Minimal Codewords and Secret Sharing

James L. Massey
Signal and Information Processing Laboratory
Swiss Federal Institute of Technology
ETH-Zentrum
CH-8092 Zürich

Abstract: The use of a linear code to "split" secrets into equal-size shares is considered. The determination of which sets of shares can be used to obtain the secret leads to the apparently new notion of minimal codewords in a linear code. It is shown that the minimal codewords in the *dual code* completely specify the access structure of the secret-sharing scheme, and conversely.

1. Introduction

In an (S, T) threshold secret-sharing scheme as introduced by Shamir [1], a q-ary secret is "split" into S q-ary shares in such a manner that any T shares uniquely determine the secret but any T - 1 or fewer shares provide no information about the secret. Shamir constructed such (S, T) threshold schemes (where $1 \le T \le S < q$) by taking the secret to be the constant term in a monic polynomial of degree T over the finite field GF(q) whose T - 1 other coefficients are selected uniformly at random; the S shares are the values of this polynomial at any S specified and distinct non-zero elements of GF(q). McEliece and Sarwate [2] gave a generalization of Shamir's construction in terms of maximum-distance-separable (MDS) codes, i.e., linear q-ary (N, K) codes with minimum distance d = N - K + 1. The secret is chosen as the first digit of a codeword; the next K - 1 digits are chosen uniformly at random over GF(q) and the codeword then computed. The S = N - 1 shares are all the codeword digits after the first. The threshold is T = K because the digits in any K positions of a codeword in an MDS code uniquely determine the full codeword, i.e., any K positions are an *information set*. Shamir's construction is equivalent to using a (possibly punctured) Reed-Solomon code as the MDS code.

Attempts to modify the *access structure*, i.e., the sets of shares that determine the secret have generally employed shares of unequal size, which is practically undesirable and leads to complicated analyses. Thus, we will abide by Shamir's restriction that *all shares have equal size.*

As an example of the problem in which we are interested, suppose we wish to "split" an m-bit secret into m-bit shares for four users (Alice, Bob, Carol and David) such that Alice and Bob can determine the secret, as also can Bob and Carol and David, but no coalition of users not containing one of these two authorized coalitions as a subset can obtain any information about the secret. A solution to this problem is to choose the secret to be the first digit of a codeword in the (5, 3) code over $GF(2^m)$ with parity-check matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}, \qquad (1.1)$$

next to choose the second and fourth digits uniformly at random in GF(2^m) and to compute the full codeword [which is possible because positions 1, 2 and 4 form an information set], and finally to give the digits in positions 2, 3, 4 and 5 to Alice, Bob, Carol and David, respectively, as their shares of the secret. Letting [v_1, v_2, v_3, v_4, v_5] be the codeword, we see from (1.1) that the constraints on the codeword digits are

$$v_1 + v_2 + v_3 = 0 \tag{1.2}$$

and

$$v_2 + v_4 + v_5 = 0. \tag{1.3}$$

From (1.2), it follows that Alice (v_2) and Bob (v_3) can indeed determine the secret v_1. Adding (1.2) and (1.3) gives

$$v_1 + v_3 + v_4 + v_5 = 0, \tag{1.4}$$

which shows that Bob (v_3) and Carol (v_4) and David (v_5) can also find the secret v_1. It can be readily checked from (1.2) and (1.3) that no coalition of users not containing one of these two authorized coalitions as a subset can determine the secret -- the reason will become apparent in the sequel.

We will show that the linear code formulation of the general access problem leads naturally to what appears to be the new concept of "minimal codewords" in a linear code. In Section 2 we introduce this concept and its key properties. It is then a simple matter in Section 3 to treat the general access problem.

2. Minimal Codewords and their Properties

We will say that the q-ary N-tuple **v'** *covers* the q-ary N-tuple **v** if, in each position where **v** is non-zero, **v'** is also non-zero. We define a codeword **v** in a q-ary (N, K) code V to be *minimal* if (i) **v** is a non-zero codeword whose leftmost non-zero component is a 1 and (ii) **v** covers no other codeword **v'** whose leftmost non-zero component is a 1.

It follows immediately from the definition that *all minimum-weight codewords* in V *with leftmost non-zero component 1 are minimal codewords.* In an MDS code, the $\binom{N}{N-K+1}$ minimum-weight codewords with leftmost digit 1 are all the minimal codewords, as follows from the fact that any codeword of weight more than N - K + 1 (i.e., with fewer than K - 1 zero components) covers codewords of weight N - K + 1 (i.e., with K - 1 zero components) whose leftmost digit is 1.

No two distinct minimal codewords can be non-zero in the same set of positions for, were this to be the case, their difference (appropriately scaled) would be a codeword with leftmost non-zero component 1 that both of the former codewords would cover, which would contradict their minimality.

