A construction for authentication / secrecy codes from certain combinatorial designs

D. R. Stinson
Department of Computer Science
University of Manitoba

Abstract

If we agree to use one of v possible messages to communicate one of k possible source states, then an opponent can successfully impersonate a transmitter with probability at least k / v, and can successfully substitute a message with a fraudulent one with probability at least (k - 1) / (v - 1). We wish to limit an opponent to these bounds. In addition, we desire that the observation of any two messages in the communication channel will give an opponent no clue as to the two source states. We describe a construction for a code which achieves these goals, and which does so with the minimum possible number of encoding rules (namely, v·(v - 1) / 2). The construction uses a structure from combinatorial design theory known as a perpendicular array.

1. Authentication and secrecy

In this paper, we study the properties of codes with respect to secrecy and authentication. We are interested in the *unconditional*, or *theoretical*, security provided by such codes. That is, we assume that any opponents have unlimited computational resources. The theory of unconditional secrecy is due to Shannon [12]. More recently, Simmons has developed an analogous theory of unconditional authentication.

We shall use the model of authentication theory as described by Simmons in [13], [14], and [15]. In this model, there are three participants: a transmitter, a receiver, and an opponent. The *transmitter* wants to communicate some information to the *receiver*, whereas the *opponent* wants to deceive the receiver. The opponent can either impersonate the receiver, making him accept a fradulent message as authentic; or, modify a message which has been sent by the transmitter.

More formally, we have a set of k source states S, a set of v messages M, and a set of b encoding rules E. A *source state* $s \in S$ is the information that the transmitter wishes to communicate to the receiver. The transmitter and reciever will have secretly chosen an *encoding rule* $e \in E$

C. Pomerance (Ed.): Advances in Cryptology - CRYPTO '87, LNCS 293, pp. 355-366, 1988.

beforehand. An encoding rule e will be used to determine the *message* e(s) to be sent to communicate any source state s. It is possible that more than one message can be used to determine a particular source state (this is called *splitting*). However, in order for the receiver to be able to uniquely determine the source state from the message sent, there can be at most one source state which is encoded by any given message $m \in M$ (i.e. $e(s) \neq e(s')$ if $s \neq s'$).

We are interested in the security of such a code with respect to both *secrecy* and *authentication*. Suppose an opponent observes i distinct messages being sent over the communications channel (where $i \geq 0$). He knows that the same key is being used to transmit the i messages, but he does not know what that key is. If we consider the code as a secrecy system, then we make the assumption that the opponent can only observe the messages being sent. Our goal is that the opponent be unable to determine any information regarding the i source states from the i messages he has observed.

In [8] and [11], the following scenario for authentication is investigated. As before, an opponent observes i distinct messages. The opponent then sends a message m' to the receiver, hoping to have it accepted as authentic (this message m' must be distinct from the i messages already sent). In [8], Massey calls this a *spoofing attack* of order i. We remark that the special cases $i = 0$ and $i = 1$ have been studied extensively by Simmons and other people (see [1], [13], [14], [15], and [16]). The case $i = 0$ is called the *impersonation* game, and the case $i = 1$ is called the *substitution* game.

For any i, there will be a probability distribution on the set of i source states which occur. We ignore the order in which the i source states occur, and assume that no source state occurs more than once. Also, we assume that *any* set of i source states has a non-zero probability of occurring. Given a set of i source states S, we define p(S) to be the probability that the source states in S occur.

Given the probability distributions on the source states described above, the receiver and transmitter will choose a probability distribution for E, called an *encoding strategy*. If splitting occurs, then they will also determine a *splitting strategy* to determine $m \in M$, given $s \in S$ and $e \in E$ (this corresponds to non-deterministic encoding).

Once the transmitter / receiver have chosen encoding and splitting strategies, we can define for each $i \geq 0$ a probability denoted Pd_i, which is the probability that the opponent can deceive the transmitter / receiver with a spoofing attack of order i.

In this paper, we consider only codes without splitting. We shall use the following notation. Given any encoding rule e, we define $M(e) = \{e(s): s \in S\}$, i.e. the set of messages permitted by encoding rule e. For a set $\mathbf{M}'$ of distinct messages, and an encoding rule e, define $f_e(\mathbf{M}') = \{s: e(s) \in \mathbf{M}'\}$, i.e. the set of source states which will be encoded under encoding rule e by a message in $\mathbf{M}'$. As well, for a set $\mathbf{M}'$ of distinct messages, define $E(\mathbf{M}') = \{e \in E: \mathbf{M}' \subseteq M(e)\}$, i.e. the set of encoding rules under which all the messages in $\mathbf{M}'$ are permitted. It is useful to think of a code as being represented by a b x k matrix, where the rows are indexed by encoding rules, the columns are indexed by source states, and the entry in row e and column s is e(s).

