ON THE RELATION BETWEEN A-CODES AND CODES CORRECTING INDEPENDENT ERRORS

Thomas Johansson* Gregory Kabatianskii** Ben Smeets*

*Department of Information Theory
Lund University
Box 118
S-221 00 Lund, Sweden

**Inst for Problems of Information Transmission
Russian Academy of Sciences
Ermolovoy 19, Moscow, GSP-4
Russia

July 23, 1993

Abstract – In this paper we show an explicit relation between authentication codes and codes correcting independent errors. This relation gives rise to several upper bounds on A-codes. We also show how to construct A-codes starting from error correcting codes. The latter is used to show that if P_S exceeds P_I by an arbitrarily small positive amount, then the number of source states grows exponentially with the number of keys but if $P_S = P_I$ it will grow only linearly.

1 Introduction

The *authentication channel* was introduced by Simmons as a model for a communication situation with two trusting participants called the *transmitter* and the *receiver* who want to protect themselves against the actions of an active *opponent* who can insert its own messages or who can change messages already sent by the transmitter. To protect their communication, the transmitter and the receiver have agreed *secretly* upon a certain encoding rule e which is taken from a finite set $\mathcal{E}$ of n possible rules. This rule e enables them to transmit a piece of information s, hereafter called source state, which is taken from the finite set S of k possible source states. Each encoding rule maps s to a message m. In this paper we investigate only authentication codes (A-codes) for which the messages are pairs (s, z), $z \in \mathcal{Z}$. The coordinate z is called a tag or an authenticator. In authentication theory these codes are called appended authenticator schemes or Cartesian A-codes. People working in coding theory would use the term systematic codes. Furthermore, in this paper we will only deal with unconditionally secure A-codes.

T. Helleseth (Ed.): Advances in Cryptology - EUROCRYPT '93, LNCS 765, pp. 1-11, 1994.

Authentication theory deals with the analysis and design of A-codes and has since the publication of the paper by Gilbert, McWilliams and Sloane, [1], and the paper by Simmons, [2], developed into a discipline on its own. Various bounds on the probabilities of success for the various attacks by the opponent have been established and many constructions for obtaining A-codes are known. For an overview of the results between 1974 and 1991 we refer to [3] and [4].

At several occasions the view has been expressed that unconditionally secure authentication codes are in a strict mathematical sense dual to error detecting and correcting codes, see for example [3, page 397, 3rd par.]. However, the true nature of this duality seems never to be addressed. It is just this aspect that is one of the main subjects of this paper. Particularly we will show how to obtain a code for correcting independent errors (CIE-code) from an A-code and vice versa. By establishing this connection we obtain new results in authentication theory.

Another subject which we address is the question of how large we can make $\mathcal{S}$ for given n, P_I, and P_S, the latter two being the probability of a successful imitation attack, respectively, the probability of a successful substitution attack. In this paper we compute P_S as the *maximum probability* over all substitution messages. The above question has practical relevance if we want to authenticate, for example, long data files. It has previously been shown that if $P_S = P_I$, then the number of source states is linearly bounded by the number of keys, [5]. However, our results show that if $P_S = P_I + \varepsilon$, for arbitrary $\varepsilon > 0$, then the number of source states grows exponentially with the number of keys !

The organization of our paper will be as follows. First, in Section 2, we discuss how we can derive a CIE code from an A-code. We give some examples which illustrate the implications of the established relation. In this section we also give the bound on $|\mathcal{S}|$ when $P_S = P_I$. In Section 3 we show how we can construct an A-code starting with a CIE code.

2 From systematic A-codes to CIE-codes

In this section we will describe how we can derive a CIE-code from a systematic A-code and how we can state the probabilities P_I and P_S in terms of the properties of the CIE-code.

