\Pr(s)|X_S^\Delta|}{|A(s)|}, \\ &\leq \sum_{s \in \mathcal{S}} \Pr(s)\epsilon, \\ &= \epsilon, \end{aligned}$$

as required. □

This definition of strong security for an AMD code is stronger than the corresponding notions of security against substitution against an authentication code: in (1) the adversary's success probability is defined with respect to a specific source distribution, whereas for a strong AMD code we require success probability at most ϵ when attacking any possible source distribution. However, we note that this stronger notion of security has also been considered in the context of authentication codes [14, 15].

2 A design-theoretic perspective on splitting authentication codes

The notion of a *splitting BIBD* was introduced in [9] for the purpose of classifying splitting authentication codes that were optimal with respect to certain bounds on their parameters. In this section we introduce the related but weaker notion of a *splitting set system*. This is essentially a way of describing a splitting authentication code using design-theoretic notation that will allow us to illuminate the fundamental connection between splitting authentication codes with perfect secrecy and AMD codes, as well as describe a wider class of authentication codes with useful properties, including perfect secrecy.

Definition 2.1. An (v, b, m)-*splitting set system* consists of a finite set $\mathcal{V}$ of *points* with $|\mathcal{V}| = v$, together with a family $\mathcal{B}$ of *blocks* where $|\mathcal{B}| = b$ and each block $b_i \in \mathcal{B}$ consists of a list of m pairwise disjoint subsets $(b_{i,1}, b_{i,2}, \ldots, b_{i,m})$, with $b_{i,j} \subset \mathcal{V}$ for $j = 1, 2, \ldots, m$. If $v = b$ then the splitting set system is said to be *symmetric*.

Consider the special case where each of the subsets $b_{i,j}$ has size k, and let $\ell = km$. In this setting an (n, b, m)-splitting set system is known as a $(v, m \times k, \lambda)$-*splitting balanced incomplete block design* (splitting BIBD) if it satisfies the following condition:

- for every pair $P, Q \in \mathcal{V}$ with $P \neq Q$ there are precisely λ blocks b_i with $P \in b_{i,j}$ and $Q \in b_{i,j'}$ for some j, j' with $j \neq j'$.

Splitting BIBDs were introduced in [9], where they were shown to be equivalent to certain optimal splitting authentication codes. More generally, every splitting authentication code with equiprobable encoding gives rise to a splitting set system (and *vice versa*) by making the following identifications:

- the set of points $\mathcal{V}$ is simply the set of messages $\mathcal{T}$ of the authentication code;
- for each key $k_i \in \mathcal{K}$ we obtain a block b_i by letting $b_{i,j} = e_{k_i}(s_j)$ for $j = 1, 2, \ldots m$.

Note that it may be the case that two different keys give rise to the same encodings of the sources. In this case the splitting set system would have repeated blocks, which nonetheless correspond to distinct keys. In what follows, however, we restrict our attention to splitting set systems without repeated blocks.

This equivalent description gives a useful language for illustrating how the combinatorial properties of an authentication code with equiprobable encoding determine its security properties.

Example 2.1. For an authentication code with equiprobable sources and equiprobable encoding we can reformulate the expression for the success probability ϵ_σ of an adversary's strategy σ given in (1) in terms of the language of splitting set systems. For $b_i \in \mathcal{B}$, the set $X^\sigma_{b_i,s_j}$ is the set of points $P \in b_{ij}$ for which $\sigma(P) \in b_{ij'}$ for some $j' \neq j$, and ϵ_σ becomes

$$\epsilon_\sigma = \sum_{i=1}^{b} \sum_{j=1}^{m} \frac{|X^\sigma_{b_i,s_j}|}{bm|b_{i,j}|}. \tag{6}$$

In the c-regular case, (6) becomes

$$\epsilon_\sigma = \frac{1}{bmc} \sum_{i=1}^{b} \sum_{j=1}^{m} |X^\sigma_{b_i,s_j}|.$$

Now suppose our splitting set system is a $(v, m \times c, \lambda)$-splitting BIBD. For any $P \in \mathcal{V}$, if $\sigma(P) = P$ then $P \notin X^\sigma_{b_i,s_j}$ for any i, j. However, for each of the v points $P \in \mathcal{V}$, if $\sigma(P) \neq P$, then there are λ blocks b_i with $P \in b_{i,j}$ and $\sigma(P) \in b_{i,j'}$ for some j, j' with $j' \neq j$. Hence

$$\sum_{i=1}^{b} \sum_{j=1}^{m} |X^\sigma_{b_i,s_j}| \leq \lambda v,$$

with equality occurring for any strategy σ with $\sigma(P) \neq P$ for any $P \in \mathcal{V}$. Hence for any such σ we have

$$\epsilon_\sigma = \frac{\lambda v}{bmc}.$$

In [9] it was shown that, for a $(v, m \times c, \lambda)$-splitting BIBD, we have $\lambda = (bmk(mc-c))/(v(v-1))$. Thus we can express ϵ_σ as $(mc-c)/(v-1)$. We observe that this is precisely the expression given by (3) for this authentication code, and hence we see it has optimal substitution probability.

Example 2.2. We now determine the impersonation probability. The expression (4) can be interpreted as saying an adversary who attempts impersonation by sending point P succeeds with probability equal to the number of blocks that contain P, divided by the total number of

blocks. For a $(v, m \times c, \lambda)$-splitting BIBD, each point is contained in $\lambda(v-1)/((m-1)c)$ points [9], and so this probability becomes

$$\frac{\lambda(v-1)}{(m-1)cb}.$$

Again expressing λ in terms of the other parameters we can rearrange this expression to determine that the impersonation probability is mc/v.

On the other hand, the expression for the optimal impersonation probability given in (5) is equivalent to the size of the smallest block divided by the number of points. For a $(v, m \times c, \lambda)$-splitting BIBD this is simply mc/v, hence we see that a splitting BIBD gives rise to an authentication code with optimum impersonation probability. (This result was proved as part of Theorem 5.5 of [9] in the case where $\lambda = 1$.)

2.1 Automorphism groups of splitting set systems

We have seen that the additional structure of a splitting BIBD makes it easier to analyse the properties of the corresponding authentication codes, and to derive further results in those codes having some desirable properties. In a similar vein, we now turn our attention to the question of how the presence of certain symmetries can impact the properties of authentication codes.

Definition 2.2. An *automorphism* of a splitting set system is a bijection $\theta\colon \mathcal{V} \to \mathcal{V}$ that preserves incidence in the sense that if $b_i = (b_{i,1}, b_{i,2}, \ldots, b_{i,m}) \in \mathcal{B}$ then $b_i^\theta = (b_{i,1}^\theta, b_{i,2}^\theta, \ldots, b_{i,m}^\theta) \in \mathcal{B}$, where $b_{i,j}^\theta = \{P^\theta : P \in b_{i,j}\}$ for $j = 1, 2, \ldots, m$.