It is only slightly less obvious that *the set of minimal codewords is a spanning set for the code*, i.e., every codeword is a linear combination of minimal codewords. For if $\mathbf{v}'$ is a non-zero codeword but not minimal, it covers a minimal codeword $\mathbf{v_1}$. Thus, there is a scalar c_1 such that $\mathbf{v}'' = \mathbf{v}' - c_1\mathbf{v_1}$ is a codeword of weight strictly less than that of $\mathbf{v}'$. Repeating this argument for $\mathbf{v}''$ instead of $\mathbf{v}'$, etc., we must eventually find $\mathbf{v}' - c_1\mathbf{v_1} - \ldots - c_n\mathbf{v_n} = \mathbf{0}$ where $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ are minimal codewords and $n \leq N$. But then $\mathbf{v}'$ is a linear combination of the minimal codewords $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$, as was to be shown.

We now note in the construction used in the previous argument that $\mathbf{v}'$ covers $\mathbf{v}''$ as well as $\mathbf{v_1}$. Thus, we have in fact proved the much stronger result that *every non-zero non-minimal codeword is a linear combination of those minimal codewords that are covered by this non-zero codeword.* This immediately implies the following result that we will need in the sequel.

Lemma: Every non-minimal non-zero codeword covers a minimal codeword whose leftmost non-zero component (necessarily a 1) occurs in the same position as the leftmost non-zero component of this non-minimal codeword.

3. Access Structures and Minimal Codewords

We assume hereafter that we use a secret-sharing system determined by a linear q-ary (N, K) code V in the manner that (i) the secret is chosen as the first digit of a codeword; (ii) the digits in K - 1 positions, selected so that together with the first position they form an information set [which is possible for any code for which the first digit in every codeword is not always 0] are selected uniformly at random over GF(q) and the full codeword then computed; and (iii) the S = N - 1 shares are all the codeword digits after the first.

We now define the *access structure* of a secret-sharing scheme to be the class consisting of all sets of shares such that the shares in each set uniquely determine the secret but, if any share is removed from this set, the remaining shares give no information about the secret. A coalition of users can then determine the secret if and only if their set of shares contains a subset that is in the access structure. Our main result is the following.

Proposition: The access structure of the secret-sharing scheme corresponding to the linear q-ary (N, K) code V is specified by those minimal codewords in the *dual code* $V^{\perp}$ whose first component is a 1 in the manner that the set of shares specified by each such minimal codeword in the dual code is the set of shares corresponding to those locations after the first where this minimal codeword is non-zero.

Example: Because the dual code $V^{\perp}$ is the row space of any parity-check matrix H for a linear code V, it follows that the 2^m-ary dual code $V^{\perp}$ corresponding to H in (1.1) has only two minimal codewords with first component 1, namely

$$\mathbf{v_I} = [1, 1, 1, 0, 0] \tag{3.1}$$

and

$$\mathbf{v_{II}} = [1, 0, 1, 1, 1]. \tag{3.2}$$

This can be seen by noting from (1.1) that $\mathbf{v_I}$ is the only codeword in $V^{\perp}$ with first component 1 that has zeroes in either position 4 or position 5; that $\mathbf{v_{II}}$ is the only codeword in $V^{\perp}$ with first component 1 that has a 0 in position 2; and that there is no codeword with first component 1 that has a zero in position 3. Thus, the access structure of this secret-sharing scheme consists of the sets $\{v_2, v_3\}$ and $\{v_3, v_4, v_5\}$, as we claimed in Section 1.

It remains to prove the proposition. Suppose that $\{v_2, v_3, ..., v_t\}$ is any set of shares that determines the secret v_1, where we have numbered the shares in this set consecutively for ease of notation. Equivalently, there exist scalars $c_2, c_3, ..., c_t$ such that

$$v_1 = c_2v_2 + c_3v_3 + ... + c_tv_t \tag{3.3}$$

holds for all $\mathbf{v} = [v_1, v_2, ..., v_N]$ in V. Again equivalently,

$$\mathbf{v}' = [1, -c_1, -c_2, ... -c_t, 0, ..., 0] \tag{3.4}$$

is a codeword in $V^{\perp}$ whose leftmost non-zero digit is in position 1. But $\{v_2, v_3, ..., v_t\}$ is in the access structure if and only if $c_2, ..., c_t$ are all non-zero for every choice of these scalars such that (3.3) is satisfied. Equivalently, $\mathbf{v}'$ is in $V^{\perp}$ and covers no vector in $V^{\perp}$ whose first component is a 1. It now follows from the Lemma that this is in turn equivalent to the condition that $\mathbf{v}'$ be a minimal codeword in $V^{\perp}$ with leftmost non-zero component 1, as was to be shown.

4. Concluding Remark

We have reduced the problem of realizing an access structure to the problem of constructing a linear code (the dual code $V^{\perp}$) whose minimal codewords with first component 1 correspond to the sets of shares in the access structure. This reduction suggests a host of new problems in secret sharing that can now be attacked with the well developed tools of algebraic coding theory.

References

[1] A. Shamir, "How to share a secret", *Comm. ACM*, Vol. 22, pp. 612-613, November 1979.

[2] R. J. McEliece and D. V. Sarwate, "On sharing secrets and Reed-Solomon codes", *Comm. ACM*, Vol. 24, pp. 583-584, September 1981.