Consider a systematic A-code, i.e., a triple $(\mathcal{S}, \mathcal{E}, \mathcal{Z})$, where for any $s \in \mathcal{S}$ and $e \in \mathcal{E}$, the message $m = (s, z) \in \mathcal{S} \times \mathcal{Z}$ is defined by letting $z = e(s)$. We restrict ourselves to a uniform distribution of e. From definitions in [2] we derive that the probability of a successful impersonation attack, P_I, is given by the formula

$$P_I = \max_{\substack{s \in \mathcal{S} \\ z \in \mathcal{Z}}} \frac{|\{e \in \mathcal{E}\,;\, e(s) = z\}|}{|\mathcal{E}|}. \tag{1}$$

Let us consider A-codes with the property that

$$|\{e \in \mathcal{E}\,;\, e(s) = z\}| = \begin{cases} |\mathcal{E}|\,P_I, & \text{or,} \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

and let us call such codes *I-equitable A-codes.* These codes are optimal against the impersonation attack in the sence that only these codes meet the trivial bound $P_I \geq |\mathcal{S}| / |\mathcal{M}|$ with equality, [3]. In the sequel we assume that our A-codes are I-equitable and that $P_I = 1/q$, where q is a power of a prime.

Let $n = |\mathcal{E}|$ and let us enumerate the elements of $\mathcal{E}$ as $e_1, e_2, \ldots, e_n$. Consider the words (vectors)

$$\mathbf{v}^{(s)} = (e_1(s), e_2(s), \ldots, e_n(s)),$$

and the corresponding set of words

$$V = \left\{\mathbf{v}^{(s)}\, ;\, s \in \mathcal{S}\right\}.$$

For given s let us also enumerate the values $e_j(s) \in \mathcal{Z}$ by the elements $b_1, b_2, \ldots, b_q$ from some q-ary alphabet B. It follows from property (2) that such an enumeration is possible. For each word $\mathbf{v}$ we define its composition as

$$\mathrm{comp}(\mathbf{v}) = (c_1, c_2, \ldots, c_q), \quad \text{where } c_i = \frac{1}{n}\left|\{j\, ;\, v_j = b_i\}\right|. \tag{3}$$

It follows from (2) that

$$\mathrm{comp}\left(\mathbf{v}^{(s)}\right) = (P_I, \ldots, P_I) = (\frac{1}{q}, \ldots, \frac{1}{q}),$$

i.e., all words of V have a constant composition.

Now recall from the introduction that the probability P_S is given by

$$P_S = \max_{\substack{s \in \mathcal{S} \\ z \in \mathcal{Z}}} \max_{\substack{s \neq \hat{s} \in \mathcal{S} \\ \hat{z} \in \mathcal{Z}}} \frac{|\{e \in \mathcal{E}\, ;\, e(s) = z, e(\hat{s}) = \hat{z}\}|}{|\{e \in \mathcal{E}\, ;\, e(s) = z\}|}. \tag{4}$$

From (2) we get the inequality

$$\left|\left\{j\, ;\, v_j^{(s)} = b, v_j^{(\hat{s})} = \hat{b}\right\}\right| \leq P_S P_I \cdot n. \tag{5}$$

By letting $b = \hat{b}$ and letting b run through B, we have

$$\mathrm{d}\left(\mathbf{v}^{(s)}, \mathbf{v}^{(\hat{s})}\right) \geq n - qP_I P_S \cdot n = n(1 - P_S), \tag{6}$$

where $\mathrm{d}(\mathbf{x}, \mathbf{y})$ denotes the usual Hamming distance between the vectors $\mathbf{x}$ and $\mathbf{y}$. But in general, we can let $b = c_1\hat{b} + c_2$ for arbitrary $c_1 \neq 0, c_2 \in B$, if we consider B as a finite field. Then (5) gives

$$\mathrm{d}\left(\mathbf{v}^{(s)}, c_1\mathbf{v}^{(\hat{s})} + c_2\mathbf{1}\right) \geq n - qP_I P_S \cdot n = n(1 - P_S). \tag{7}$$

This means that if we assume $P_S \neq 1$, we form from each codeword $\mathbf{v}^{(s)}$ in the code V new codewords by all the affine transformations $\phi : v \mapsto c_1 v + c_2$, where $c_1 \neq 0, c_2 \in B$. If $P_S \neq 1$ no two codewords from this transformation can be the same and the distance property of (6) still holds, *provided* we also assume that $P_S \geq P_I = 1/q$. Since all codewords have constant composition we also add multiples of the codeword $\mathbf{1}$ without changing the minimum distance. Thus we have a code V' given by

$$V' = \left\{c_1 \mathbf{v}^{(s)} + c_2 \mathbf{1} \,;\, \mathbf{v}^{(s)} \in V, c_1 \neq 0, c_2 \in B\right\} \cup \{c\mathbf{1} \,;\, c \in B\}$$

with the same distance property as V, i.e., the minimum distance d of the code V' is bounded by $d \geq n(1 - P_S)$. The number of codewords in V' is then $q(q-1)\,|\mathcal{S}| + q$.