Definition 2.3. We say that a splitting set system $(\mathcal{V}, \mathcal{B})$ is *group generated* if there is an abelian subgroup $\mathcal{G}$ of its automorphism group that acts regularly on $\mathcal{V}$. That is, $(\mathcal{V}, \mathcal{B})$ is group generated if and only if there is an abelian subgroup $\mathcal{G}$ of its automorphism group with the property that for every pair of points $P, Q \in \mathcal{V}$ there is precisely one element $g \in \mathcal{G}$ such that $P^g = Q$.

Note that this definition extends readily to the case where $\mathcal{G}$ is nonabelian, but for the purposes of this paper, we restrict our attention to abelian groups.

Example 2.3. The *cyclic splitting designs* defined by Huber in [8] are a examples of a group generated splitting set systems. Here the groups in question are cyclic groups.

Consider the action of $\mathcal{G}$ on $\mathcal{B}$ that is induced by the action of $\mathcal{G}$ on $\mathcal{V}$. We refer to the orbits of blocks under this action as *block orbits* of the splitting set system. To simplify the presentation and analysis, in this paper we will assume that all block orbits have size $|\mathcal{G}|$, i.e. that $\mathcal{G}$ acts semiregularly on $\mathcal{B}$.[1]

The following lemma sets out some useful combinatorial properties of the block orbits:

Lemma 2.1. *Let Ω be a block orbit of a group generated (v, b, m)-splitting set system with $|\Omega| = v$. Then the blocks in Ω satisfies the following properties:*

1. *For any $j = 1, 2, \ldots, m$ the set $b_{i,j}$ has the same size for all $b_i \in \Omega$. Denote this size by $|b_{i,j}| = c_j^\Omega$.*
2. *Every block $b_i \in \Omega$ contains the same number of points. Denote this number by $\ell^\Omega = \sum_{j=1}^m c_j^\Omega$.*

[1] If the orbit sizes are not uniform then we can not guarantee perfect secrecy by taking a uniform distribution on the blocks. A closer attention to the probabilities and a careful application of the Orbit-Stabliser Theorem is required to analyse this case.

3. *Every point $P \in \mathcal{V}$ occurs in c_j^Ω of the sets $b_{i,j}$ with $b_i \in \Omega$.*

4. *Every point $P \in \mathcal{V}$ occurs in ℓ^Ω of the blocks in Ω.*

Proof. Let $\mathcal{G}$ be the abelian subgroup of the automorphism group that acts regularly on $\mathcal{V}$. By definition, $\mathcal{G}$ acts regularly on Ω.

1. Let $b_i \in \Omega$, and set $c_j^\Omega = |b_{i,j}|$ for $j = 1, 2 \ldots, m$. Since $\mathcal{G}$ acts regularly on Ω, it follows that for any $b_{i'} \in \Omega$ we have $b_{i'} = b_i^g$ for some $g \in \mathcal{G}$ and hence $b_{i',j} = b_{i,j}^g$. This implies that $|b_{i',j}| = c_j^\Omega$ for all $b_{i'} \in \Omega$.

2. This follows immediately from 1.

3. The number of pairs (P, b_i) where $b_i \in \Omega$ and $P \in b_{i,j}$ is $|\Omega| c_j^\Omega = v c_j^\Omega$. The collection of sets $b_{i,j}$ with $b_i \in \Omega$ is a union of orbits under the action of $\mathcal{G}$, and hence is fixed when acted on by any element of $\mathcal{G}$. As $\mathcal{G}$ acts regularly on $\mathcal{V}$, it follows that the multiset of points contained in the multiset union $\bigcup_{b_i \in \Omega} b_{i,j}$ contains each element of $\mathcal{V}$ an equal number of times. For, if some element P_1 occurred more times than the element P_2, then acting on $\{b_{i,j} : b_i \in \Omega\}$ with the unique element $g \in \mathcal{G}$ for which $P_1^g = p_2$ would not fix $\{b_{i,j} : b_i \in \Omega\}$. Thus we conclude that each point occurs $v c_j^\Omega / v = c_j^\Omega$ times in this union, and furthermore these occurrences are all in distinct sets $b_{i,j}$. (By construction, no set $b_{i,j}$ contains repeated elements.)

4. This follows immediately from 3.

□

Property 3 of Lemma 2.1, considered together with Theorem 2.3 of [11] (which applies only to c-splitting authentication codes) implies that a c-splitting authentication code arising from a group generated splitting set system has both perfect secrecy and optimal impersonation probability. By restricting our attention to group-generated splitting set systems we can prove results analogous to Lemma 2.2 and Theorem 2.3 of [11] for authentication codes that are not necessarily c-splitting.

Theorem 2.2. *The messages of an authentication code corresponding to a group generated splitting set system are distributed uniformly, and this distribution is independent of the distribution of the sources.*

Proof. Suppose $(\mathcal{V}, \mathcal{B})$ is a group generated (v, b, m)-splitting set system for which $\mathcal{G}$ is an abelian subgroup of the automorphism group that acts regularly on $\mathcal{V}$. Suppose that there are h block orbits, so we have $b = hv$. Fix a source s_j. A point $P \in \mathcal{V}$ occurs c_j^Ω times in sets $b_{i,j}$ with $b_i \in \Omega$. Each of these instances arises with probability $(b c_j^\Omega)^{-1}$, hence the total probability of obtaining the message P when the source is s_j and the key corresponds to a block in the orbit Ω is b^{-1}. Summing over all h orbits, we see that the total probability of obtaining the message P when the source is s_j is $hb^{-1} = v^{-1}$. As $\Pr(P \mid s_j) = v^{-1}$ for all $P \in \mathcal{V}$ and all $s_j \in \mathcal{S}$ we conclude that the messages are uniformly distributed, independently of the source, as required. □

This result leads directly to the following corollary.

Corollary 2.3. *An authentication code corresponding to a group generated splitting set system has perfect secrecy.*

Corollary 2.4. *An authentication code corresponding to a group generated (v, b, m)-splitting set system has optimal impersonation probability if and only if each block has the same number of points.*

Proof. By Lemma 2.1, a block b_i occurring in some block orbit Ω_h has size ℓ_h^Ω. Hence the impersonation probability of the authentication code is at least $\min_{i\in\{1,2,\ldots,h\}} \ell^{\Omega_h}/v$.