Theorem 1 [8, p. 12]. In an authentication system without splitting,

$$Pd_i \geq \frac{k-i}{v-i}.$$

Proof: Suppose the opponent observes the i messages in the set $\mathbf{M}' = \{m_1, \ldots m_i\}$ in the channel. For $m \in \mathbf{M} \setminus \mathbf{M}'$, let payoff($\mathbf{M}'$, m) denote the probability that message m would be accepted as authentic. Then we have

$$\text{payoff}(m, \mathbf{M}') = \frac{\sum_{e \in E(\mathbf{M}' \cup \{m\})} p(e) \cdot p(S = f_e(\mathbf{M}'))}{\sum_{e \in E(\mathbf{M}')} p(e) \cdot p(S = f_e(\mathbf{M}'))}.$$

It is not difficult to calculate

$$\sum_{m \in \mathbf{M} \setminus \mathbf{M}'} \text{payoff}(m, \mathbf{M}') = k - i.$$

Hence, there exists some $m \in \mathbf{M}' \setminus \mathbf{M}$ such that payoff(m, $\mathbf{M}'$) $\geq (k - i) / (v - i)$. For every set $\mathbf{M}'$ of i source states, the opponent can choose such an m. This proves that $Pd_i \geq (k - i) / (v - i)$.•

Following Massey [8], we say that the authentication system is *L-fold secure against spoofing* if $Pd_i = (k - i) / (v - i)$ for $0 \leq i \leq L$.

When we consider the secrecy properties of a code, we desire that no information be conveyed by the observation of the messages which are transmitted. We say that a code has *perfect L-fold secrecy* if, for every set M_1 of at most L messages observed in the channel, and for every set S_1 of at most $|M_1|$ source states, we have $p(S_1 \mid M_1) = p(S_1)$. That is, observing a set of at most L messages in the channel does not help the opponent determine the L source states.

The purpose of this note is to give a simple construction for a system which achieves perfect 2-fold secrecy and is 1-fold secure against spoofing, and does so with the minimum possible number of keys, as given by the following bound.

Theorem 2 If a code achieves perfect L-fold secrecy and is (L - 1)-fold secure against spoofing, then

$$b \geq \binom{v}{L}.$$

Proof: Let $\mathbf{M}_1$ be a set of $i \leq L - 1$ messages which are permitted under a particular encoding rule. Let x be any message not in $\mathbf{M}_1$. Suppose there is no encoding rule under which all messages in $\mathbf{M}_1 \cup \{x\}$ are valid. Then a suitable modification of the proof of Theorem 1 shows that we would have $Pd_i > (v - i) / (k - i)$, a contradiction. Hence, it follows that every L-subset of messages is valid under at least one encoding rule.

Now, pick any L-subset of messages $\mathbf{M}_2$. In order to achieve perfect L-fold secrecy, the messages in $\mathbf{M}_2$ must encode every possible L-subset of source states. Hence, $\mathbf{M}_2$ is a valid set of messages under at least $\binom{k}{L}$ encoding rules. Now, if we count L-subsets of messages, we get

$$b \cdot \binom{k}{L} \geq \binom{v}{L} \cdot \binom{k}{L}.$$

This completes the proof.•

In the remainder of the paper, we shall study the existence of codes where the bound of Theorem 2 is met with equality. Hence, we define on *optimal L-code* to be a code which achieves perfect L-fold secrecy, is (L - 1)-fold secure against spoofing, and has precisely $\binom{v}{L}$ encoding rules.

We give a construction for optimal 2-codes in Section 2. We remark that a construction for codes which provide perfect 1-fold secrecy and are 1-fold secure against spoofing was given in [16].

2. Constructions for optimal 2-codes using perpendicular arrays

Our interest is in constructing optimal L-codes. We can do this for L = 2 using a type of combinatorial design known as a perpendicular array. These arrays were first studied in [9], and have been investigated by several researchers in combinatorial design theory since then (see [5], [6] and [7]). A *perpendicular array* PA(n, k) is a $v \cdot (v - 1) / 2$ by k array, A, of the symbols $\{1, \ldots, v\}$, which satisfies the following property:

> for any two columns i and j of A, and for any two distinct symbols $x, y \in \{1, \ldots, v\}$, there is a unique row r such that $\{A(r, i), A(r, j)\} = \{x, y\}$.