Summary:
Given an I-equitable A-code with parameters $|\mathcal{E}|$, $|\mathcal{S}|$ and P_S, there exists a corresponding CIE-code with parameters $(n, M, d) = (|\mathcal{E}|\,, q(q-1)\,|\mathcal{S}| + q, |\mathcal{E}|\,(1 - P_S))$.

We can now apply upper bounds on the code V' to get upper bounds on the maximum number of source states in an I-equitable A-code with given P_S. For example by using the well-known Plotkin bound [7] we can prove

Theorem 1 **([5])**:For an I-equitable A-code for which $P_I = P_S = 1/q$, the number of source states is upper bounded as

$$(q-1)\,|\mathcal{S}| \;\leq\; |\mathcal{E}| - 1. \tag{8}$$

■

Proof: The code obtained from the A-code has parameters $(n, M, d) = (|\mathcal{E}|\,, q(q-1)(|\mathcal{S}| + q, \theta \cdot |\mathcal{E}|)$, where $\theta = 1 - P_S = (q-1)/q$. Let $A_q(n, d)$ be the maximum number of codewords in an (n, d)-code. By the Plotkin bound, [7, pages 170-171], we have

$$A_q(n, \theta n) \leq q A_q(n-1, \theta n) \leq q \frac{\theta n}{\theta n - \theta(n-1)} = qn = q\,|\mathcal{E}|\,.$$

From this we have $q(q-1)\,|\mathcal{S}| + q \leq q\,|\mathcal{E}|$ and the result follows. □

Remark: This bound shows essentially what is happening for the case $P_S = \frac{1}{q}$, namely, the size of the source state space is, at the best, bounded by the size of the key space. Consequently we have to tolerate a large key when we want to authenticate many source states.

Let us return to equation (6). This is only a weak corollary to equation (5). One of the serious weaknesses of this result is that we calculate in (6) the total number of positions in which two words differ but the key property in the computation of P_S of an A-code is that the *ratio* of the number of positions in which two words differ must be almost the same for all given values of the coordinates, see equation (5).

Let us look at some examples which illustrate this.

Example 1: Let $q = 2$, then $P_I = 1/2$ and the vectors $\mathbf{v}^{(s)}$ of V have composition $(1/2, 1/2)$, i.e. they are binary vectors of constant weight $n/2$. We can rewrite (5) as

$$\left|\left\{j \,;\, v_j^{(s)} = b, v_j^{(\hat{s})} = \hat{b}\right\}\right| \leq P_S \frac{n}{2}.$$

Letting $b = \widehat{b}$ we have $\mathrm{d}\left(\mathbf{v}^{(s)}, \mathbf{v}^{(\hat{s})}\right) \geq n - qP_IP_Sn = n(1-P_S)$. On the other hand by letting $b = \widehat{b} \oplus 1$ (complement), we have $\mathrm{d}\left(\mathbf{v}^{(s)}, \mathbf{v}^{(\hat{s})} \oplus \mathbf{1}\right) \geq n(1-P_S)$, where $\mathbf{1}$ is the word (vector) $(1, 1, \ldots, 1)$. The latter inequality can also be written as $\mathrm{d}\left(\mathbf{v}^{(s)}, \mathbf{v}^{(\hat{s})}\right) \leq n - n(1-P_S) = P_sn$. It can be shown that if for some binary code V, $\alpha n \leq d(\mathbf{v}, \mathbf{v}') \leq (1-\alpha)n$ and $\mathbf{v}$, $\mathbf{v}'$ have weight $n/2$, that this code gives us an A-code with $P_I = 1/2$ and $P_S = \alpha$. Hence in the case $P_I = 1/2$ we have a one-to-one correspondence between I-equitable A-codes and binary weight $n/2$ codes in the so-called antipodal Hamming space, (we define $\mathrm{d}_{\text{antipodal}}(\mathbf{x}, \mathbf{y}) = \min(\mathrm{d}(\mathbf{x}, \mathbf{y}), n - \mathrm{d}(\mathbf{x}, \mathbf{y}))$ and $\mathbf{x} \equiv \mathbf{y}$ if and only if $\mathbf{x} = \mathbf{y} \oplus \mathbf{1}$). In particular, if