Let $P \in \mathcal{V}$. By Lemma 2.1, the number of keys that can give rise to P as an encoding of source j is given by

$$\sum_{s=1}^{h} c_j^{\Omega_s},$$

and the total number of keys that can give rise to t as an encoding of some source is thus

$$\sum_{j=1}^{m}\sum_{s=1}^{h} c_j^{\Omega_s} = \sum_{s=1}^{h} \ell^{\Omega_s}.$$

We observe that, if ℓ^{Ω_s} is some constant ℓ for each s, then this expression is simply $h\ell$, and so the impersonation probability is $h\ell/(hv) = \ell/v$, which is optimal. However, if ℓ^{Ω_s} varies with s, then

$$\begin{aligned}\sum_{s=1}^{h} \ell^{\Omega_s} &> \sum_{s=1}^{h} \min_{s\in\{1,2,\ldots,h\}} \ell^{\Omega_s},\\ &= h \min_{s\in\{1,2,\ldots,h\}} \ell^{\Omega_s},\end{aligned}$$

and so the impersonation probability is strictly greater than $\min_{s\in\{1,2,\ldots,h\}} \ell^{\Omega_s}/v$ and hence it is not optimal. □

Lemma 2.2 of [11] showed that a c-splitting authentication code for m sources, v messages and b keys has optimal impersonation probability if and only if each message P is contained in bcm/v blocks b_i, which can be seen as a result that is in some sense dual to Corollary 2.4 in the setting of group-generated c-splitting authentication codes. We note that in the case of a c-splitting group-generated splitting set system, each point is contained in cm of the blocks in each orbit by Lemma 2.1. Summing over all h orbits, this implies that each point is contained in a total of $hcm = bcm/v$ blocks, and hence the fact that it has optimal impersonation probability follows from Lemma 2.2 of [11].

These results show that group-generated splitting set systems give rise to a class of authentication codes with interesting and useful properties. They are also a natural class to consider from a point of view of seeking good constructions of splitting authentication codes: the literature contains examples of group-generated splitting BIBDs such as those arising from external difference families [9], and those in [16, 17, 8]. (We observe that as they are group-generated, the splitting authentication codes constructed in [16] provide perfect secrecy even though this property is not considered in that paper.)

2.2 AMD codes and group generated splitting set systems

In this section we explore a close connection between group-generated splitting set systems that allows us to view an AMD code as a special case of an authentication code with perfect secrecy. We consider the cases of weak and strong AMD codes separately; we will see that these

correspond respectively to authentication codes that require a uniform source distribution, or to those permit any distribution on the sources.

Starting with a weak AMD code, we can obtain a splitting set system by constructing its *development*. (This can be seen as a generalisation of Theorem 3.4 of [9].)

Definition 2.4. Let $A(s_1), A(s_2), \ldots, A(s_m) \subset \mathcal{G}$ be the sets of valid encodings of the sources of a weak AMD code. The *development* of this AMD code is the splitting set system obtained by setting $\mathcal{V} = \mathcal{G}$, and letting $\mathcal{B}$ be the set of all blocks of the form $(g + A(s_1), g + A(s_2), \ldots, g + A(s_m))$ for some $g \in \mathcal{G}$.

(Note that to simplify the presentation and analysis, we restrict our attention to the case where the development contains $|\mathcal{G}|$ distinct blocks.) This construction allows us to interpret a weak AMD code as a traditional authentication code. The results of Theorem 2.5 show that this interpretation is both useful and natural, as they demonstrate that the greatest success probability of any adversary in attacking the AMD code corresponds directly to the greatest success probability of any substitution strategy for attacking the authentication code. Immediate consequences of this interpretation include the fact that known bounds on the parameters of authentication codes apply also to the parameters of AMD codes. Indeed bounds from the literature on the parameters of AMD codes can be seen to be special cases of existing bounds for authentication codes. Further consequences of this connection will be discussed later in this section.

Theorem 2.5. *The development of a weak (m, n, ϵ)-AMD code $(\mathcal{S}, \mathcal{G}, \mathcal{A}, \mathcal{E})$ with equiprobable encoding is a group generated (n, b, m) splitting set system $(\mathcal{V}, \mathcal{B})$. In the case where it has n distinct blocks, it has $\mathcal{G}$ as a subgroup of its automorphism group that acts regularly on $\mathcal{V}$ and on $\mathcal{B}$. The corresponding splitting authentication code has perfect secrecy and optimal impersonation probability and its substitution probability is at most ϵ when the sources are chosen uniformly.*

Proof. By construction, the development of the (m, n, ϵ)-AMD code $(\mathcal{S}, \mathcal{G}, \mathcal{A}, \mathcal{E})$ is a splitting set system $(\mathcal{V}, \mathcal{B})$ whose points are the elements of $\mathcal{G}$; hence, the number of points is n. Again, by construction we see that addition by an element of $\mathcal{G}$ gives an automorphism of $(\mathcal{V}, \mathcal{B})$. Since $\mathcal{G}$ acts regularly on itself by addition it follows that $\mathcal{G}$ is a subgroup of the automorphism group of $(\mathcal{V}, \mathcal{B})$ that acts regularly on $\mathcal{V}$, hence the splitting set system is group generated. The blocks of $\mathcal{B}$ lie in a single orbit, so if $|\mathcal{B}| = n$ then the action of $\mathcal{G}$ on $\mathcal{B}$ is regular. Corollary 2.3 shows that the corresponding splitting authentication code has perfect secrecy, and Corollary 2.4 shows that it has optimal impersonation probability.

We now determine the substitution probability of the authentication code corresponding to $(\mathcal{V}, \mathcal{B})$. Let b_1 denote the block $(A(s_1), A(s_2), \ldots, A(s_m))$. Consider a substitution strategy σ. We have

$$\epsilon_\sigma = \sum_{j=1}^{m} \frac{1}{mb|A(s_j)|} \sum_{i=1}^{b} |X^\sigma_{b_i,s_j}|.$$

The sum $\sum_{i=1}^{b} |X^\sigma_{b_i,s_j}|$ counts all pairs $(P, b_i) \in \mathcal{V} \times \mathcal{B}$ with $P \in b_{i,j}$ and $\sigma(P) \in b_{i,j'}$ for some $j' \neq j$. We compute this expression in a different way. Let $P \in \mathcal{V}$, and set $\Delta_P = \sigma(P) - P$. For each point $Q \in b_{1,s_j}$ there is a unique element $g_Q \in \mathcal{G}$ with $Q + g_Q = P$. Let b_q denote the block $g_Q + b_1$. Then $P \in b_{q,j}$. We claim that $P \in X^\sigma_{b_q,s_j}$ if and only if $Q \in X^{\Delta_P}_{s_j}$: when $Q \in X^{\Delta_P}_{s_j}$ we

have $Q + \Delta_P \in b_{1,j'}$ for some $j' \neq j$, in which case the element

$$\begin{aligned} g_Q + (Q + \Delta_P) &= (g_Q + Q) + \sigma(P) - P, \\ &= P + \sigma(P) - P, \\ &= \sigma(P) \end{aligned}$$

lies in $b_{q,j'}$ (Figure 1).