We have the following construction using perpendicular arrays.

Theorem 3 If there exists a PA(v, k), where $k > 2$, then there is a code for k source states with v messages and $v \cdot (v - 1) / 2$ encoding rules, which achieves perfect 2-fold secrecy and is 0-fold secure against spoofing.

Proof: We construct an encoding rule from each row r of the perpendicular array A: for each row $r = (x_1, \ldots, x_k)$ of A, we define an encoding rule $e_r(s) = (x_s: 1 \leq s \leq k)$. We shall use each encoding rule with probability $2 / (v \cdot (v - 1))$.

Let's first verify that $Pd_0 = k / v$. This follows immediately from the following easily proved property of perpendicular arrays: if $k > 2$, then every symbol occurs exactly $(v - 1) / 2$ times in each column of a PA(v, k).

Next, we check that we have perfect 2-fold secrecy. Again, this is an almost immediate consequence of the definition of perpendicular array. Given any two messages m and m', and given any two source states s and s', there is exactly one encoding rule e such that $\{e(s), e(s')\} = \{m, m'\}$. Hence, we have $p(S = \{s, s'\} \mid \{m, m'\}) = p(S = \{s, s'\})$.•

The following example illustrates that a code constructed by means of Theorem 2 will not necessarily be 1-fold secure against spoofing.

Example 1 The following is a PA(5, 3):

0	1	2
1	2	3
2	3	4
3	4	0
4	0	1
0	3	1
1	4	2
2	0	3
3	1	4
4	2	0

Suppose the source probability distribution is (p_1, p_2, p_3), where $p_1 > p_2 > p_3$. What happens if the opponent observes the message 0 being transmitted? The conditional probability distribution on the encoding rules, given that message 0 is observed, is:

$$\left(\frac{p_1}{2}, 0, 0, \frac{p_3}{2}, \frac{p_2}{2}, \frac{p_1}{2}, 0, \frac{p_2}{2}, 0, \frac{p_3}{2}\right).$$

If the opponent substitutes message 0 with message 1, it will be accepted as authentic with probability

$$\frac{p_1}{2} + \frac{p_2}{2} + \frac{p_1}{2}$$

$$= \frac{1}{2} + \frac{p_1 - p_3}{2}.$$

In fact, the opponent's optimal substitution strategy is to replace any message m by the message (m + 1) mod 5. This yields

$$Pd_1 = \frac{1}{2} + \frac{p_1 - p_3}{2}.$$

A special type of perpendicular array will allow us to attain $Pd_1 = (k - 1) / (v - 1)$. A PA(v, k) A is said to be *cyclic* (and is denoted CPA(v, k)) if any cyclic permutation of the columns of A yields an array which can be obtained from A by means of a suitable permutation of the rows of A. That is, if $(x_1, \ldots, x_k)$ is a row of A, then $(x_2, \ldots, x_k, x_1)$ is also a row of A.

Example 2 A cyclic PA(5, 5).

0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
3	4	0	1	2
4	0	1	2	3
0	2	4	3	1
1	3	0	4	2
2	4	1	0	3
3	0	2	1	4
4	1	3	2	0

We have the following

Theorem 4 If there exists a cyclic PA(v, k), then there is an optimal 2-code for k source states with v messages.

Proof: Let A be a CPA(v, k). Construct the code as in Theorem 2. We need only verify that $Pd_1 = (k - 1) / (v - 1)$. Let m and m' be two distinct messages. We have

$$\text{payoff}(m, m') = \frac{\sum_{e \in E(m, m')} p(e) \cdot p(S = f_e(m))}{\sum_{e \in E(m')} p(e) \cdot p(S = f_e(m))}$$

$$= \frac{\sum_{e \in E(m, m')} p(S = f_e(m))}{\sum_{e \in E(m')} p(S = f_e(m))}$$

$$= \frac{\sum_{e \in E(m, m')} p(S = f_e(m))}{\frac{v-1}{2}}.$$

Now, there are k·(k - 1) / 2 rows r of A for which m, m' occur in row r. For each source state j, there are exactly (k - 1) / 2 encoding rules e_r where m, m' occur in row r and $e_r(j) = m$. Then,

$$\sum_{e \in E(m, m')} p(S = f_e(m)) = \frac{k-1}{2}.$$

Hence, payoff(m, m') = (k - 1) / (v - 1), as desired. This is true for any two messages m, m'. Hence, no matter what the opponent's substitution strategy, he will deceive the receiver with this probability.