$$S(\mathcal{E}, P_I, P_S) = \max |\mathcal{S}|, \tag{9}$$

for A-codes with given $\mathcal{E}$, P_I and P_S, we have

$$\lim_{n \to \infty} \frac{\log S(n, 1/2, P_S = \alpha)}{n} = f(\alpha) = \lim_{n \to \infty} \frac{1}{n} \log A_2(n, \alpha n),$$

where $f(\alpha)$ is a final answer in coding theory. If the Varshamov-Gilbert (V-G) bound, [6], is tight in the binary case, then $f(\alpha) = 1 - H(\alpha)$, where $H(\alpha) = -\alpha \log(\alpha) - (1-\alpha)\log(1-\alpha)$ is the binary entropy function.

Note that when $P_S \to 1/2$, then $f(P_S) \to 0$ as it should since when $P_S = P_I = 1/2$ we have by Theorem 1 that $\lim_{n \to \infty} \log |\mathcal{S}| / n \leq \lim_{n \to \infty} \log(n-1)/n = 0$. ◘

Example 2: Let $q = 3$, and let the alphabet $B = \{0, -1, 1\}$. Consider two different A-codes and their corresponding 3-ary codes of which we only list two codewords.

$$V : \begin{cases} 1 & 1 & 1 & \text{-1} & \text{-1} & \text{-1} & 0 & 0 & 0 \\ 1 & 0 & \text{-1} & 1 & 0 & \text{-1} & 1 & 0 & \text{-1} \end{cases} \quad d = 6, \qquad P_I = 1/3, \quad P_S = 1/3,$$

$$\tilde{V} : \begin{cases} 1 & 1 & 1 & \text{-1} & \text{-1} & \text{-1} & 0 & 0 & 0 \\ \text{-1} & \text{-1} & 0 & 0 & 0 & 1 & 1 & 1 & \text{-1} \end{cases} \quad d = 9, \qquad P_I = 1/3, \quad P_S = 2/3.$$

We can discover that $\tilde{V}$ is a bad A-code by permuting the assignment of the symbols of B. Thus we get 3! words from the first and and 3! words from the second word and by pairwise checking their distances we find that $d = 3$. This renumeration technique can be used to improve the bound of equation (7). It is important to note that we can only check the distance between words that stem from different codewords in $\tilde{V}$, since the minimum distance of two words that stem from the same codeword may be less than $|\mathcal{E}|(1-P_S)$.

If we want to consider all the obtained codewords as a code with minimum distance unchanged, we must check that the minimum distance of any two words from the same codeword in $\tilde{V}$ is at least $|\mathcal{E}|(1-P_S)$. Thus it will depend on the value of P_S which permutations we can apply. For example, if $P_S \geq (q-2)/q$ we can apply all the $q!$ permutations without changing the minimum distance and thus get a code with parameters

$$(n, M, d) = (|\mathcal{E}|, q!\,|\mathcal{S}| + q, |\mathcal{E}|(1-P_S)), \quad \text{for } P_S \geq \frac{q-2}{q}.$$

However, the following example shows that even with this renumeration we do not get a tight bound.

$$V' : \begin{cases} 1 & 1 & 1 & \text{-1} & \text{-1} & \text{-1} & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & \text{-1} & \text{-1} & \text{-1} & 0 \end{cases} \quad d = 4$$

After renumeration we get

$$\begin{cases} 1\ 1\ 1\ \text{-}1\ \text{-}1\ \text{-}1\ 0\ 0\ 0 \\ 1\ 1\ 1\ \text{-}1\ \text{-}1\ 0\ 0\ 0\ \text{-}1 \end{cases} \quad d=2,$$

but we have $P_S = 1$! □

Example 3: We now give a new A-code construction.

Consider the following construction of a systematic A-code.