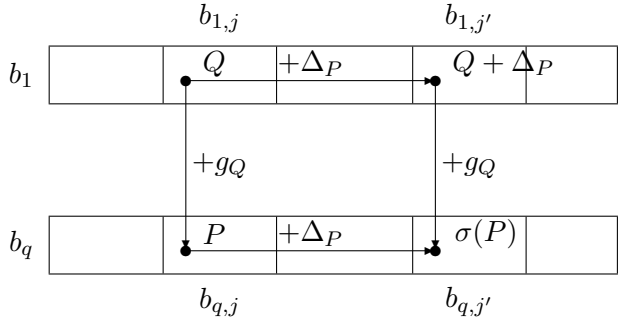

Figure 1: $P \in X^{\sigma}_{b_q,s_j}$ when $Q \in X^{\Delta_P}_{s_j}$

Thus the point P lies in precisely $|X^{\Delta_P}_{s_j}|$ of the sets $X^{\sigma}_{b_i,s_j}$ and we conclude that

$$\sum_{i=1}^{b} |X^{\sigma}_{b_i,s_j}| = \sum_{P \in G} |X^{\Delta_P}_{s_j}|,$$

which implies that the success probability of σ is

$$\begin{aligned} \sum_{j=1}^{m} \frac{1}{mn|A(s_j)|} \sum_{P \in \mathcal{G}} |X^{\Delta_P}_{s_j}| &= \frac{1}{n} \sum_{P \in \mathcal{G}} \left(\sum_{j=1}^{m} \frac{|X^{\Delta_P}_{s_j}|}{m|A(s_j)|} \right), \\ &\leq \frac{1}{n} \sum_{P \in \mathcal{G}} \epsilon, \\ &= \epsilon, \end{aligned}$$

since an adversary who chooses Δ_P in attacking the AMD code has success probability at most ϵ.

□

The following result can be seen as a weak converse of the above result:

Theorem 2.6. *A group generated (v, v, m) splitting set system $(\mathcal{V}, \mathcal{B})$ with a single block orbit of size v, for which the corresponding authentication code has substitution probability at most ϵ, gives rise to an (m, v, ϵ)-AMD code.*

Proof. Let $\mathcal{G}$ be the subgroup of the automorphism group of $(\mathcal{V}, \mathcal{B})$ that acts regularly on $\mathcal{V}$. Fix a point P. For every point $P' \in \mathcal{V}$, there is a unique element g of $\mathcal{G}$ for which $P^g = P'$; we can thus identify these points with these group elements. Pick a block $b_1 \in \mathcal{B}$; then the sets $b_{1,j} \subset \mathcal{G}$,

for $j = 1, 2, \ldots, m$, are pairwise disjoint and can be regarded as the sets $A(s_1), A(s_2), \ldots, A(s_m)$ of an AMD code $\mathcal{A}$. As there is a single block orbit of size v, $\mathcal{G}$ acts regularly on $\mathcal{B}$ and so we observe that $(\mathcal{V}, \mathcal{B})$ is in fact the development of $\mathcal{A}$. Consider a substitution strategy σ_Δ defined by setting $\sigma(P) = P + \Delta$ for each $P \in \mathcal{V}$. We observe that for any $i \in 1, 2, \ldots, v$ we have $|X^{\sigma_\Delta}_{b_i,s_j}| = |X^{\Delta}_{s_j}|$ by construction, since addition of group elements preserves differences. Thus we have

$$\begin{aligned}
\epsilon_{\sigma_\Delta} &= \sum_{j=1}^{m} \frac{1}{mvc_j} \sum_{i=1}^{v} |X^{\sigma_\Delta}_{b_i,s_j}|, \\
&= \sum_{j=1}^{m} \frac{1}{mv|A(s_j)|} \sum_{i=1}^{v} |X^{\Delta}_{s_j}|, \\
&= \frac{1}{v} \sum_{i=1}^{v} \epsilon_\Delta, \\
&= \epsilon_\Delta.
\end{aligned}$$

Hence we see that $\mathcal{A}$ is a weak (m, v, ϵ)-AMD code, as $\epsilon_{\sigma_\Delta} \leq \epsilon$ for all $\Delta \in \mathcal{G}$. □

This correspondence is not specific to the case of weak AMD codes: if we replace the weak AMD code by a strong AMD code, we obtain an authentication code that works for any source distribution.

Theorem 2.7. *The development of a strong (m, n, ϵ)-AMD code $(\mathcal{S}, \mathcal{G}, \mathcal{A}, \mathcal{E})$ with equiprobable encoding is a group generated (n, b, m) splitting set system $(\mathcal{V}, \mathcal{B})$. In the case where it has n distinct blocks, it has $\mathcal{G}$ as a subgroup of its automorphism group that acts regularly on $\mathcal{V}$ and on $\mathcal{B}$. The corresponding splitting authentication code has perfect secrecy and it has substitution probability at most ϵ for any source distribution.*

Proof. The proof follows that of Theorem 2.5 exactly, except for the determination of the substitution probability. In this case we have

$$\epsilon_\sigma = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\Pr(s_j)|X^{\sigma}_{b_i,s_j}|}{n|A(s_j)|}.$$

As before, we have $\sum_{i=1}^{n} |X^{\sigma}_{b_i,s_j}| = \sum_{P \in \mathcal{G}} |X^{\Delta_P}_{s_j}|$, and so

$$\begin{aligned}
\epsilon_\sigma &= \sum_{j=1}^{m} \frac{\Pr(s_j)}{n|A(s_j)|} \sum_{P \in \mathcal{G}} |X^{\Delta_P}_{s_j}|, \\
&= \sum_{j=1}^{m} \frac{\Pr(s_j)}{n} \sum_{P \in \mathcal{G}} \epsilon_{s_j,\Delta_P}, \\
&\leq \sum_{j=1}^{m} \frac{\Pr(s_j)}{n} \sum_{P \in \mathcal{G}} \epsilon, \\
&= \epsilon.
\end{aligned}$$

□

Theorem 2.8. *A group generated (v, v, m) splitting set system $(\mathcal{V}, \mathcal{B})$ with a single block orbit of size v, for which the corresponding authentication code has substitution probability at most ϵ for any source distribution, gives rise to a strong (m, v, ϵ)-AMD code.*