Cyclic perpendicular arrays have been the subject of several recent papers, such as [2], [5] and [6]. It is not difficult to see that the existence of a CPA(v, k) requires that v and k be odd, and that 2k | v·(v - 1). For k = 3 and 5, these conditions are necessary and sufficient for existence, with one exception.

Theorem 5

1) ([6]) A CPA(v, 3) exists if and only if $v \equiv 1$ or 3 modulo 6.
2) ([5]) A CPA(v, 5) exists if and only if $v \equiv 1$ or 5 modulo 10, $v \neq 15$.

For k > 5, only sporadic results are known. One class of cyclic PAs is given by

Theorem 6 ([2]) k is odd and $v \equiv 1$ modulo 2k is a prime power, then there is a CPA(v, k).

We will discuss the construction of Theorem 6 in some detail in Section 3, but let's first observe that, although the existence of a CPA(v, k) is sufficient for the existence of an optimal 2-code, it is not necessary. We shall say that a PA(v, k) is *pair-column balanced* if, for every pair of symbols x, y, the following property is satisfied:

> Among the rows containing x and y, x and y each occur (k - 1) / 2 times in each column.

Lemma 7 If there exists a pair-column balanced PA(v, k), then there exists an optimal 2-code for k source states with v messages.

Proof: The proof is identical to that of Theorem 4.•

Example 3 (van Rees [17]) A pair-column balanced PA(11, 3). Develop the following 5 rows modulo 11:

0	1	2
0	9	7
0	3	6
0	4	8
0	5	10

We have the following result concerning pair-column balanced PA(v, 3).

Theorem 8 There exists a pair-column balanced PA(v, 3) for all odd $v \geq 3$, $v \neq 5, 17$. Further, there does not exist a pair-column balanced PA(5, 3).

Proof: This result follows easily from the following recursive pairwise balanced design construction for PAs ([7, Lemma 4.1]). Let v be a positive integer, and let $K \subseteq \{2, \ldots, v - 1\}$. A (v, K)-PBD (*pairwise balanced design*) is a set X of v elements (points) and a set **B** of subsets of X (blocks), such that every (unordered) pair of points occurs in a unique block $B \in \mathbf{B}$, and $|B| \in K$ for every $B \in \mathbf{B}$. Suppose we have a (v, K)-PBD, and for every $n \in K$, there exists a PA(n, k). Then we can construct a PA(v, k), by taking a PA(|B|, k) on symbol set B, for every $B \in \mathbf{B}$. It is easy to check if every "input" PA(|B|, k) is pair-column balanced, then so is the resulting PA(v, k).

Now, we already have that there is a pair-column balanced PA(n, 3) for every $n \equiv 1$ or 3 mod 6, and for n = 11. In [4, Theorem 3.3], it is shown that there exists a (v, {3, 11})-PBD for all $v \equiv 5$ modulo 6, $v \geq 23$. The above construction produces pair-column balanced PA(v, 3) for all these values of v. There remain only v = 5 and v = 17 to consider. An exhaustive search ([17]) has shown that no pair-column balanced PA(5, 3) exists. The case v = 17 remains open.•

Apparently, no examples of pair-column balanced PA(v, 5) are known, other than the cyclic PAs.

3. Implementing optimal 2-codes

In this section, we describe the construction and implementation of the optimal 2-codes guaranteed by Theorem 6.

Suppose k is odd and $v \equiv 1$ modulo $2k$ is a prime power. The perpendicular array PA(v, k) is constructed as follows ([6]). Let ω be a primitive element in the finite field GF(v), and let $\alpha = \omega^{(n-1)/k}$. For each $i = 1, \ldots, (v-1)/2k$, for each $j = 0, \ldots, k-1$, and for each $\beta \in GF(v)$, define a row

$$\beta + \omega^i\alpha^j \qquad \beta + \omega^i\alpha^{1+j} \qquad \beta + \omega^i\alpha^{2+j} \qquad \ldots \qquad \beta + \omega^i\alpha^{k-1+j}$$

This defines $v \cdot (v-1)/2$ rows, which is the right number, at least. It is not difficult to see that the resulting array is indeed a PA(v, k). Also, it is clearly cyclic: the rows obtained by varying j, for fixed i and β, are all cyclic shifts.