Construction 1: Let $\mathcal{S} = \{s = (s_1, ..., s_k)\,;\ s_i \in \mathbb{F}_q\}$. Now define the source state polynomial to be

$$s(x) = s_1 x + s_2 x^2 + \ldots + s_k x^k.$$

Let $\mathcal{E} = \{e = (a,b)\,;\ a,b \in \mathbb{F}_q\}$. For the transmission of source state s we generate the message m which is obtained as

$$m = (s, a + s(b)) = (s_1, s_2, \ldots, s_k, a + s(b)).$$

Theorem 2: Construction 1 gives an A-code which has parameters:

$$P_I = \frac{1}{q}, \quad P_S = \frac{k}{q}.$$

■

Proof: Here $|\mathcal{E}| = q^2$ and from (1) we have

$$P_I = \max_{\substack{s\in\mathcal{S}\\ z\in\mathcal{Z}}} \frac{|\{e \in \mathcal{E}\,;\ e(s) = z\}|}{|\mathcal{E}|} = \max_{\substack{s\in\mathcal{S}\\ z\in\mathcal{Z}}} \frac{|\{a,b\,;\ s(b) + a = z\}|}{q^2}.$$

For a given value of b is a uniquely determined by $a = z - s(b)$ for any value of (s,z) and thus $P_I = q/q^2 = 1/q$. For the substitution attack, we have from (4) that

$$P_S = \max_{\substack{s\in\mathcal{S}\\ z\in\mathcal{Z}}} \max_{\substack{s\neq\hat{s}\in\mathcal{S}\\ \hat{z}\in\mathcal{Z}}} \frac{|\{a,b\,;\ s(b) + a = z, \hat{s}(b) + a = \hat{z}\}|}{|\{a,b\,;\ s(b) + a = z\}|} =$$

$$= \max_{\substack{s\in\mathcal{S}\\ z\in\mathcal{Z}}} \max_{\substack{s\neq\hat{s}\in\mathcal{S}\\ \hat{z}\in\mathcal{Z}}} \frac{|\{a,b\,;\ s(b) + a = z, (s-\hat{s})(b) + (z-\hat{z}) = 0\}|}{q}.$$

Now a is uniquely determined by $a = z - s(b)$ and since $(s-\hat{s})(x) + (z-\hat{z})$ is a non-zero polynomial of degree at most k it has at most k zeros. Thus for any $(s,z), (\hat{s},\hat{z})$ we have $|\{a,b\,;\ s(b) + a = z, (s-\hat{s})(b) + (z-\hat{z}) = 0\}| \leq k$ and $P_S = k/q$. □

□

Consider again an I-equitable A-code. It is also possible to associate a binary code to our A-code. Let us assign to every source state $s \in \mathcal{S}$ q binary vectors of length n

$$\mathbf{b}^{(s,0)}, \mathbf{b}^{(s,1)}, \ldots, \mathbf{b}^{(s,q-1)}, \quad \text{where } b_i^{(s,j)} = \begin{cases} 1, & \text{if } e_i(s) = j, \\ 0, & \text{otherwise,} \end{cases}$$

i.e., the characteristic functions. It follows from the properties of our A-code that these vectors have weight n/q, $n = |\mathcal{E}|$, and any two distinct vectors have not more than $P_S(n/q)$ "common" 1-s. Hence the distance d of the code obtained from these vectors is at least $2(1 - P_S)(n/q)$ and thus

$$S(n = |\mathcal{E}|, 1/q, P_S) \leq \frac{1}{q} A_2 \left(n, \underbrace{2(1 - P_S)\frac{n}{q}}_{2\delta}, \underbrace{\frac{n}{q}}_{w} \right), \tag{10}$$

see (9) and $A_2(n, d, w)$ in [6]. For $w - \delta =$ constant, $w/n =$ constant and $n \to \infty$ we have, [6, page 527],

$$A_2(n, 2\delta, w) \leq \left\lfloor \frac{n}{w} \left\lfloor \frac{n-1}{w-1} \cdots \left\lfloor \frac{n - (w - \delta)}{\delta} \right\rfloor \cdots \right\rfloor \right\rfloor \approx \left(\frac{n}{w} \right)^{w - \delta + 1}.$$

We see that the A-codes in Construction 1 are asymptotically optimal because we have $|\mathcal{S}| = q^k$, $n = |\mathcal{E}| = q^2$, $w = q$, $w - \delta + 1 = k + 1$ and

$$\frac{1}{q} A_2(q^2, 2(q - k), q) \sim q^{k+1-1} = |\mathcal{S}|, \quad \text{for } k \text{ fixed and } q \to \infty.$$

Summarizing, we have shown how we can derive CIE codes from A-codes and how we can apply some of the known results of coding theory to get new results for the parameters of feasible A-codes.