Proof. Consider the source distribution in which source s_j is chosen with probability 1, and define the substitution strategy σ_Δ as in the proof of Theorem 2.6. In this case we have

$$\begin{aligned}\epsilon_{\sigma_\Delta} &= \frac{1}{vc_j}\sum_{i=1}^{v}|X^{\sigma_\Delta}_{b_i,s_j}|,\\ &= \frac{1}{v|A_(s_j)|}\sum_{i=1}^{v}|X^{\Delta}_{s_j}|,\\ &= \epsilon_{s_j,\Delta}.\end{aligned}$$

Thus, for any choice of source s_j, we have $\epsilon_{s_j,\Delta} = \epsilon_{\sigma_\Delta} \leq \epsilon$ and so the AMD code is a strong (m, v, ϵ)-AMD code, as required. □

In [10], the notion of an *R-optimal* weak (resp. strong) AMD code was defined. These are weak (resp. strong) AMD codes for which the success probability of the worst-case adversarial choice of Δ is equal to that of the average-case choice. A c-regular weak or strong (m, n, ϵ)-AMD code is R-optimal if $\epsilon = c(m-1)/(n-1)$ (see [10]).

Corollary 2.9. *The development of a c-regular R-optimal weak (m, n, ϵ)-AMD code has optimal substitution probability when the sources are uniformly distributed. The development of a c-regular R-optimal strong (m, n, ϵ)-AMD code has optimal substitution probability for any source distribution.*

Proof. Let $\mathcal{A}$ be an R-optimal weak (m, n, ϵ)-AMD code. By Theorem 2.5 we know that its development is an authentication code with substitution probability $c(m-1)/(n-1)$ when the sources are chosen uniformly. For this authentication code, the expression in (3) is also equal to $c(m-1)/(n-1)$, and hence we conclude that this substitution probability is optimal. □

In fact the bound in (3) is not tight in general for authentication codes that are not c-splitting, as shown by the following example:

Example 2.4. Consider the weak $(4, 10, 1/2)$-AMD code $\mathcal{A}$ in $\mathbb{Z}_{10}$ that is defined by the sets $A(1) = \{0\}, A(2) = \{5\}, A(3) = \{1, 9\}, A(4) = \{2, 3\}$. This was shown in [10] to be R-optimal. However, we note that the expression of (3) for the corresponding authentication code is

$$\frac{6-2}{10-1} = \frac{4}{9} < \frac{1}{2}.$$

Inspired by the R-bound [10] for AMD codes, we now establish a bound on the substitution probability for splitting authentication codes that coincides with (3) in the c-splitting case, but which is tighter for authentication codes that are not c-splitting. Note that, although we make use of a cyclic group in the proof, we are not assuming that the authentication code is group generated, since the group elements are, in general, not automorphisms of the authentication code.

Theorem 2.10. *Let $(\mathcal{V}, \mathcal{B})$ be a (v, b, m)-splitting set system arising from an authentication code with substitution probability ϵ. Then*

$$\epsilon \geq \sum_{i=1}^{b}\frac{1}{b}\left(\frac{|b_i| - \sum_{j=1}^{m}\Pr(s_j)|b_{i,j}|}{v-1}\right).$$

Proof. We prove this result by showing that there exists a substitution strategy whose success probability is at least this value.

We identify the points of $\mathcal{V}$ with the elements of $\mathbb{Z}_v = \{0, 1, 2 \ldots, v-1\}$. For $r \in \{1, 2, \ldots, v-1\}$ we define a substitution strategy σ_r by setting $\sigma_r(P) = P + r \pmod{v}$ for all $P \in \mathcal{V}$. We now compute the mean $\overline{\epsilon_{\sigma_r}} = \sum_{r=1}^{v-1} \frac{1}{v-1} \epsilon_{\sigma_r}$ as follows:

$$\begin{aligned}
\overline{\epsilon_{\sigma_r}} &= \sum_{r=1}^{v-1} \frac{1}{v-1} \epsilon_{\sigma_r}, \\
&= \frac{1}{v-1} \sum_{r=1}^{v-1} \sum_{i=1}^{b} \sum_{j=1}^{m} \frac{|X_{b_i,s_j}^{\sigma_r}| \Pr(s_j)}{b|b_{i,j}|}, \\
&= \frac{1}{(v-1)b} \sum_{j=1}^{m} \Pr(s_j) \sum_{i=1}^{b} \frac{1}{|b_{i,j}|} \sum_{r=1}^{v-1} |X_{b_i,s_j}^{\sigma_r}|.
\end{aligned}$$

Consider the set $X_{b_i,s_j}^{\sigma_r}$. There are $|b_{i,j}|$ elements in $b_{i,j}$ and $|b_i| - |b_{i,j}|$ elements in $\bigcup_{j' \neq j} b_{i,j'}$. For each pair of elements $P \in b_{i,j}$ and $Q \in b_{i,j'}$ with $j' \neq j$ there is a unique value of r in $\{1, 2, \ldots, r-1\}$ with $\sigma_r(P) = Q$. In this case we thus have $P \in X_{b_i,s_j}^{\sigma_r}$. Hence we see that $\sum_{r=1}^{v-1} |X_{b_i,s_j}^{\sigma_r}|$ is equal to the number of such pairs, which is $|b_{i,j}|(|b_i| - |b_{i,j}|)$. Thus we have

$$\begin{aligned}
\overline{\epsilon_{\sigma_r}} &= \frac{1}{(v-1)b} \sum_{j=1}^{m} \Pr(s_j) \sum_{i=1}^{b} \frac{1}{|b_{i,j}|} |b_{i,j}|(|b_i| - |b_{i,j}|), \\
&= \frac{1}{(v-1)b} \sum_{j=1}^{m} \Pr(s_j) \sum_{i=1}^{b} (|b_i| - |b_{i,j}|), \\
&= \frac{1}{(v-1)b} \sum_{i=1}^{b} \sum_{j=1}^{m} \Pr(s_j)(|b_i| - |b_{i,j}|), \\
&= \frac{1}{(v-1)b} \sum_{i=1}^{b} \left(|b_i| - \sum_{j=1}^{m} \Pr(s_j)|b_{i,j}| \right).
\end{aligned}$$

Since this quantity is the mean of the success probabilities ϵ_{σ_r}, we conclude that there is at least one value of $r \in \{1, 2, \ldots, v-1\}$ for which ϵ_{σ_r} is greater than or equal to this quantity. □