Hence, we can pick a random encoding rule e by generating a random 3-tuple (i, j, β) of the form described above. A source state s ($0 \le s \le k-1$) would then be encoded as

$$e(s) = \beta + \omega^i\alpha^{s+j}.$$

Also, given an encoded message m, we can solve for the source state s by means of the equation

$$\alpha^{s+j} = (m - \beta) / \omega^i.$$

Observe that this requires the calculation of a logarithm in the finite field GF(v). This is made easier by the knowledge that $0 \le s \le k-1$. As well, if v is a prime, and $v-1$ has only small prime factors, then an algorithm of Pohlig and Hellman ([10]) can be used which has computational complexity $O((\log v)^2)$.

Finally, let's consider the sizes of messages and encoding rules, as a function of the authentication security. Recall that we have $Pd_0 = k/v$ and $Pd_1 = (k-1)/(v-1)$, where we have k source states, m messages, and $v \cdot (v-1)/2$ encoding rules. Since we want Pd_0 and Pd_1 to be small, we might consider taking $v = k \cdot (k-1) + 1$ (assuming, of course, that it is a prime power). In this case, we

have $Pd_0 \approx 1 / (k - 1)$ and $Pd_1 = 1 / k$. We require roughly $2 \cdot \log_2 k$ message bits in order to transmit $\log_2 k$ bits of source. If we send two messages, then we transmit $2 \cdot \log_2 k$ bits of source with $4 \cdot \log_2 k$ bits of message, with perfect secrecy. The encoding rule requires about $4 \cdot \log_2 k$ bits.

It is interesting to compare these space requirements with that of the well-known "one-time pad". The one-time pad achieves perfect secrecy by requiring one bit of key (analogous to our encoding rules) for every bit of source. It is also well-known that this much key is required if perfect secrecy is to be attained. In the example we have constructed above, we have 2 bits of "key" for every bit of source communicated (when two messages are sent). If we send only one message, then we have 4 bits of "key" for every bit of source, but we gain a high degree of security with respect to authentication. It bears repeating that these results cannot be achieved with less "key" (Theorem 2).

References

1. Ernest F. Brickell, A few results in message authentication, Congressus Numerantium 43 (1984), 141-154.

2. A. Granville, A. Moisiadis and R. Rees, Nested Steiner n-gon systems and perpendicular arrays, preprint.

3. E. Gilbert, F. J. MacWilliams and N. J. A. Sloane, Codes which detect deception, Bell System Tech. J. 53 (1974), 405-424.

4. C. Huang, E. Mendelsohn and A. Rosa, On partially resolvable t-partitions, Annals Disc. Math. 12 (1983), 169-183.

5. C. C. Lindner and D. R. Stinson, Steiner pentagon systems, Discrete Math. 52 (1984), 67-74.

6. C. C. Lindner and D. R. Stinson, The spectrum for the conjugate invariant subgroups of perpendicular arrays, Ars Combinatoria 18 (1984), 51-60.

7. C. C. Lindner, R. C. Mullin and G. H. J. van Rees, Separable orthogonal arrays, Utilitas Math., to appear.

8. J. L. Massey, Cryptography - A selective survey, in "Digital Communications" (1986), 3-21.

9. R. C. Mullin, P. J. Schellenberg, G. H. J. van Rees and S. A. Vanstone, On the construction of perpendicular arrays, Utilitas Math. 18 (1980), 141-160.

10. S. C. Pohlig and M. E. Hellman, An improved algorithm for computing logarithms over GF(p) and its cryptographic significance, IEEE Trans. on Inform. Theory 24 (1978), 106-110.

11. P. Schobi, Perfect authentication systems for data sources with arbitrary statistics, preprint.

12. C. E. Shannon, Communication theory of secrecy systems, Bell System Tech. J. 28 (1949), 656-715.

13. Gustavus J. Simmons, A game theory model of digital message authentication, Congressus Numerantium 34 (1982), 413-424.

14. Gustavus J. Simmons, Message Authentication: A game on hypergraphs, Congressus Numerantium 45 (1984), 161-192.

15. Gustavus J. Simmons, Authentication theory / Coding theory, in "Advances in Cryptology: Proceedings of CRYPTO 84", Lecture Notes in Computer Science, vol. 196, 411-432, Springer Verlag, Berlin, 1985.

16. D. R. Stinson, Some constructions and bounds for authentication codes, J. of Cryptology, to appear.

17. G. H. J. van Rees, private communication.