3 Construction of an A-code from a q-ary linear code

Assume that we have a code C over $\mathbb{F}_q$ with the property

$$\forall \mathbf{c} \in C, \lambda \in \mathbb{F}_q \qquad \mathbf{c} + \lambda \mathbf{1} \in C. \tag{11}$$

Assume that the code C has parameters (n, M, d).

If $b - \widehat{b}$ is constant we compute, see (5),

$$\begin{aligned} \left| \left\{ j \,;\, c_j = b, \widehat{c}_j = \widehat{b} \right\} \right| &= \left| \left\{ j \,;\, c_j - \widehat{c}_j = b - \widehat{b} \right\} \right| \\ &= n - \mathrm{d}\left(\mathbf{c} - \widehat{\mathbf{c}}, (b - \widehat{b})\mathbf{1} \right) \\ &\leq n - d. \end{aligned}$$

However we have

$$d\Big(\mathbf{c}-\hat{\mathbf{c}},(b-\hat{b})\mathbf{1}\Big)=0,\ \text{if and only if } \mathbf{c}-\hat{\mathbf{c}}=\lambda\mathbf{1}.$$

Thus the words of C whose difference is a multiple of $\mathbf{1}$, would result in an A-code for which $P_S = 1$. We have to factor these words out of C.

The code C can be partitioned into equivalence classes by introducing the equivalence relation R as

$$\mathbf{a}\ R\ \mathbf{b} \text{ if } \mathbf{a}-\mathbf{b}=\lambda\mathbf{1} \text{ for some } \lambda\in\mathbb{F}_q.$$

Clearly each equivalence class contains q elements. Now let $[u]$ denote the equivalence class containing the codeword u. We form the quotient set

$$\hat{U}=C/\{\mathbf{1}\}=\{[u]\,;\,u\in C\}.$$

Now U is the code obtained by replacing each equivalence class by a specific representative. Clearly U has parameters $(n, M/q, d)$.

Now extend the code U to a new code V of length nq where

$$V=\{(\mathbf{u},\mathbf{u}+\alpha_1\mathbf{1},\mathbf{u}+\alpha_2\mathbf{1},\ldots,\mathbf{u}+\alpha_{q-1}\mathbf{1})\,;\,u\in U,\mathbb{F}_q=\{0,\alpha_1,\ldots,\alpha_{q-1}\}\}. \qquad (12)$$

By reversing the reasoning in the previous section we can use the code V to construct an A-code. Each codeword in V corresponds to a source state. Thus denote V as

$$V=\Big\{\mathbf{v}^{(s)}\,;\,s\in\mathcal{S}\Big\},$$

where $\mathbf{v}^{(s)} = (e_1(s), e_2(s), \ldots, e_{nq}(s))$. Now the property (12) gives that $\text{comp}(\mathbf{v}) = (1/q, \ldots, 1/q)$ for any $\mathbf{v} \in V$, so $P_I = 1/q$. To find the probability of successful substitution, let $\mathbf{v}^{(s)}$ and $\mathbf{v}^{(\hat{s})}$ be the two codewords corresponding to two distinct source states that maximize P_S. Assume that we have observed s with the tag b from $\mathbf{v}^{(s)}$ and replace it with $\hat{s}$ and tag $\hat{b}$ from $\mathbf{v}^{(\hat{s})}$. Then

$$\begin{aligned}
P_S &= \frac{\Big|\Big\{j\,;\,v_j^{(s)}=b, v_j^{(\hat{s})}=\hat{b}\Big\}\Big|}{\Big|\Big\{j\,;\,v_j^{(s)}=b\Big\}\Big|}\\
&= \frac{\Big|\Big\{j\,;\,v_j^{(s)}=b, v_j^{(s)}-v_j^{(\hat{s})}=b-\hat{b}\Big\}\Big|}{\Big|\Big\{j\,;\,v_j^{(s)}=b\Big\}\Big|}\\
&= \frac{\Big|\Big\{j\,;\,u_j^{(s)}-u_j^{(\hat{s})}=b-\hat{b}\Big\}\Big|}{n}.
\end{aligned}$$

Thus P_S is the maximal value of the composition values in $\text{comp}\Big(\mathbf{u}^{(s)}-\mathbf{u}^{(\hat{s})}\Big)$. Also

$$\text{comp}\Big(\mathbf{v}^{(s)}-\mathbf{v}^{(\hat{s})}\Big)=\text{comp}\Big(\mathbf{u}^{(s)}-\mathbf{u}^{(\hat{s})}\Big).$$