We note that the quantity $\sum_{j=1}^{m} \Pr(s_j)|b_{i,j}|$ is the average, over all sources s_j, of the size of the set b_{ij} of possible encodings of s_j when the key is b_i. The corresponding bound in [1] has instead the maximum over all sources s_j of the size of $b_{i,j}$. For authentication codes that are not c-splitting, this new bound is thus tighter. This new bound now corresponds directly to the R-bound for an AMD code, in both the weak and strong cases. R-optimal AMD codes can be viewed as those where the success probability of the worst case choice of δ (i.e. the most successful δ) is equal to that of the average case (so that in fact the success probability of each choice of δ is the same.) A similar interpretation holds for this new bound, making it a rather natural one:

Theorem 2.11. *Let $(\mathcal{V}, \mathcal{B})$ be a (v, b, m)-splitting set system arising from an authentication code whose substitution probability ϵ attains the bound of Theorem 2.10. Then any substitution strategy σ for which $\sigma(P) \neq P$ for all $P \in \mathcal{V}$ has $\epsilon_\sigma = \epsilon$.*

Proof. The value $\overline{\epsilon_{\sigma_r}}$ was an average value taken over the $v-1$ substitution strategies of the form σ_r. We need to show that no other substitution strategies can be more successful. The key thing to note about the set of substitution strategies $\{\sigma_1, \sigma_2, \ldots, \sigma_{v-1}\}$ is that for each pair $P, Q \in \mathcal{V}$ with $P \neq Q$ there is precisely one strategy σ_r in that set with $\sigma_r(P) = Q$. Suppose instead that we wish to calculate the average success probability over the set Γ of all strategies σ for which $\sigma(P) \neq P$ for all $P \in \mathcal{V}$. For any pair $P, Q \in \mathcal{V}$ with $P \neq Q$, there are $(v-1)^{v-1}$ elements $\sigma \in \Gamma$ with $\sigma(P) = Q$. Note that $|\Gamma| = (v-1)^v$. So when we repeat the calculation we did before, the average will turn out the same, since the sum will be $(v-1)^{v-1}$ times larger, but we are dividing by $(v-1)^v$ instead of by $(v-1)$. The bound is only tight if the worst-case probability is equal to the average case, which implies that all the strategies in Γ are equiprobable. □

Observe in addition that the bound will only be tight if $|b_i| - \sum_{j=1}^{m} \Pr(s_j)|b_{i,j}|$ is constant, independent of i. In the case of uniform sources this becomes $(m|b_i| - 1)/m$, so we require the size of the blocks to be the same.

3 Constructions for Authentication Codes with Splitting and Perfect Secrecy

We observe that the literature contains examples of group-generated authentication codes that are not AMD-codes, for example splitting BIBDs that are the development of more than one base block [16, 17]. Group-generated splitting BIBDs have perfect secrecy by Corollary 2.3. In this section we consider a combinatorial property that allows us to determine when a splitting BIBD has perfect security.

3.1 Equitably Ordered Splitting BIBDs

We recall from Section 2 that a $(v, m \times c, 1)$-*splitting BIBD* is a set system consisting of a set $\mathcal{V}$ of v points and a set $\mathcal{B}$ of blocks of size mc, which satisfies the following properties:

1. each block B can be partitioned into u subsets of size c, which are denoted B_i, $1 \leq i \leq m$, and
2. given any two distinct points x and y, there is a unique block B such that $x \in B_i$ and $y \in B_j$, where $i \neq j$.

A $(v, m \times c, 1)$-splitting BIBD has replication number r and b blocks, where

$$\begin{aligned} r &= \frac{v-1}{(m-1)c} \quad \text{and} \\ b &= \frac{v(v-1)}{m(m-1)c^2}. \end{aligned}$$

Of course r and b must be integers if a $(v, m \times c, 1)$-splitting BIBD exists.

Splitting BIBDs were defined in [9] as a method of constructing authentication codes with splitting. They have been studied in a number of research papers since then. Here, our interest is in constructing authentication codes with splitting that also provide perfect secrecy. This can be accomplished if the splitting BIBD satisfies an additional property.

A $(v, m \times c, 1)$-splitting BIBD is *equitably ordered* if the multiset equation

$$\bigcup_{B \in \mathcal{B}} B_i = \frac{r}{m} X$$

is satisfied for all i, $1 \leq i \leq m$.

Theorem 3.1. *If a splitting BIBD is equitably ordered, then it yields an authentication code with perfect secrecy.*

The obvious necessary condition for a splitting BIBD to be equitably ordered is that

$$r \equiv 0 \bmod m.$$

Now, assuming an equitable ordering, we have $v = r(m-1)c + 1$ and $r = tm$ for some integer t, so

$$v = tm(m-1)c + 1.$$

Then

$$\begin{aligned} b &= \frac{(tm(m-1)c+1)(tm(m-1)c)}{m(m-1)c^2} \\ &= \frac{(tm(m-1)c+1)t}{c} \\ &= t^2m(m-1) + \frac{t}{c}, \end{aligned}$$

so $t = cs$, $r = csm$ and

$$v = sm(m-1)c^2 + 1,$$

for some integer s. That is, a splitting BIBD can be equitably ordered only if

$$v \equiv 1 \bmod (m(m-1)c^2). \tag{7}$$

It is easy to obtain $(v, m \times c, 1)$-splitting BIBDs that can be equitably ordered if they are generated by base blocks over an abelian group.

Lemma 3.2. *Suppose that a $(v, m \times c, 1)$-splitting BIBD is generated by base blocks over an abelian group of order v, and suppose every orbit of blocks has size v. Then the splitting BIBD can be equitably ordered.*

Proof. Under the stated hypotheses, the splitting BIBD is generated from

$$\frac{v-1}{m(m-1)c^2}$$

base blocks. Each base block gives rise to v blocks in the design. We can arbitrarily order each base block. Then the development of each base block yields exactly c copies of each point in each of the m sets. Therefore, we get

$$\frac{v-1}{m(m-1)c} = \frac{r}{m}$$

copies of each point in each of the m sets. □

Example 3.1. A $(25, 3 \times 2, 1)$-splitting BIBD is presented in [7]. It has points in $\mathbb{Z}_{25}$ and it is generated from the base block

$$\{\{0,1\},\{2,4\},\{12,20\}\}.$$

If we order the base block as

$$(\{0,1\},\{2,4\},\{12,20\})$$

and maintain this ordering as the block is developed, we obtain the blocks

$$\begin{gathered}(\{0,1\},\{2,4\},\{12,20\})\\(\{1,2\},\{3,5\},\{13,21\})\\\vdots\\(\{24,0\},\{1,3\},\{11,19\}).\end{gathered}$$

Then each point occurs twice in the union of the first sets, second sets and third sets.