Let α be the index element for the maximal composition value. Consider instead $\text{comp}\Big(\mathbf{u}^{(s)}-(\mathbf{u}^{(\hat{s})}-\alpha\mathbf{1})\Big)$ in the code C. Here the maximum composition value is located at the index element 0 and then the maximum composition value is actually $1 - d/n$. Thus the maximum value of $\text{comp}\Big(\mathbf{v}^{(s)}-\mathbf{v}^{(\hat{s})}\Big)$ over all pairs in the code V is $1 - d/n$ and we have $P_S = 1 - d/n$.

Summarizing we have:

Theorem 3: Given a code C with parameters (n, M, d) such that if $\mathbf{c} \in C$, then $\mathbf{c}+\lambda\mathbf{1} \in C$ for all $\lambda \in \mathbf{F}_q$. Then there exists a corresponding A-code with parameters

$$|\mathcal{S}| = Mq^{-1}, \quad |\mathcal{E}| = nq, \quad P_I = 1/q \text{ and } P_S = 1 - d/n.$$

■

We refer to this construction as the *q-twisted* construction.

Now suppose we have a *linear* $(n, k+1)$ q-ary code C' with the property

$$\mathbf{1} \in C'. \tag{13}$$

It follows from the linearity of the code that property (11) holds. Thus we again get in the end a code V with the parameters given in Theorem 3 by applying the q-twisted construction.

Example 4: Let C be a Reed-Solomon (R-S) code of length q. Let $L = \{$all polynomials of degree $< k+1$ in $\mathbf{F}_q[x]\}$. Then the R-S code C can be described as

$$C = \{(f(0), f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_{q-1}))\,;\ f \in L, \mathbf{F}_q = \{0, \alpha_1, \ldots, \alpha_{q-1}\}\}\,.$$

If we now form the quotient code U, it will be as above but over all non-constant polynomials of degree $< k+1$ in $\mathbf{F}_q[x]$.

After the extension, the parameters are $|\mathcal{E}| = q^2$, $|\mathcal{S}| = q^k$, and since $d = q-(k+1)+1 = q-k$, we have $P_S = 1 - d/q = k/q$. We see that we obtained an A-code with the same parameters as in Construction 1. In fact, looking more closely, the two codes are the same up to renaming of source states and encoding rules. □

Let $A_q(n,d)$ be as usual, [7], the maximum number of codewords in a q-ary code of length n and with distance d. Now let $A_q^*(n,d)$ be the corresponding quantity if we add the property that $\mathbf{c} \in C$ implies $\mathbf{c}+\lambda\mathbf{1} \in C$ for all $\lambda \in \mathbf{F}_q$. Then

$$S(n, 1/q, P_S) \geq A_q^*\left(\frac{n}{q}, (1-P_S)\frac{n}{q}\right)/q,$$

where $S(n, 1/q, P_S)$ denotes the maximum number of source states as in (9).

For the special case of linear codes we have the following lemma:

Lemma 4: If there exists a linear code C with a codeword $\mathbf{c}$ such that every element in $\mathbf{c}$ is nonzero, then there also exists a linear code C' with the same parameters and such that $\mathbf{1} \in C'$. ■

Proof: Let the code C have generator matrix G. We can now do some elementary manipulations on G without changing the minimum distance. These operations include multiplication of columns by a nonzero scalar. Thus we multiply each column of G with the inverse of $\mathbf{c}$'s element in that column, getting a new code C'. Then $\mathbf{1} \in C'$. □

For nonlinear codes we can prove the Varshamov-Gilbert (V-G) bound even with the special restriction that $c \in C$ implies $c + \lambda\mathbf{1} \in C$.

Lemma 5: The maximum number of codewords in a q-ary code C of length n and minimum distance d such that $c \in C$ implies $c + \lambda\mathbf{1} \in C$, satisfies

$$A_q^*(n,d) \geq \frac{q^n}{V_q(n,d-1)},$$

where $V_q(n,d-1) = \sum_{i=0}^{d-1} \binom{n}{i}(q-1)^i$ is the size of a usual Hamming sphere around a codeword. ■

Proof: Consider the code with cardinality $A_q^*(n,d)$. If $A_q^*(n,d)V_q(n,d-1) < q^n$ there exists a point c which does not lie in any of the spheres. But this implies that $c + \lambda\mathbf{1}$ does not lie in any sphere either! This because if $c + \lambda\mathbf{1}$ is in the sphere around c' then $\mathrm{d}(c + \lambda\mathbf{1}, c') < d$ and thus $\mathrm{d}(c, c' - \lambda\mathbf{1}) < d$ and c is in the sphere around $c' - \lambda\mathbf{1}$ which is a contradiction. Thus we can add these q points as codewords and still have a code for which we have that if $c \in C$ then $c + \lambda\mathbf{1} \in C$, which contradicts the maximality of $A_q^*(n,d)$. □

Actually, this slightly strengthen the usual bound since $A_q^*(n,d)$ must be divisible by q.

Any known asymptotic bound[1] for q-ary codes is also valid for this extra condition, [8], since this is not a strong restriction when $n \to \infty$. Thus we have the same asymptotic behavior and the asymptotic V-G bound gives

$$S(|\mathcal{E}|, 1/q, P_S) \gtrapprox q^{g_q(P_S)|\mathcal{E}|/q}, \quad \text{for } P_S > 1/q, \tag{14}$$

where $g_q(x) = -x\log_q(x) - (1-x)\log_q(1-x) + x\log_q(q-1)$. Thus if we allow P_S to be larger than $P_I = 1/q$, then we can get A-codes that have an exponential number of source states.

Example 5: For $q = 2$, it follows from Example 1 that $|S|$ can be very close to $2^{\xi n}$, where $\xi \approx 1$ when $P_S \approx 1$. □

Example 6: Let us consider again Construction 1. Unfortunately, the upper bound (10) gives us not a coincidence of lower and upper bound for the case $|\mathcal{E}| \to \infty$, $P_S > 1/q$ fixed. □

For the binary case we give the upper bound (10). From the work of Wegman of Carter, [9, in the abstract], we have the following bound

$$\log|\mathcal{E}| \geq \log q + \log\log|S| - \log\log q.$$

From this we would obtain

$$|S| \leq q^{|\mathcal{E}|/q}. \tag{15}$$

But this bound is *not correct* since it does not depend on the P_S at all. Moreover, for $q = 2$. we get from (15) that $|S| \leq 2^{n/2}$, but we know from Example 5 that the size of S can be very close to 2^n.

[1] The V-G bound is known to be not optimal for $q = 49$

4 Conclusion

We have shown a relation between authentication codes and codes correcting independent errors and we have also shown how to construct A-codes from CIE codes. This is used to show that if P_S exceeds P_I by an arbitrarily small positive amount, then the number of source states in S grows exponentially with the number of keys and if $P_S = P_I$ it will grow only linearly. Furthermore we falsified a statement in Reference [9].

References

[1] E.N. Gilbert, F.J. MacWilliams and N.J.A. Sloane, "Codes which detect deception", Bell Syst. Tech. J., Vol. 53, no. 3, 1974, pp. 405-424.

[2] G.J. Simmons, "Authentication theory/coding theory", in *Advances in Cryptology, Proceedings of CRYPTO 84*, G.R. Blakley and D. Chaum, Eds. Lecture Notes in Computer Science, No. 196. New York, NY: Springer, 1985, pp. 411-431.

[3] G.J. Simmons, "A survey of Information Authentication", in *Contemporary Cryptology, The science of information integrety*, ed. G.J. Simmons, IEEE Press, New York, 1992.

[4] D.R. Stinson, "The combinatorics of authentication and secrecy codes", Journal of Cryptology, Vol. 2, no. 1, 1990, pp. 23-49.

[5] D.R. Stinson, "Universal hashing and authentication codes" *Proceedings of Crypto 91*, Santa Barbara, USA, 1991, pp. 74-85.

[6] F.J. McWilliams, N. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, New-York, 1977.

[7] S. Roman, *Coding and Information Theory*, Springer-Verlag, New York, 1992.

[8] M.A. Tsfasman, S.G. Vlăduţ, *Algebraic-Geometric Codes*, Kluwer Academic Publ., Dortrecht, 1991.

[9] M.N. Wegman, J.L. Carter, "New hash functions and their use in authentication and set equality", J. Computer and System Sciences, Vol. 22, 1981, pp. 265-279.