Using the technique of Lemma 3.2, we can construct equitably ordered $(v, 2\times c, 1)$-splitting BIBDs.

Theorem 3.3. *For any $c \geq 2$, an equitably ordered $(v, 2\times c, 1)$-splitting BIBD exists if and only if $v \equiv 1 \bmod (2c^2)$.*

Proof. Necessity follows from (7). For sufficiency, we use a construction from [7]. There, it is shown that a $(2c^2t+1, 2\times c, 1)$-splitting BIBD can be constructed from t base blocks defined over $\mathbb{Z}_{2c^2t+1}$. Applying Lemma 3.2, we have the desired result. □

We use the following general recursive approach to construct various families of $(v, m\times c, 1)$-splitting BIBDs that are equitably ordered. This construction will make use of group divisible designs. We note that the term "group" here is a historical usage that does not refer to an algebraic group. To avoid confusion, we will refer to the groups of a group divisible design as "design groups" to clarify that we are talking about particular sets of points in the design, rather than an algebraic group. A *group-divisible design* consists of a set of points $\mathcal{V}$, a set G of *groups* that forms a partition of $\mathcal{V}$, and a set of blocks $\mathcal{B}$ such that no block contains more than one point from the same design group, and every pair of points from different design groups is in a unique block. A group divisible design is an m-GDD if every block has size m. The *type* of a GDD is the multiset of its design group sizes. The type of a GDD is usually described using an exponential notation.

Suppose that there is an m-GDD on $sm(m-1)c$ points, such that

$$|G| \equiv 0 \bmod (m(m-1)c)$$

for every design group G. The replication number r_x of any point $x \in G$ is

$$\begin{aligned} r_x &= \frac{sm(m-1)c - |G|}{m-1}\\ &= \frac{sm(m-1)c - s'm(m-1)c}{m-1} \quad \text{for some intger } s'\\ &= mc(s-s')\\ &\equiv 0 \bmod m. \end{aligned}$$

We show that an m-GDD satisfying the above properties can be equitably ordered, by using a technique from [15]. We first construct the point vs block bipartite incidence graph for the m-GDD. Each "block" vertex has degree m and each "point" vertex x has degree $r_x \equiv 0 \bmod m$. Split each "point" vertex x into r_x/m vertices of degree m. Now we have an m-regular bipartite graph, which therefore can be m-edge-coloured. Say the colours are $1, 2, \ldots, m$. For each block, this specifies an ordering of the points in such a way that every point occurs equally often in each position. Therefore the blocks of the GDD have been equitably ordered.

Next, we take c copies of every point in the GDD and replace every (ordered) block by the trivial $(mc, m \times c, 1)$-splitting GDD of type c^m. That is, each ordered block $(x_1, x_2, \ldots, x_m)$ is replaced by

$$(\{x_1\} \times \{1, \ldots, c\}, \{x_2\} \times \{1, \ldots, c\}, \ldots, \{x_m\} \times \{1, \ldots, c\}).$$

This yields an $(sm(m-1)c^2, m \times c, 1)$-splitting GDD that is equitably ordered.

Suppose further that there is a $(c|G|+1, m \times c, 1)$-splitting BIBD that is equitably ordered, for every design group G in the m-GDD. Note that

$$c|G| + 1 \equiv 1 \bmod (m(m-1)c^2),$$

so the necessary numerical condition (7) is satisfied. Then we obtain a $(v, m \times c, 1)$-splitting BIBD by simply taking the blocks in the $(sm(m-1)c^2, m \times c, 1)$-splitting GDD along with all the blocks in the various $(c|G|+1, m \times c, 1)$-splitting BIBDs. Since each of these designs is equitably ordered, the resulting $(v, m \times c, 1)$-splitting BIBD is equitably ordered.

Summarizing the discussion above, we have the following.

Theorem 3.4. *Suppose that $v = sm(m-1)c^2 + 1$ and suppose there is an m-GDD on $(v-1)/c$ points, such that the following conditions hold for every design group G:*

1. *$|G| \equiv 0 \bmod (m(m-1)c)$ and*
2. *there is a $(c|G|+1, m \times c, 1)$-splitting BIBD that is equitably ordered.*

Then there is a $(v, m \times c, 1)$-splitting BIBD that is equitably ordered.

We now construct several families of equitably ordered $(v, m \times c, 1)$-splitting BIBDs, for fixed m and c, using Theorem 3.4.

Theorem 3.5. *There exists a $(v, 3 \times 2, 1)$-splitting BIBD that is equitably ordered if and only if $v \equiv 1 \bmod 24$.*

Proof. The necessary condition $v \equiv 1 \bmod 24$ follows from (7). We prove sufficiency using the same approach as [7]. Let $v = 24s + 1$. For the case $s = 1$, an equitably ordered $(25, 3 \times 2, 1)$-splitting BIBD (from [7]) was presented in Example 3.1. For $s = 2$, the $(49, 3 \times 2, 1)$-splitting BIBD presented in [7] can be equitably ordered by Lemma 3.2. For $s \geq 3$, we proceed as follows. A 3-GDD of type 12^s exists for all $s \geq 3$. As we have already mentioned, there is an equitably ordered $(25, 3 \times 2, 1)$-splitting BIBD. Therefore, from Theorem 3.4, we obtain a $(24s + 1, 3 \times 2, 1)$-splitting BIBD that is equitably ordered. □

Theorem 3.6. *There exists a $(v, 4 \times 2, 1)$-splitting BIBD that is equitably ordered if and only if $v \equiv 1 \bmod 48$, with the possible exception of $v = 49$.*

Proof. The necessary condition $v \equiv 1 \bmod 48$ follows from (7). Let $v = 48s + 1$. A $(49, 4 \times 2, 1)$-splitting BIBD is not known to exist, so we cannot handle the case $s = 1$. For $s = 2, 3, 4, 5, 6, 7$ and 9, $(48s + 1, 3 \times 2, 1)$-splitting BIBDs are given in [7] that are generated from base blocks over groups. Therefore, using Lemma 3.2, we have equitably ordered splitting BIBDs for these values of s.

For $s = 8$ and $s \geq 10$, we use 4-GDDs on $24s$ points with design group sizes divisible by, and greater than, 24. A 4-GDD of type $48^{s/2}$ exists for all even $s \geq 8$ (see [2]). A 4-GDD of type $12^{(s-3)/2}18^1$ exists for all odd $s \geq 11$ (see [6]). Giving weight 4 to every point and applying the Fundamental GDD Construction ([3] Section IV.2.1), we obtain a 4-GDD of type $48^{(s-3)/2}72^1$ for all odd $s \geq 11$. Hence, from Theorem 3.4, we obtain a $(48s + 1, 4 \times 2, 1)$-splitting BIBD that is equitably ordered, for $s = 8$ and for all $s \geq 10$. □

Theorem 3.7. *There exists a $(v, 3 \times 3, 1)$-splitting BIBD that is equitably ordered if and only if $v \equiv 1 \bmod 55$, with the possible exception of $v = 33$.*

Proof. For $(v, 3 \times 3, 1)$-splitting BIBDs, this result was shown by Wang [16]. We use a slightly different recursive construction to construct splitting BIBDs that are equitably ordered. First, the necessary condition $v \equiv 1 \bmod 54$ follows from (7). Let $v = 54s + 1$. A $(55, 3 \times 3, 1)$-splitting BIBD is not known to exist, so we cannot handle the case $s = 1$. For $s = 2, 3, 4, 5$ and 7, $(54s + 1, 3 \times 2, 1)$-splitting BIBDs are given in [16] that are generated from base blocks over groups. Therefore, using Lemma 3.2, we have equitably ordered splitting BIBDs for these values of s.

For $s = 6$ and $s \geq 8$, we use 4-GDDs on $18s$ points with design group sizes divisible by, and greater than, 18. A 3-GDD of type $36^{s/2}$ exists for all even $s \geq 6$, and a 3-GDD of type $36^{(s-3)/2}54^1$ exists for all odd $s \geq 9$ (see [4]). Hence, from Theorem 3.4, we obtain a $(54s + 1, 3 \times 3, 1)$-splitting BIBD that is equitably ordered, for $s = 6$ and for all $s \geq 8$. □

Theorem 3.8. *There exists a $(v, 3 \times 4, 1)$-splitting BIBD that is equitably ordered if and only if $v \equiv 1 \bmod 96$.*

Proof. The necessary condition $v \equiv 1 \bmod 96$ follows from (7). We prove sufficiency using the same approach as [17]. Let $v = 96s + 1$. For $s = 1, 2$, $(96s + 1, 3 \times 2, 1)$-splitting BIBDs are given in [17] that are generated from base blocks over groups. Therefore, using Lemma 3.2, we have equitably ordered splitting BIBDs $s = 1, 2$.

For $s \geq 3$, we proceed as follows. A 3-GDD of type 24^s exists for all $s \geq 3$. From Theorem 3.4, we obtain a $(96s + 1, 3 \times 4, 1)$-splitting BIBD that is equitably ordered. □

4 Discussion and Conclusion

Theorems 2.5 and 2.6 show that a weak AMD code is in fact a special case of a group-generated authentication code for uniformly distributed sources. The fact that these authentication codes have perfect secrecy gives a new perspective on the potential context in which an AMD code might be applied. The traditional description of a weak AMD code involves an adversary who is unable to see an encoded message, but who can add a group element to that unknown message. Thus an AMD code can only be applied in a context where these rather specific properties arise. When treating the AMD code as an authentication code with perfect secrecy, it can be applied in any context where an authentication code might be useful. Here the adversary sees the encoded message, but it is independent of the source and hence provides no information about the source.

This perspective also allows us to identify those properties of an AMD code that do not hold for more general authentication codes. For example, choosing a group element to add to the encoded message defines a substitution strategy for the authentication code. We observe that each substitution strategy arising this way has the property that the probability of its success conditioned on the event of the key being k is the same for all $k \in \mathcal{K}$. This property could facilitate an analysis of success probabilities of substitution attacks in a context where the adversary learns partial information about the choice of key, for example.

We have established that group-generated splitting set systems in general, and AMD codes in particular, are useful classes of splitting authentication codes with perfect secrecy. It is interesting to see whether they can be further exploited in the construction of splitting authentication codes with perfect secrecy that achieve optimal or near-optimal security against substitution attacks, and whether an explicit focus on the perfect secrecy property can inspire new applications for AMD codes.

References

[1] C. Blundo, A. De Santis, K. Kurosawa, and W. Ogata. On a fallacious bound for authentication codes. *J. Cryptol.*, 12(3):155–159, 1999.

[2] A. Brouwer, A. Schrijver, and H. Hanani. Group divisible designs with block-size four. *Discrete Math.*, 20:1 – 10, 1977.

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

[4] C. J. Colbourn, D. G. Hoffman, and R. Rees. A new class of group divisible designs with block size three. *J. Combin. Theory, Series A*, 59(1):73 – 89, 1992.

[5] R. Cramer, Y. Dodis, S. Fehr, C. Padró, and D. Wichs. Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. In N. P. Smart, editor, *EUROCRYPT '08*, volume 4965 of *LNCS*, pages 471–488. Springer, 2008.

[6] G. Ge and A. C. Ling. Group divisible designs with block size four and group type $g^u m^1$ for small g. *Discrete Math.*, 285(1):97 – 120, 2004.

[7] G. Ge, Y. Miao, and L. Wang. Combinatorial constructions for optimal splitting authentication codes. *SIAM J. Discrete Math.*, 18(4):663–678, 2005.

[8] M. Huber. Information theoretic authentication and secrecy codes in the splitting model. In *22nd International Zurich Seminar on Communications (IZS)*. Eidgenössische Technische Hochschule Zürich, 2012.

[9] W. Ogata, K. Kurosawa, D. R. Stinson, and H. Saido. New combinatorial designs and their applications to authentication codes and secret sharing schemes. *Discrete Math.*, 279(1):383 – 405, 2004. In Honour of Zhu Lie.

[10] M. B. Paterson and D. R. Stinson. Combinatorial characterizations of algebraic manipulation detection codes involving generalized difference families. *Discrete Math.*, 339(12):2891–2906, 2016.

[11] M. B. Paterson and D. R. Stinson. On the equivalence of authentication codes and robust (2, 2)-threshold schemes. *J. Math. Cryptol.*, 15(1):179–196, 2021.

[12] G. J. Simmons. Authentication theory/coding theory. In G. R. Blakley and D. Chaum, editors, *CRYPTO '84*, volume 196 of *LNCS*, pages 411–431. Springer, 1984.

[13] M. D. Soete. New bounds and constructions for authentication/secrecy codes with splitting. *J. Cryptol.*, 3(3):173–186, 1991.

[14] D. R. Stinson. Some constructions and bounds for authentication codes. *J. Cryptol.*, 1(1):37–52, 1988.

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

[16] J. Wang. A new class of optimal 3-splitting authentication codes. *Des. Codes, Cryptogr.*, 38(3):373–381, 2006.

[17] J. Wang and R. Su. Further results on the existence of splitting BIBDs and application to authentication codes. *Acta Appl. Math.*, (3):791–803, 2010.